



## Where do successful populations originate from?

Peter Andras\*, Adam Stanton

School of Computing and Mathematics, Keele University, Newcastle-under-Lyme, Staffordshire ST5 5BG, UK



### ARTICLE INFO

#### Article history:

Received 1 November 2020

Revised 16 April 2021

Accepted 22 April 2021

Available online 1 May 2021

#### Keywords:

Computational modelling

Socio-technical evolution

Socio-biological simulation

Human geography

Geography of speciation

### ABSTRACT

In order to understand the dynamics of emergence and spreading of socio-technical innovations and population moves it is important to determine the place of origin of these populations. Here we focus on the role of geographical factors, such as land fertility and mountains in the context of human population evolution and distribution dynamics. We use a constrained diffusion-based computational model, computer simulations and the analysis of geographical and land-quality data. Our analysis shows that successful human populations, i.e. those which become dominant in their socio – geographical environment, originate from lands of many valleys with relatively low land fertility, which are close to areas of high land fertility. Many of the homelands predicted by our analysis match the assumed homelands of known successful populations (e.g. Bantus, Turkic, Maya). We also predict other likely homelands as well, where further archaeological, linguistic or genetic exploration may confirm the place of origin for populations with no currently identified urheimat. Our work is significant because it advances the understanding of human population dynamics by guiding the identification of the origin locations of successful populations.

© 2021 Elsevier Ltd. All rights reserved.

## 1. Introduction and background

Identifying the place of origin or homeland, the urheimat, of current and historical populations is a fascinating research challenge (De Barros Damgaard et al., 2018; Diamond and Bellwood, 2003; Driem, 1993; Gray et al., 2009; Mathieson et al., 2018; Piazza et al., 1995). Populations may be defined by their language and/or material and customary culture. Recently many studies have used genomic analysis of ancient humans to identify the most likely human population migration paths that led to the population of various parts of the world (Lazaridis et al., 2014; Nielsen et al., 2017; Slatkin and Racimo, 2016). Computational simulations have proved powerful in the evaluation of hypotheses about historical human population migrations (Fedotov et al., 2008; Isern et al., 2017; Montenegro et al., 2016).

Modern human populations evolved during the last 200,000 years (Campbell and Tishkoff, 2010; Cavalli-Sforza, 1997; Nielsen et al., 2017; Slatkin and Racimo, 2016) through many cultural changes that led to the emergence of successful populations that grew, spread, and became dominant over larger territories for a period of time (Bouckaert et al., 2012; Cavalli-Sforza, 1997; Chiaroni et al., 2008; Diamond and Bellwood, 2003; Mellars,

2006; Piazza et al., 1995). Archaeological and genetic evidence indicates that African human populations grew rapidly around 80,000 years ago, followed by spreading to Asia and Europe between 60,000 and 40,000 years ago (Mellars, 2006; Owens and King, 1999; Slatkin and Racimo, 2016). Linguistic and genetic evidence indicates that Indo-Europeans are likely to have originated from the plains between the Dniester and the Volga rivers (Diamond and Bellwood, 2003; Mathieson et al., 2018), with deeper roots in the Caucasus and Northern Iran (Mathieson et al., 2018), around 6500 B.C.E. and spread from here to become dominant in Europe and South Asia. In Africa the Bantu spread out from the border areas of Nigeria and Cameroon around 5000 years ago and became dominant in most of Sub-Saharan Africa (Campbell and Tishkoff, 2010). There are several populations that can trace their origins to north-eastern Central Asia (e.g. Mongols, Turks) (De Barros Damgaard et al., 2018), to the Ural-Volga area (e.g. Finns, Hungarians) (Piazza et al., 1995), the Fertile Crescent (e.g. Jews, Arabs) (Kitchen et al., 2009), to the mountains of Yunnan (e.g. Tibetans) (Driem, 1993), or to Taiwan (e.g. Malays, Polynesians) (Gray et al., 2009). Genetic and linguistic markers may not point to the same origins for a given population (Cavalli-Sforza, 1997), but both indicate earlier successful populations (note that the genetic or linguistic markers themselves do not determine the success of the population, which depends on socio-technical/cultural features of the populations, the markers only help tracing the origins of the populations).

\* Corresponding author at. School of Computing and Mathematics, Keele University, Newcastle-under-Lyme, Staffordshire ST5 5BG, UK.

E-mail address: [p.andras@keele.ac.uk](mailto:p.andras@keele.ac.uk) (P. Andras).

Successful populations make better use of available environmental and human resources than other populations. Such populations are likely to gain resources at the expense of other populations that are less resource-efficient. Consequently, successful populations expand and spread. Less successful populations may get displaced and may become extinct. The better use of resources may include milk consumption, agricultural or military innovations, or more resilient social organization. Here we assume that the emergence of new successful populations is due to cultural (socio-technical) innovations that allow a population to become better than its neighbors in efficient utilization of environmental and human resources.

## 2. Mathematical and computational modelling

Geographical constraints are important for the evolution of human populations (Coop et al., 2009). The fragmentation of the environment, due to mountains, rivers, sea shore, and variable availability of food resources, has been recognised as an important factor for the evolution and speciation of animal populations (Debinski and Holt, 2000; Ryberg et al., 2013). It is reasonable to assume that such factors are important for the emergence of successful human populations. Our hypothesis is that the geographical features that matter most are the distances from mountain ridges and high-fertility lands, and the fertility of the land where the population resides. This hypothesis is based on the following intuitive reasoning. Being near to mountains creates natural defence lines for the population, making it harder for other populations to invade and displace the current population (Jeong et al., 2016). Being near to high-fertility land provides the opportunity to spread into resource-rich areas, which can support population growth. In general, it is assumed that a population that can utilise their resources more efficiently (e.g. better agricultural practices that give higher yield of produce; more effective fighting practices and tools, which increase the likelihood of winning in armed conflicts), will dominate and potentially displace another population that is less efficient in the use of their resources. Residing in fertile lands offers resources to grow the population but reduces the growth potential difference between populations with similar resource utilization efficiency. Residing in harsh lands offers less resource for growth, but supports the relatively quicker growth of a population that is more efficient in resource utilization than others, contributing to the emergence of locally dominant populations. This paper aims to test the stated hypothesis using computational modelling and the analysis of real world geographical data.

Usually the modeling of the spreading of populations (humans, animals) relies on reaction-diffusion equations that define the spreading (diffusion) of the population, taking into consideration constraints on the spreading, and also the effects of the environment (reaction part) in terms of population loss and gain (e.g. mortality, reproduction) (Fedotov et al., 2008; Flather and Bevers, 2002; García-Ramos and Rodríguez, 2002; Volpert and Petrovskii, 2009). The general equation for such models is

$$\frac{\partial p(x, t)}{\partial t} = \nabla(D(p, x) \cdot \nabla p(x, t)) + F(p, x, t) \quad (1)$$

where  $p(x, t)$  is the population size at time  $t$  at spatial position  $x$ , the first component on the right side represents the diffusion effect and the second the population change that depends on the population size, location and time, which is the reaction term of the equation.  $D(p, x)$  represents the diffusivity of the environment and includes the spatial constraints. The equation may include a stochastic component in either part to represent random variation of parameters. In general, the explicit description of population spreading in space and time cannot be calculated. An approximate solution may be

found by simulating a discretized reaction-diffusion process (Flather and Bevers, 2002).

To study the impact of geographical constraints on emergence of successful populations we formulated a stochastic reaction-diffusion model of population spreading with multiple populations in the presence of such constraints (Andras, 2015) and implemented it through a computer simulation (Isern et al., 2017; Montenegro et al., 2016). Geographical constraints are included as simulated mountain ridges in the diffusivity specification, and the variability of the simulated land fertility is included in the reaction part of the equation. We built a discretized simulation of the model using a two dimensional grid of squares for spatial positions (Andras, 2015). The size of the  $j$ -th population at time  $t$  and location  $x$  is denoted as  $p_j(x, t)$ . The characteristic resource utilization ability of this population is  $\pi_j$ , represented as a bit string. Each grid location may be occupied by multiple populations. The update equation for the population size at a given location is:

$$p_j(x, t + 1) - p_j(x, t) = \sum_{\tau \in T} (r_{x+\tau}^{t,j} \cdot p_j(x + \tau, t) - r_{x,\tau}^{t,j} \cdot p_j(x, t)) + (f_{x,t}(\pi_j) - 1) \cdot p_j(x, t) \quad (2)$$

where  $T = \{(-1, 0), (1, 0), (0, -1), (0, 1)\}$ ,  $f_{x,t}$  are the effective resource utilization efficiency functions for a given location and time, considering land fertility and the simultaneous presence of other populations, and  $r_{x,\tau}^{t,j}$  are stochastic diffusivity parameters. Here the effective resource utilization efficiency includes the effect of competition between populations. Having multiple populations at the same location means that they need to share the locally available resources, while they have different efficiency of using their share of resources. The  $f_{x,t}$  determines the local reaction term of the diffusion – reaction system, and it shows how the local population grows by itself (i.e. ignoring the diffusion aspect given by population movements). Populations that are more efficient in their resource utilisation will have a higher growth rate and difference between the growth rates of different populations will also be influenced by the availability of resources at the location of the populations. We note that the choice of  $T = \{(-1, 0), (1, 0), (0, -1), (0, 1)\}$  for the diffusion neighbourhood is for convenience and simplicity (this is similar to the von-Neumann neighbourhood in the context of cellular automata). In effect this means that there is no direct population diffusion in diagonal direction, but this does not change the generality of the simulations, since diagonal direction diffusion will still happen, just not in one step, but in two steps.

The stochastic diffusivity parameters are given as

$$r_{x,\tau}^{t,j} = \begin{cases} 0 & \text{if } p_j(x, t) < b_{x,\tau} \\ \zeta_{x,\tau}^{t,j} & \text{if } p_j(x, t) \geq b_{x,\tau} \end{cases} \quad (3)$$

where  $\zeta_{x,\tau}^{t,j} = \zeta_{x,\tau,dtm}^{t,j} + \zeta_{x,\tau,rd}^{t,j}$  is the sum of a deterministic and a random component generated from a uniform distribution over  $[0, \omega]$ ,  $\omega < 1$ ,  $\tau \in T$ . The  $b_{x,\tau}$  are the model parameters representing diffusivity barriers between locations  $x$  and  $x + \tau$ ,  $b_{x,\tau} = b_{x+\tau,-\tau}$ , i.e. these implement simulated mountains, high  $b_{x,\tau}$  means a higher barrier value, which requires a larger population size to make crossing (diffusing through) likely, if the population size is not large enough to cross the barrier then  $r_{x,\tau}^{t,j} = 0$ . The deterministic component of the diffusivity parameter,  $\zeta_{x,\tau,dtm}^{t,j}$ , represents the impact of past population movements and the attraction of current populated areas, i.e. if the population migrated in a certain direction before, it is likely to follow the migration in the same direction and will not reverse quickly its migration direction, if there are areas with high population, indicating the good availability of resources, this will attract the migration of populations in the direction of such areas. In other

words, part of the  $j$ -th population moves from location  $x$  to  $x + \tau i$  if the population size is sufficient, i.e.  $p_j(x, t) \geq b_{x,\tau}$ .

Bit changes in  $\pi_j$  are cultural innovations, which imply the emergence of a new population with modified resource utilization ability. The new population starts at the location of the ancestor population and a part of this population converts to the new population. Populations may die out at any location. A population is successful if at any time across all locations it represents at least 0.5% of the total population of the simulated world. Further details about the model and simulation are provided in the materials and methods section in the supplementary information.

We considered the following features of successful populations: resource utilization efficiency, the resource utilization innovation, i.e. the difference between  $\pi_{population}$  and  $\pi_{ancestor}$ , and the time persistence, i.e. the number of time turns between the emergence and disappearance of the population. The impact of natural constraints was measured as the distance of the population's origin location to mountain ridges and the origin location's land fertility.

### 3. Implementation and simulation of the computational model

The computational model and simulation is based on a similar model and simulation described earlier (Andras, 2015). The simulation results reported here were generated with a modified version of this earlier model. The changes that we made implement more closely than the previous version the specific aspects of human migration, such as spreading of information about socio-economic centres and cognitive decision making about the direction of migratory movement. These factors imply that socio-economic centres exert attraction on migratory pathways and migratory movement directions are persistent for some period of time (i.e. there are no random reversals of the movement for a while). Technical details of the computational model and of the simulations are reported below.

The simulations were implemented in Delphi XE3 (Embarcadero Technologies). The locations of simulated mountain ridges and fertile land areas were set randomly. The work reported here is based on simulations of lands defined as  $100 \times 60$  square grids, with 20 mountain ridges and 10 locations of low fertility land areas with gradual change of land fertility. The data analysis is based on 60 distinct simulations. The size of the simulated land grid is set to be sufficiently large, while at the same time to allow sufficiently fast execution of the simulations. The numbers of mountain ridges and low land fertility land locations are set to again to be sufficiently large to generate a sufficiently complexly structured simulation environment. These numbers have been varied to assess their impact on the simulations and this sensitivity analysis is reported in Section 6.

Simulated mountain ridges are defined as continuous lines between two randomly selected grid locations in the  $100 \times 60$  grid location world. The height of the ridge is set randomly to be  $h$  by sampling a uniform distribution over the allowed value range. The range of the possible random values for  $h$  is set by the average difficulty of crossing mountain ridges  $\zeta$  – the actual range in the simulations is  $[0, 100 \cdot \zeta]$  and  $\zeta = 10$ , this value of  $\zeta$  is set to generate a sufficiently difficult environment, other values have been considered in the context of the sensitivity analysis discussed in Section 6. A mountain ridge  $M$  is implemented by setting the diffusivity values for the grid locations belonging to  $M$ , i.e.  $b_{x,\tau} = h$  for  $\forall x \in M, \tau \in T$  and  $b_{x+\tau,-\tau} = h$ . If mountain ridges cross each other the diffusivity values at the crossing grid locations are set to the highest  $h$  value associated with the crossing mountain ridges. We note that our setting of mountain ridges is the simplest way of

setting them randomly, it is not aimed to reproduce any closer approximation of the natural arrangement of mountain ridges. The simplest random setting of mountain ridges is considered sufficient for the purpose of the simulations presented here.

Simulated low fertility land areas are defined by picking randomly the centre grid location of the area, which has to be neighbouring or within a mountain ridge. At the centre the land fertility is set to be the lowest, and gradually the land fertility increases with the distance from the centre. If  $x$  is the centre grid location of a low fertility area, then the  $s$  size parameter of the area is set randomly, and the extent of low fertility is calculated for all grid locations  $L = \{x + \rho | (-s, -s) \leq \rho \leq (s, s)\}$  as  $\lambda(x + \rho) = 1/|\rho|$ . If low fertility land areas overlap the calculated extents of reduction in land fertility are added together resulting in a lower land fertility than in each of non-overlapping parts of these areas. The land fertility of a location  $x$  is  $1/\lambda(x)$ , if  $\lambda(x) \geq 0.1$  and it is 10 otherwise.

If the size of a certain population gets to zero or becomes negative at a spatial location, this population is considered extinct at this location and it gets removed from the respective location.

Each population has a set of cultural features, its socio-technical repertoire, that determine its characteristic resource utilisation ability  $\pi$ . The set of features is represented as a binary vector with 100 bits,  $g_1, \dots, g_{100}$ , where if the value of the bit  $g_i$  is 1, it means that  $i$ -th cultural feature is present in the socio-technical repertoire of the population, while if  $g_i = 0$  then this feature is not part of the socio-technical repertoire of the population. Then  $\pi$ , the resource utilisation ability of the population, is defined as  $\pi = \sum_{i=1}^{10} \sum_{k=1}^{10} g_{k+(i-1) \cdot 10} \cdot 2^k$ . Here our aim is to model in a simple way the combination of cultural features that determine resource utilisation practices. For example, these features may relate to animal husbandry (e.g. the use of animal milk as a frequent food source), hunting techniques (e.g. the use of a certain kind of arrow and bow), religious practices (e.g. using or avoiding human sacrifice), and so on. Naturally some cultural features are more impactful than others in terms of resource utilisation efficiency and some cultural features may be different, but equally impactful. Our model assumption here is that there are a set of groups of features such that any two feature in the same group have a different impact and that for each impact level there is one corresponding feature in each group of cultural features. To have a sufficient number of groups of features and a sufficient number of features in each group, we chose 10 both as the number of groups and number of features. This is an arbitrary choice, however, it provides a sufficiently large overall number of cultural features and groups of such features. The simplest way to achieve the difference of impact between features within the same group is that each cultural feature has an associated impact that is a power of a common base number, and the simplest choice for the base number is 2. This choice means that each cultural feature will exist fully or not at all in the resource utilisation repertoire of a population. Alternatively, we could have chosen a larger base, to allow several levels of development of each cultural feature (e.g. using animal milk as a drink; using animal milk to make a small range of milk products such as butter and soft cheese; using animal milk to make a wide range of milk products some of which can be stored for longer such as hard cheeses and fermented milk), however in this case the values of  $g_i$  would need to indicate the level of the cultural feature and would no longer be binary values (e.g. choosing the base 4 would allow three levels of expertise in the respective cultural feature, 1, 2, 3 the 0 indicating the lack of the cultural feature).

At the end of each time turn of the simulation each population at each grid location has the chance  $\eta = 0.00004$  to undergo a cultural innovation through a flip of one of the bits of the corresponding genome. If a such change happens the population splits as

described in the paper and a new population with a new socio-technical repertoire originates at the grid location where the cultural change happened. The amount of resource utilisation innovation due to the cultural change is  $l = \pi_{new} - \pi_{ancestor}$ . Due to the way we defined  $\pi$  the values of resource utilisation ability innovations in the simulations are always a power of 2. Note that the change, or cultural innovation, may happen in both direction, i.e. by adding a cultural feature to the socio-technical repertoire or by dropping a cultural feature from the pre-existing repertoire. The value of  $\eta$  is set to be sufficiently small, such that cultural innovations do not happen too often.

The functions  $f_{x,t}$  that determine the effective resource utilisation efficiency of the populations in a given location and at a given time are defined as follows. Let us assume that there are  $n$  populations at the grid location  $x$  at time  $t$ . First we calculate for the  $j$ -th population  $\theta_{x,t}^j = 1/(1 + e^{\beta \cdot (0.5 - \pi_j / (\sum_{i=1}^n \pi_i))})$ , where  $\beta$  is the model parameter that determines the steepness of competition between the co-located populations – we set  $\beta = 12$ , other values are explored in the context of sensitivity analysis discussed in Section 6. Then we calculate  $f_{x,t}(\pi_j) = \theta_{x,t}^j / \lambda(x) - \mu$ , if  $\lambda(x) \geq 0.1$  and  $f_{x,t}(\pi_j) = 10 \cdot \theta_{x,t}^j - \mu$ , if  $\lambda(x) < 0.1$ , where  $\lambda(x)$  is the extent of low fertility land at the grid location  $x$  (i.e. higher value of  $\lambda(x)$  implies lower land fertility), and  $\mu$  is the mortality rate that is assumed to be the same for all populations, in the simulation we used  $\mu = 0.02$ . The mortality rate is set to be sufficiently small and also sufficiently large to allow both growth and shrinking of populations in the context of the simulations and depending on the local competition between the co-located populations.

To calculate the random and deterministic components of the non-zero diffusivity parameters, i.e.,  $\xi_{x,t,\tau}^t = \xi_{x,t,\tau,dtm}^t + \xi_{x,t,\tau,rd}^t$ , we proceed as follows. Let  $J_{x,t}$  be the set of indices of all populations present at the location  $x$  at time  $t$  and let  $P(x,t) = \sum_{i \in J_{x,t}} p_i(x,t)$  be the total population at this location and time. For each direction  $\tau \in T = \{(-1, 0), (1, 0), (0, -1), (0, 1)\}$  we calculate the attractiveness of the direction  $\alpha(\tau)$  by looking at the total population in each grid location within the corresponding quadrant of the world  $Q_{x,\tau}$  (see Fig. 3) within a set radius (equal to 5 in the implementation). In effect we find out which of the four possible directions has higher population locations, which attract the movement of the population from its current location. Let  $P_{x,t,\tau}^{max} = \max_{x' \in Q_{x,\tau}} P(x', t)$  the maximum value of the total populations in these quadrants. Let  $\bar{P}_{x,t,\tau}^{max} = P_{x,t,\tau}^{max} / (\sum_{\tau' \in T} P_{x,t,\tau'}^{max} + P(x,t))$ , then we calculate  $\alpha(x,t,\tau) = 1/(1 + e^{-k_1 \cdot (\bar{P}_{x,t,\tau}^{max} - k_2)})$ , which is the attractiveness of the direction  $\tau$ , and  $k_1, k_2$  are the steepness and inflection point parameters – the values for these were set to 100 and 0.3, respectively, in the simulations in order to allow the clear differentiation between attractive and non-attractive locations. This means that given the sigmoidal expression of  $\alpha(x,t,\tau)$  the likely diffusion of the population is larger in the direction where there is a high population location in that direction within the neighbourhood of the current location, compared to another direction where the location with the largest total population has a smaller total population. However,  $\alpha(x,t,\tau)$  is set to 0 if the current population migrated to its current location from this direction recently – the recent migration is defined in this simulation as migration that happened in the last 5 time turns. This alteration of  $\alpha(x,t,\tau)$  implements the implicit maintenance of the direction of migration, by preventing the migration back to the recent earlier location. The deterministic component of the diffusivity parameter is then calculated as  $\xi_{x,t,dtm}^t = z \cdot \chi_{x,t,\tau} \cdot \alpha(x,t,\tau)$  and the random component is calculated as  $\xi_{x,t,rd}^t = (1 - z) \cdot \chi_{x,t,\tau} \cdot \varsigma$  where  $\varsigma$  is a random number from  $[0, 1]$  with uniform distribution,  $\chi_{x,t,\tau}$  is set to 0.4 as default value and is

altered to 0.6 if  $P_{x,t,\tau}^{max} > 1.5 \cdot P(x,t)$  and to 0.016 if  $P_{x,t,\tau}^{max} < P(x,t)$ , reflecting that more populated areas are more attractive as migration destination than less populated areas, and  $z$  is the parameter that balances the contribution of the deterministic and random components of the diffusivity parameter, which is set to 0.9 in the simulations (i.e. the deterministic component has a stronger influence than the random component). The values of these parameters are set on the basis of numerical simulation experiments such that the impact of the attractiveness of more populated locations does influence the movement of populations, but does not dominate it fully. Fig. 4

In addition to the above calculations, the diffusivity parameter is set to 0 if the movement of the population in a certain direction would mean the moving backward of the population to their earlier location (i.e. where they came from) if there are not yet sufficient number of time turns passed since this earlier move of the population. In the simulations we require 5 time turns to be passed before the population can move in the backward direction. All key parameters involved in the description of the simulations are listed in Table 11 with a brief explanation of the meaning of the parameters.

All simulations started with a random set of original populations that occupied in average 1.5% of all available grid locations. Initially there is at most one population in each grid location. The initial occupancy level is set such that initially the simulations have a sufficient, but not very large number of populations (around 90 populations over the simulated land area of  $100 \times 60$  grid locations), which allows an initial growth and spreading of these populations until their areas grow to the point where they start interfering with each other. This allows sufficient variety among the initial populations and also a variable density and area arrangement of the populations by the time they get to the point where they interfere with other populations. Each simulation was run for 30,000 time turns. The maximal number of populations at one grid location was set to be 8. If the number of residing populations would go above this limit at a certain grid location, due to inward migration, the population with the lowest resource utilisation efficiency was made extinct at this grid location. This limit on the number of different populations in one grid location is set to avoid competition between too many populations and it is approximately in line with the upper limit of the number of distinct larger populations co-existing in ancient towns and tribal alliances.

One may question the relevance of these simulations for human populations. While these are based in general on the numerical simulation of constrained diffusion – reaction systems, the particular nature of the simulated system and its constraints make these relevant for the simulation of spreading and competition biological populations. Furthermore, the setting of the rules and constraints of the simulations make these particularly relevant for modelling of dynamics of human populations (e.g. the attraction impact of highly populated areas on setting the direction of migration, the limit on the number of co-existing populations in locations). In general, the same simulation framework could be adapted to the simulation of the spreading and dynamics of other kinds of biological populations, but these would require different rules and constraints (e.g. bacteria would be driven by presence and diffusion of attractive and repellent molecules and tens – hundreds of bacterial species may co-exist at a given location).

#### 4. Main results: Areas of likely origin of successful populations

Using the simulations we found that the level of innovation and the time persistence of successful populations negatively correlate with the origin location's distance from mountain ridges, and that the resource utilization efficiency positively correlates with the

origin's distance from mountain ridges (Fig. 1A–C). These imply that successful populations originating closer to mountain ridges exist for longer and they make more impactful innovations. This supports our assumptions about the positive impact of closeness to natural barriers on the ability of the populations to persist and develop significant cultural innovations.

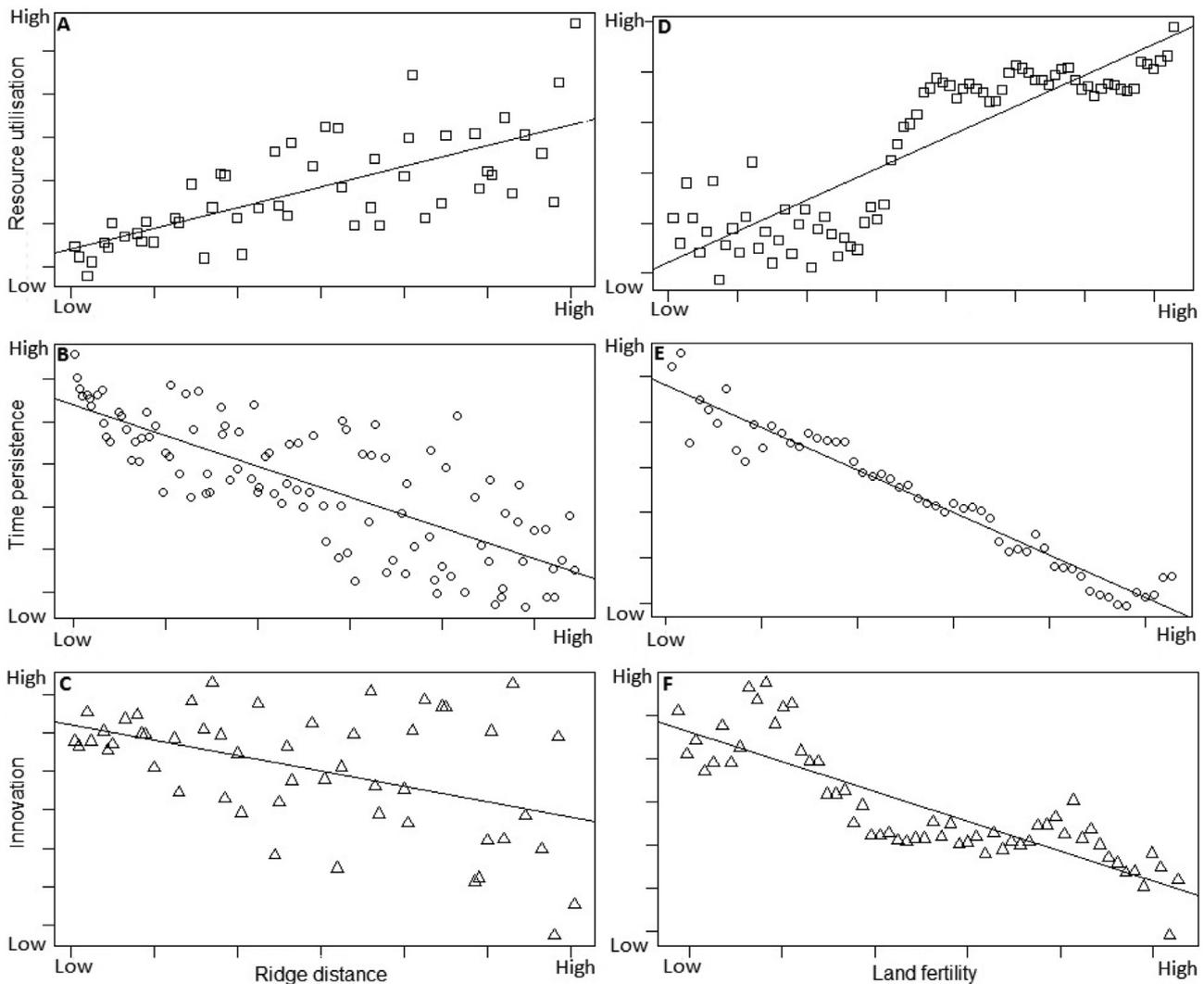
Our results show that resource utilization efficiency correlates positively, while the time persistence and level of innovation correlate negatively with the origin's land fertility for successful populations (Fig. 1D–F). These imply that successful populations originating from harsher lands (low land fertility) last longer and innovate more, while those that originate from fertile lands are more efficient in using resources. These support our assumptions about the effect of land fertility on the innovation, persistence and resource utilization ability of populations.

Our simulations confirm that successful populations are more likely to originate from areas with isolated valleys and low land fertility, where they develop high-impact cultural innovations, followed by rapid expansion and emergence to dominance of the same populations or their descendants from neighbouring fertile lands. For further validation we considered the high-resolution altitude and land-fertility data for Earth (Fischer et al., 2012;

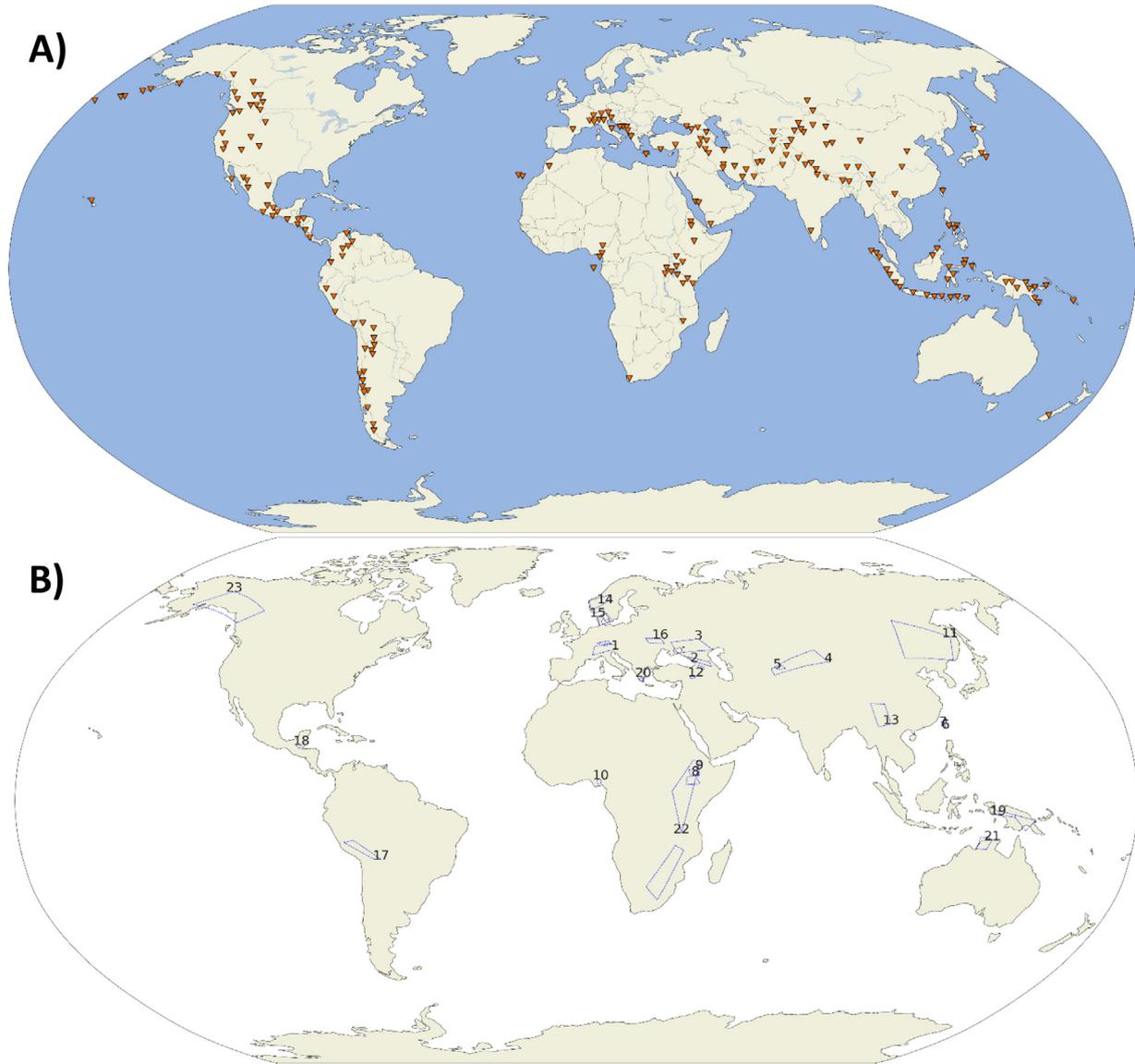
NOAA, 2016) and determined those areas which fit the above description. We used the local Laplacian transform of the altitude data with expansion of identified valley areas and the overlap of these with high-fertility land areas – see details of the analysis of the topographic data in the next section. The identified areas are expected to match areas that are assumed to be the urheimat for populations that became dominant around the world across history (e.g. Indo-Europeans (Piazza et al., 1995), Bantu (Campbell and Tishkoff, 2010)). The identified areas are shown in Fig. 2A, together with assumptions about urheimats of historical populations shown in Fig. 2B. The match between the predicted and actual locations is very good. To assess the statistical validity of the predicted areas as true successful population origins, we considered random locations on fertile lands and calculated the match of these with predicted areas. The comparison results (Table 1) show that our predictions match the accepted urheimat areas significantly better than random locations.

### 5. Analysis of the world map and assumed areas of origin

We used accurate topographical mapping of the surface of Earth to identify rugged, mountainous regions. We then identified where



**Fig. 1.** The relationship between natural constraints (mountain ridge distance, land fertility) and features of successful populations (resource utilization efficiency, time persistence, innovation). The panels show the best linear approximation of these relationships as well. The corresponding correlations and related p-values are as follows: A)  $c = 0.7154$ ,  $p = 3.65 \times 10^{-9}$ ; B)  $c = -0.7769$ ,  $p = 5.06 \times 10^{-22}$ ; C)  $c = -0.4591$ ,  $p = 7 \times 10^{-4}$ ; D)  $c = 0.8774$ ,  $p = 1.21 \times 10^{-25}$ ; E)  $c = -0.9696$ ,  $p = 1.04 \times 10^{-34}$ ; F)  $c = -0.848$ ,  $p = 4.48 \times 10^{-17}$ .



**Fig. 2.** Geographical locations. A) The locations on the map of the world that were identified as possible origins of successful populations in accordance with our results based on the simulations. B) The assumed areas of origin (urheimat) for a set of populations that became dominant in parts of the world across human history: 1 – Celts, 2 – Early Indo-Europeans, 3 – Indo-Europeans, 4 – Turkic/Mongols (highlands), 5 – Turkic/Mongols, 6 – Austronesians (highlands), 7 – Austronesians, 8 – Semitic/Kushitic (highlands), 9 – Semitic/Kushitic, 10 – Bantu, 11 – Korean/Japanese/Manchu, 12 – Anatolian Farmers, 13 – Tibeto-Burmese/Tai-Kadai, 14 – Germanics, 15 – Germanics (lowlands), 16 – Slavs, 17 – Inca, 18 – Maya, 19 – Papua, 20 – Greeks, 21 – Australian Aborigines, 22 – Hominids, 23 – Early American Natives (for further details see the text and Table 2).

these regions border fertile lowland areas. Where these areas are within sufficiently close distance of each other, we marked them as potential origin points for successful populations.

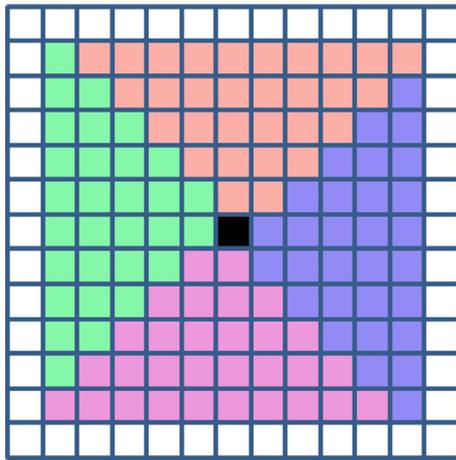
For the ground-truth terrain data, we used the ETOPO1 Global Relief Model (NOAA, 2016). This is a 1 arc-minute global relief model build from global and regional data sets. For the land fertility data, we used the Harmonized World Soil Database v1.2 (Fischer et al., 2012). A version of this database, pre-processed to indicate soil quality is available at 5 arcminute resolution from the same source. Land parcels are categorized according to characteristics such as organic carbon content, pH, water capacity, salinity, and nutrients, putting each on a scale from 1 (high quality) to 4 (low quality) – we note that the ratings 5 – 7 in the database identify non productive areas (i.e. non-soil, permafrost, water). We up-sampled this data to match the resolution of the ETOPO1 terrain dataset.

To find rugged regions, we first applied the 2D Laplacian operator with the convolution matrix

$$L = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

to the relief map to highlight peaks and valleys in the landscape. We then considered the absolute magnitude of each element, normalized across the whole map and applied the threshold  $t = 0.3$  element-wise across the array of the map. The threshold was set such that we achieved a good match with our subjective estimation of expected valley areas (i.e. where there are mountain ranges on the map, we expect to find valley areas).

Next we applied a greyscale dilation of size  $l = 40$  pixels (each pixel is equivalent to 1 arc-minute, or approximately 2 km in both



**Fig. 3.** The quadrant of the grids world considered for the calculation of attractiveness of population movement directions. The black grid location in the middle is the current location of the population. The colored sections indicate the relevant quadrants for the corresponding directions.

directions) to the output. The greyscale dilation is effectively a maximum filter over a sliding window, which spreads the identified areas beyond the mountainous regions towards lowland areas, up to around 80 km from their focal point – i.e. each pixel value is revised to be equal to the maximal original pixel value within the distance of 40 pixels.

We then found the intersection of these regions with high-quality soil regions (where the soil quality index is 1 for all pixels in a contiguous area). These intersecting areas were then subjected to another round of dilation of size  $r = 40$  pixels to form larger contiguous regions. Finally, the contiguous regions were enumerated, using a von Neumann neighborhood (i.e. up, down, left and right

neighbors, ignoring the corner neighbors) structural parameter to specify connectedness.

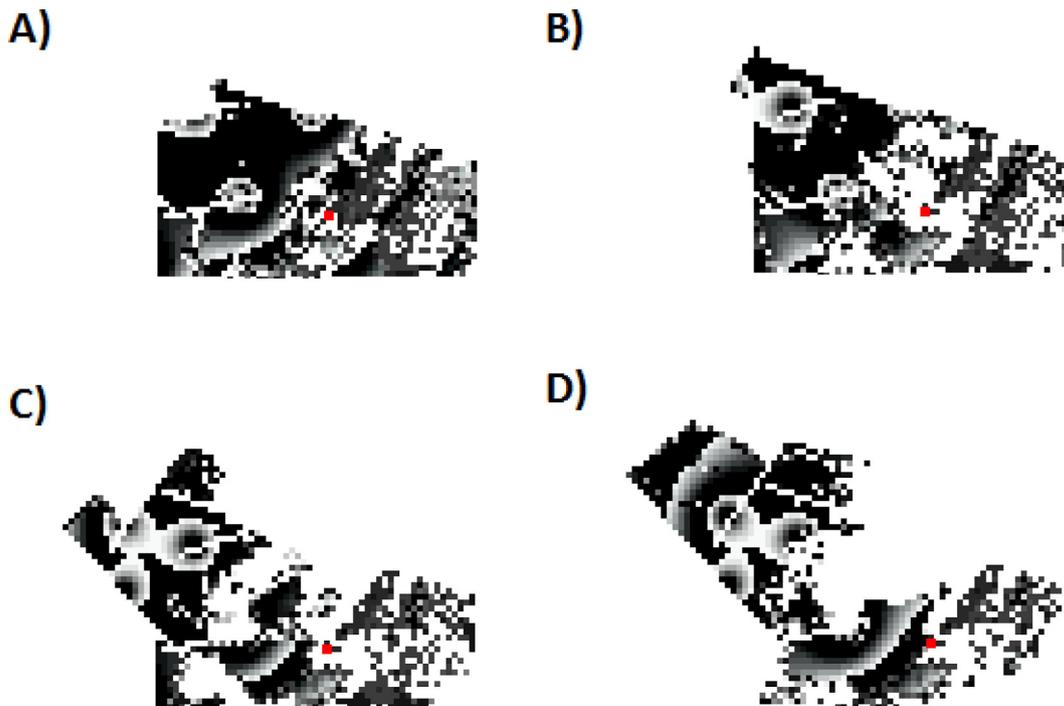
The centres of mass (COMs, i.e. the average of the location coordinates in both planar directions of the pixels belonging to the region) of these larger contiguous regions were identified as potential origin points for successful populations (see the triangles in Fig. 2A).

We identified the assumed areas of origin for a set of know historical human populations using descriptive and map-based information available in the literature. These identified origin or urheimat areas are listed in Table 2, including the relevant references.

We compared the identified areas of possible origin of successful populations with a random set of locations selected from fertile land areas (soil quality index = 1) to see whether the former are statistically significantly closer than the latter to assumed areas of origin of historically known successful populations. We used the two tailed, different standard deviations,  $t$ -test to compare the mean distance from each of the 204 identified locations to the nearest assumed area of origin, to the mean distance of 10,000 locations drawn randomly from the fertile land areas to the nearest assumed area of origin. If a location is inside of an assumed area of origin then the distance is considered to be zero. The results of the comparison are reported in Table 1 and these show that the two location distributions are significantly different (the  $p$ -value is well below 0.01), implying that the determined locations are statistically significantly closer to the assumed true origin locations of successful historical populations than randomly picked locations.

### 6. Sensitivity analysis of the computational model

In order to further check the validity of our results, we explored the parameter sensitivity of our computational model. For this purpose we used as key variables the number of mountain ridges, the



**Fig. 4.** Exemplary snapshots of the evolution of the spreading of populations in the simulations. Panels A) – D) show the evolution of the spreading of a population over a 200 time turn period, with 50 time turns between the snapshots. The snapshots show both expansion of the population in the upper-left direction and retraction of the population from the lower-right direction. The spreading shows Turing-like patterns. The darker is the grey-scale color the larger is the population count in the respective grid location. The red spot indicates the origin grid location of this population.

**Table 1**

Analysis of geographic locations. Comparison of the match of geographic locations identified in accordance with our results based on simulations and random geographic locations in fertile areas with the assumed areas of origin (urheimat) of known successful populations.

	Number of locations	Mean distance from assumed areas of origin (miles)	Standard deviation of distances from assumed areas of origin (miles)	Two-tailed, un-equal standard deviations, <i>t</i> -test p-value for the comparison of the two mean values
Identified locations	204	8.7457	8.5157	0.0006414
Random locations	10,000	10.858	8.7521	0.0006414

**Table 2**

Geographic location of urheimat areas of a set of historically known successful populations. The geographical locations are provided as quadrilaterals defined as quartets of (latitude, longitude) value pairs, and in some cases as a union of two such quadrilaterals.

Number	Population	Urheimat area	Geographical location
1	Celts (Cunliffe, 2018)	Alps and Bavaria	{{(44.64,5.75); (46.53,14.08); (47.40,6.66); (48.02,14.33)}} + {{(48.45,7.54); (49.07,11.71); (48.20,12.73); (48.05,8.76)}}
2	Early Indo-Europeans (Mathieson et al., 2018)	Southern Caucasus	{{(44.15,39.58); (42.63,41.59); (41.06,47.29); (42.50,46.91)}} + {{(41.42,42.67); (41.18,44.39); (40.11,44.48); (40.63,42.36)}}
3	Indo-Europeans (Mathieson et al., 2018; Piazza et al., 1995)	Dniester – Volga area north of the Caucasus	(48.68,34.56); 49.63,45.06; (46.27,49.32); (45.22,36.25)
4	Turkic/Mongols (highlands) (De Barros Damgaard et al., 2018)	Tien-Shan and neighbouring mountain areas	(46.29,85.48); (42.53,88.78); (38.38,68.40); (40.33,67.86)
5	Turkic/Mongols (De Barros Damgaard et al., 2018)	Fergana Valley area	(40.16,69.89); (40.74,70.20); (41.42,71.97); (40.45,72.95)
6	Austronesians (highlands) (Gray et al., 2009)	Eastern Taiwan	(22.75,120.44); (22.04,120.89); (24.86,121.20); (24.98,121.90)
7	Austronesians (Gray et al., 2009)	Western Taiwan	(23.04,120.03); (22.93,120.31); (24.39,120.69); (24.15,120.17)
8	Semitic/Kushitic (highlands) (Ehret, 1979)	South-West Ethiopian highlands	(7.15,35.65); (7.67,38.25); (5.08,39.75); (4.83,35.28)
9	Semitic/Kushitic (Ehret, 1979)	South West Ethiopia	(9.42,36.16); (9.60,39.64); (7.54,39.44); (7.70,36.69)
10	Bantu (Campbell and Tishkoff, 2010)	Boundary area of current Cameroon and Nigeria	(6.40,6.79); (6.51,7.84); (5.08,8.01); (4.98,6.79)
11	Korean/Japanese/Manchu (Hammer et al., 2006)	Northern Manchuria	(55.63,121.51); (50.43,138.53); (43.06,133.35); (43.70,116.72)
12	Anatolian Farmers (Diamond and Bellwood, 2003; Mathieson et al., 2018)	Southern Central Anatolia	(37.94,39.01); (38.16,41.28); (37.76,40.82); (37.33,39.66)
13	Tibeto-Burmese/Tai-Kadai (Driem, 1993)	Yunnan	(22.54,99.04); (23.42,103.40); (29.59,102.79); (29.72,98.20)
14	Germanics (Harke and Todd, 1994)	Western Scandinavia	(58.94,5.92); (61.39,11.78); (64.19,13.81); (62.24,5.26)
15	Germanics (lowlands) (Harke and Todd, 1994)	Denmark and Northern Germany	(53.99,8.89); (57.04,8.43); (57.93,11.93); (55.42,14.14)
16	Slavs (Mielnik-Sikorska et al., 2013)	Western Ukraine and South-West Belarus	(49.95,25.65); (49.92,31.02); (48.47,26.51); (48.25,32.18)
17	Quechua/Inca (Diamond and Bellwood, 2003)	South-East Peru	(-12.87,-75.12); (-12.24,-71.93); (-18.15,-65.57); (-18.07,-63.51)
18	Maya (Diamond and Bellwood, 2003)	Western Guatemala	(17.19,-90.99); (17.03,-89.24); (15.88,-89.86); (16.41,-90.56)
19	Papua (Bergström et al., 2017)	Central and Eastern New Guinea	{{(-3.21,135.83); (-4.36,135.39); (-5.23,140.88); (-4.49,140.91)}} + {{(-5.23,140.88); (-4.49,140.91); (-9.29,144.65); (-6.25,147.56)}}
21	Australian Aborigines (Bouckaert et al., 2018)	North West part of Northern Territories of Australia	(-11.22,130.40); (-12.07,134.19); (-14.97,132.87); (-14.80,129.75)
22	Hominids (Hammer et al., 2006)	Sothorn Ethiopia – Kenya – Western Tanzania and Central South Africa	{{(12.00,39.84); (-10.00,33.85); (2.84,30.74); (12.27,37.72)}} + {{(-15.27,34.80); (-30.49,27.06); (-26.49,23.09); (-13.77,32.24)}}
23	Early American Natives (Moreno-Mayar et al., 2018)	Eastern Beringia (Southern Alaska and Western Yukon)	(60.95, -155.26); (65.37,-144.89); (58.84,-123.79); (54.93, -130.30)

number of low fertility land areas, the average difficulty of crossing mountain ridges ( $\zeta$ ), and the steepness of competition between co-located populations ( $\beta$ ). To measure the model sensitivity we considered the variation of the key relationships between ridge distance and land fertility of the origin location and innovation, time persistence and resource utilization efficiency of populations, respectively.

The sensitivity of the computational model was assessed using 5 different values for each parameter, and 8 – 10 simulations for each parameter value (all other considered parameters were held constant at the default value). The values considered for the parameters were as follows: 5, 10, 20, 30, 60 mountain ridges; 2, 5, 10, 20, 40 low fertility land areas; 2, 10, 20, 50, 100 for the mountain ridge crossing difficulty parameter  $\zeta$ ; and 8, 12, 24, 70, 150 for the competition steepness parameter  $\beta$ . The sensitivity analysis data is shown in Tables 3–10.

We found that the number of mountain ridges has very little impact on the nature of the key relationships. The ridge distance

and resource utilization relationship reverses for high numbers of mountain ridges. In other cases the relationships maintain the same nature as reported in Table 1 and the statistical significance of the relationships increases with the number of mountain ridges. These results are summarized in Supplementary Tables 3 and 4.

For small number of low fertility areas the correlations with the origin's land fertility get reversed for innovation and also in one case for the resource utilization efficiency and in one case for time persistence. For higher numbers of low land fertility areas the relationships with the origin's land fertility are the same as the ones reported in Table 1. In terms of ridge distance relationships we found the reversal for the relationships for innovation for mid range values of the number of low fertility areas and also in one case for the resource utilization efficiency relationship. In all other cases the relationships reported in Table 1 are maintained. The results for the variation of the number of low land fertility areas are shown in Supplementary Tables 5 and 6.

**Table 3**

The key relationships for variable number of mountain ridges. Correlation values followed by corresponding p-values in brackets. Numbers in bold italic are not in agreement with those reported in Table 1 for the corresponding relationship.

Mountain ridges Relationship with origin's distance from mountain ridges	5	10	30	60
Resource utilisation	0.448 (6.8E-5)	0.4755 (2.76E-12)	<b>-0.54 (6.58E-13)</b>	<b>-0.54 (5.86E-13)</b>
Time persistence	-0.35 (1.4E-12)	-0.6189 (7E-27)	-0.677 (5.33E-18)	-0.7016 (8E-20)
Innovation	-0.15 (4.32E-2)	-0.1738 (3.65E-7)	-0.3287 (1.53E-5)	-0.3595 (1.82E-6)

**Table 4**

The key relationships for variable number of mountain ridges. Correlation values followed by corresponding p-values in brackets.

Mountain ridges Relationship with origin's land fertility	5	10	30	60
Resource utilisation	0.78 (1.53E-16)	0.7978 (2.9E-17)	0.8979 (2.45E-32)	0.9427 (3.86E-34)
Time persistence	-0.9 (8.76E-18)	-0.658 (1.82E-12)	-0.809 (3.35E-12)	-0.839 (2.25E-18)
Innovation	-0.346 (2.8E-3)	-0.6525 (2.86E-8)	-0.882 (1.06E-14)	-0.918 (1.65E-18)

**Table 5**

The key relationships for variable number of low land fertility areas. Correlation values followed by corresponding p-values in brackets. Numbers in bold italic are not in agreement with those reported in Table 1 for the corresponding relationship.

Low land fertility areas Relationship with origin's distance from mountain ridges	2	5	20	40
Resource utilisation	0.324 (2.16E-2)	0.4281 (2.15E-3)	<b>-0.3473 (2.79E-5)</b>	0.5228 (4.33E-14)
Time persistence	-0.55 (3.41E-3)	-0.6322 (1.2E-3)	-0.485 (1.02E-15)	-0.607 (2.96E-25)
Innovation	-0.44 (9.11E-3)	<b>0.2057 (1.27E-2)</b>	<b>0.6033 (2.95E-3)</b>	-0.427 (1.53E-11)

**Table 6**

The key relationships for variable number of low land fertility areas. Correlation values followed by corresponding p-values in brackets. Numbers in bold italic are not in agreement with those reported in Table 1 for the corresponding relationship.

Low land fertility areas Relationship with origin's land fertility	2	5	20	40
Resource utilisation	0.858 (3.51E-3)	<b>-0.707 (7.62E-10)</b>	0.9652 (1.62E-42)	0.9652 (1.62E-42)
Time persistence	<b>0.746 (1E-4)</b>	-0.7748 (1.3E-6)	-0.9107 (8.5E-41)	-0.9107 (4.9E-39)
Innovation	<b>0.4873 (1.9E-3)</b>	<b>0.5336 (2.7E-4)</b>	-0.7495 (1E-14)	-0.7295 (1.1E-15)

**Table 7**

The key relationships for variable ridge crossing difficulty. Correlation values followed by corresponding p-values in brackets. Numbers in bold italic are not in agreement with those reported in Table 1 for the corresponding relationship.

Ridge crossing difficulty Relationship with origin's distance from mountain ridges	2	20	50	100
Resource utilisation	<b>-0.325 (5.3E-5)</b>	0.608 (4.58E-12)	0.5227 (1.26E-9)	<b>-0.4366 (6.44E-5)</b>
Time persistence	-0.66 (8.5E-19)	-0.381 (4.8E-4)	-0.7 (8.3E-4)	-0.7551 (5.9E-7)
Innovation	<b>0.264 (6.08E-2)</b>	-0.3815 (5.43E-5)	-0.3968 (2.23E-4)	-0.4306 (2.22E-2)

**Table 8**

The key relationships for variable ridge crossing difficulty. Correlation values followed by corresponding p-values in brackets.

Ridge crossing difficulty Relationship with origin's land fertility	2	20	50	100
Resource utilisation	0.665 (2.75E-9)	0.7245 (6.31E-13)	0.8881 (5.4E-25)	0.8253 (1.54E-17)
Time persistence	-0.94 (2.9E-20)	-0.805 (3.75E-10)	-0.921 (9.86E-20)	-0.936 (1.95E-18)
Innovation	-0.39 (1.47E-3)	-0.4235 (3.55E-4)	-0.6715 (1.08E-8)	-0.6767 (1.76E-7)

**Table 9**

The key relationships for variable level of competition. Correlation values followed by corresponding p-values in brackets. Numbers in bold italic are not in agreement with those reported in Table 1 for the corresponding relationship.

Level of competition Relationship with origin's distance from mountain ridges	8	24	70	150
Resource utilisation	0.386 (5.27E-7)	0.3025 (2.05E-4)	0.4026 (1.69E-6)	0.3128 (4.26E-4)
Time persistence	<b>0.587 (4.04E-7)</b>	-0.2776 (5.8E-5)	-0.3202 (2.27E-7)	-0.5196 (5.78E-8)
Innovation	<b>0.291 (6.09E-3)</b>	<b>0.31 (1.36E-3)</b>	-0.1864 (6.59E-3)	-0.2052 (2.06E-2)

**Table 10**

The key relationships for variable level of competition. Correlation values followed by corresponding p-values in brackets. Numbers in bold italic are not in agreement with those reported in Table 1 for the corresponding relationship.

Level of competition	8	24	70	150
Relationship with origin's land fertility				
Resource utilisation	0.315 (1.77E-3)	0.4458 (4.21E-6)	0.4866 (3.75E-7)	−0.68 (1.26E-14)
Time persistence	<b>0.901 (1.5E-13)</b>	<b>0.8303 (6.81E-10)</b>	<b>0.6265 (4.35E-5)</b>	−0.665 (6.50E-13)
Innovation	<b>0.313 (1.32E-2)</b>	<b>0.3748 (2.68E-3)</b>	−0.7494 (3.43E-4)	<b>0.4398 (9.17E-6)</b>

**Table 11**

Summary of the key symbols and notations used in the mathematical notations and equations throughout the paper, together with the meaning and relevance of these.

Symbol	Meaning
$x$	Location
$t$	Time
$p(x, t)$	Population at location $x$ and time $t$
$D(p, x)$	Diffusivity of the environment, which depends on the population size and location
$F(p, x, t)$	Reaction term of the reaction – diffusion equation, representing the population change depending on the population size, location and time
$p_j(x, t)$	The $j$ -th population at location $x$ and time $t$
$\pi_j$	Characteristic resource utilization ability of the $j$ -th population
$\tau$	Direction of diffusion, which can take the values $(-1,0)$ , $(1,0)$ , $(0,-1)$ , $(0,1)$ , which correspond to the directions left, right, down and up.
$r_{x,t}^j$	Stochastic diffusivity parameter for the $j$ -th population at location $x$ and time $t$ in the direction $\tau$
$f_{x,t}$	The effective resource utilization efficiency functions for a given location and time, considering land fertility and the simultaneous presence of other populations
$b_{x,\tau}$	Barrier value for location $x$ and direction $\tau$ , this characterises diffusion barriers, which implement mountain ranges in the simulations
$\zeta_{x,\tau}^t$	Value of the diffusivity parameter $r_{x,t}^j$ if $p_j(x, t) \geq b_{x,\tau}$
$\zeta_{x,t,dtm}^t$	Deterministic component of $\zeta_{x,\tau}^t$
$\zeta_{x,t,rdm}^t$	Random component of $\zeta_{x,\tau}^t$
$M$	Mountain ridge, comprising of locations $x$ belonging to the mountain ridge
$\zeta$	Mountain ridge maximal height parameter, the maximal height of mountain ridges is $100 \cdot \zeta$
$\lambda(x)$	Level of low land fertility at location $x$
$g_1, \dots, g_{100}$	Bits of the population's socio-technical repertoire that determine the resource utilisation ability $\pi$
$\eta$	Innovation likelihood parameter
$i$	Amount of resource utilisation innovation, following an innovation event
$\theta_{x,t}^j$	Relative resource utilisation efficiency for the $j$ -th population at location $x$ and time $t$
$\beta$	Parameter characterising the competition between populations
$\mu$	Mortality rate
$P(x, t)$	Total population at location $x$ and time $t$
$J_{x,t}$	Set of indices of all population at location $x$ and time $t$
$Q_{x,\tau}$	Neighbourhood quadrant in the direction $\tau$ for the location $x$
$p_{x,t,\tau}^{\max}$	Maximum value of total population in one location within the quadrant $Q_{x,\tau}$ at time $t$
$\tilde{p}_{x,t,\tau}^{\max}$	Normalized maximum value of total population in one location within the quadrant $Q_{x,\tau}$ at time $t$
$\alpha(x, t, \tau)$	Attractiveness of moving population in the direction $\tau$ from the location $x$ at time $t$
$k_1, k_2$	Parameters determining the migration direction preference.
$z$	Parameter determining the balance between the deterministic and random components the diffusivity parameters
$\varsigma$	Random parameter determining the value of the random component of the diffusivity parameters
$\chi_{x,t,\tau}$	Parameter determining the value of the diffusivity parameters, reflecting the attractiveness for migration of the relevant movement direction $\tau$ for location $x$ at time $t$
$L$	2D Laplacian operator convolution matrix

The key relationships relative to the distance of origin from mountain ridges are reversed for resource utilization efficiency for very low and very high crossing difficulty and for innovation

for very low crossing difficulty. In all other cases the relationships reported in Table 1 are maintained and strengthened for higher crossing difficulty. In terms of the relationships relative to the origin's land fertility, the correlations reported in Table 1 are maintained and increase in strength with increasing crossing difficulty. These results are shown in Supplementary Tables 7 and 8. We note that crossing difficulty in the real world may be influenced by climate and technology (e.g. harsh climate may increase crossing difficulty, while technological advances may reduce crossing difficulty).

For low level of competition several of the key relationships with time persistence and innovation get reversed for both the origin's land fertility and its distance from mountain ridges. The resource utilization relationships are maintained as in Table 1, with the exception of the case of very high competition and the origin's distance from mountain ridges. For high competition values the innovation and time persistence relationships with the origin's distance from mountain ridges are the same as the ones reported in Table 1. Some of the key relationships for the origin's land fertility are reversed even for higher competition values. The competition steepness sensitivity analysis results are shown in Supplementary Tables 9 and 10. We note that in the real world technology may increase the steepness of competition, making the impact of access to fertile land areas on the success of populations much larger.

Our sensitivity analysis shows that in most cases the parameters do not influence very much these key relationships, although in some cases the relationships may get weakened or even reversed. This means that the proposed computational model is generally robust and the results that we reported are valid across a range of parameter settings. However, our analysis also shows that some parameters in certain cases weaken to some extent the validity of the general results, pointing to the importance of these parameters both in terms of setting of the computational model and also in terms of actual environmental and socio-technical settings that are valid in that environment.

While we have not investigated the impact of all considered parameters, as we explained in Section 3, in general we aimed to set these parameter values to be in appropriate ranges in order to generate sufficiently complex simulations that were also feasible to be run sufficiently many times. In our view the above considered parameters are the key parameters that have the potential to have the biggest impact on the simulations.

## 7. Conclusions

The work presented here uses computational simulations and predictive analysis of geographical data to support the hypothesis that successful populations of humans emerge in areas with many valleys and relatively low land fertility, which are sufficiently close to highly fertile lands. Our work predicts a number of geographical areas which are currently not associated with origins of successful human populations, but can be considered as candidate urheimats for known successful populations without identified origin location (Fig. 2).

The model and analysis presented here has implications for other fields of science as well. In particular, in the context of the process of speciation and spreading of biological species our results

imply that areas with high 'endemism' (i.e. relatively many endemic species), which are sufficiently separated by natural barriers, but provide sufficient life supporting resources, are the most likely locations of speciation events. This has been recently evidenced by analysis of experimental data (Rabosky et al., 2018). In the context of emergence of fundamentally novel technologies and technical innovations, our work suggests that these may happen most often in environments that are sufficiently protective over relatively long term (e.g. in terms of patenting protection or in terms of required in-depth technological know-how that is not easily accessible – equivalent of natural barriers), require and support sufficient blue-sky thinking (i.e. addressing fundamental questions of the technology – equivalent of drive for cultural innovations) and are also sufficiently close to places with appropriately trained workforce and venture capital funding ready to invest in new technologies (equivalent of high fertility areas).

## 8. Data and code availability

The source code of the software used for the simulations and the simulations data are available on request from the corresponding author.

The map and soil quality data that support the findings of this study are available from: ETOPO1 Global Relief Model, <https://www.ngdc.noaa.gov/mgg/global/global.html>, doi:<https://doi.org/10.7289/V5C8276M> and Harmonized World Soil Database v.1.2, <http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/en/>. The software code for the analysis of these data is available at <https://github.com/StanDeSiccle/HumanOrigins>. The code for simulations and the simulations data used for the paper are available on request from the authors.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgment

We thank Catherine Merrick (University of Cambridge) for very useful comments on the draft version of the paper.

## References

Andras, P. Environmental Factors and the Emergence of Cultural – Technical Innovations. 2015. 10.7551/978-0-262-33027-5-ch028.

Bergström, A., Oppenheimer, S.J., Mentzer, A.J., Auckland, K., Robson, K., Attenborough, R., Alpers, M.P., Koki, G., Pomat, W., Siba, P., Xue, Y., Sandhu, M.S., Tyler-Smith, C., 2017. A Neolithic expansion, but strong genetic structure, in the independent history of New Guinea. *Science* 80. <https://doi.org/10.1126/science.aan3842>.

Bouckaert, R., Lemey, P., Dunn, M., Greenhill, S.J., Alekseyenko, A.V., Drummond, A.J., Gray, R.D., Suchard, M.A., Atkinson, Q.D., 2012. Mapping the origins and expansion of the Indo-European language family. *Science* 80. <https://doi.org/10.1126/science.1219669>.

R.R. Bouckaert, C. Bower, Q.D. Atkinson. The origin and expansion of Pama-Nyungan languages across Australia. *Ecol. Evol Nat.* 2018. 10.1038/s41559-018-0489-3.

M.C. Campbell S.A. Tishkoff The Evolution of Human Genetic and Phenotypic Variation in Africa 2010 *Biol Curr* 10.1016/j.cub.2009.11.050.

L.L. Cavalli-Sforza Genes, peoples, and languages 1997 *Natl. Acad. Sci. U. S. A Proc* 10.1073/pnas.94.15.7719.

Chiaroni, J., King, R.J., Underhill, P.A., 2008. Correlation of annual precipitation with human Y-chromosome diversity and the emergence of Neolithic agricultural and pastoral economies in the Fertile Crescent. *Antiquity*. <https://doi.org/10.1017/S0003598X00096800>.

Coop, G., Pickrell, J.K., Novembre, J., Kudaravalli, S., Li, J., Absher, D., Myers, R.M., Cavalli-Sforza, L.L., Feldman, M.W., Pritchard, J.K., 2009. The role of geography in human adaptation. *PLoS Genet.* <https://doi.org/10.1371/journal.pgen.1000500>.

Cunliffe, B., 2018. *The Ancient Celts*. Oxford University Press, Oxford.

De Barros Damgaard, P., Marchi, N., Rasmussen, S., Peyrot, M., Renaud, G., Korneliussen, T., Moreno-Mayar, J.V., Pedersen, M.W., Goldberg, A., Usmanova, E., Baimukhanov, N., Loman, V., Hedeager, L., Pedersen, A.G., Nielsen, K., Afanasiev, G., Akmatov, K., Aldashev, A., Alpaslan, A., Baimbetov, G., Bazaliiskii, V.I., Beisenov, A., Boldbaatar, B., Boldgiv, B., Dorzhu, C., Ellingvag, S., Erdenebaatar, D., Dajani, R., Dmitriev, E., Evdokimov, V., Frei, K.M., Gromov, A., Goryachev, A., Hakonarson, H., Hegay, T., Khachatryan, Z., Khakhanov, R., Kitov, E., Kolbina, A., Kubatbek, T., Kukushkin, A., Kukushkin, I., Lau, N., Margaryan, A., Merkyte, I., Mertz, I.V., Mertz, V.K., Mijidдорж, E., Moiyesev, V., Mukhtarova, G., Nurmukhanbetov, B., Orozbekova, Z., Panyushkina, I., Pieta, K., Smrčka, V., Shevnina, I., Logvin, A., Sjögren, K.G., Štolcová, T., Tashbaeva, K., Tkachev, A., Tulegenov, T., Voyakin, D., Yepiskoposyan, L., Undrakhbold, S., Varfolomeev, V., Weber, A., Kradin, N., Allentoft, M.E., Orlando, L., Nielsen, R., Sikora, M., Heyer, E., Kristiansen, K., Willerslev, E., 2018. 137 ancient human genomes from across the Eurasian steppes. *Nature*. <https://doi.org/10.1038/s41586-018-0094-2>.

D.M. Debinski R.D. Holt. A survey and overview of habitat fragmentation experiments. 2000. *Biol Conserv.* 10.1046/j.1523-1739.2000.98081.x

Diamond, J., Bellwood, P., 2003. Farmers and their languages: The first expansions. *Science* 80. <https://doi.org/10.1126/science.1078208>.

G.V. Driem Language change, conjugational morphology and the sino-tibetan urheimat 1993 *Hafniensia Acta Linguist* 10.1080/03740463.1993.10415452.

Ehret, C., 1979. On the antiquity of agriculture in Ethiopia. *J. Afr. Hist.* <https://doi.org/10.1017/S002185370001700X>.

S. Fedotov D. Moss D. Campos Stochastic model for population migration and the growth of human settlements during the Neolithic transition 2008 *Rev. E – Stat. Nonlinear, Soft Matter Phys Phys* 10.1103/PhysRevE.78.026107.

Fischer, G., Nachtergaele, F., Prieler, S., van Velthuizen, H.T., Verelst, L., Wiberg, D., 2012. *Global Agro-ecological Zones-Model Documentation (GAEZ v. 3.0). Food and Agriculture Organization of the United Nations*.

Flather, C.H., Bevers, M., 2002. Patchy reaction-diffusion and population abundance: The relative importance of habitat amount and arrangement. *Am. Nat.* <https://doi.org/10.1086/324120>.

García-Ramos, G., Rodríguez, D., 2002. Evolutionary speed of species invasions. *Evolution (N. Y.)*. [https://doi.org/10.1554/0014-3820\(2002\)056\[0661:esosij\]2.0.co;2](https://doi.org/10.1554/0014-3820(2002)056[0661:esosij]2.0.co;2).

Gray, R.D., Drummond, A.J., Greenhill, S.J., 2009. Language phylogenies reveal expansion pulses and pauses in pacific settlement. *Science* 80. <https://doi.org/10.1126/science.1166858>.

Hammer, M.F., Karafet, T.M., Park, H., Omoto, K., Harihara, S., Stoneking, M., Horai, S., 2006. Dual origins of the Japanese: Common ground for hunter-gatherer and farmer Y chromosomes. *J. Hum. Genet.* <https://doi.org/10.1007/s10038-005-0322-0>.

Harke, H., Todd, M., 1994. The Early Germans. *Britannia*. 10.2307/527028.

N. Isern J. Zilhão J. Fort A.J. Ammerman Modeling the role of voyaging in the coastal spread of the Early Neolithic in the West Mediterranean 2017 *Natl. Acad. Sci. U. S. A Proc* 10.1073/pnas.1613413114.

C. Jeong A.T. Ozga D.B. Witonsky H. Malmström H. Edlund C.A. Hofman R.W. Hagan M. Jakobsson C.M. Lewis M.S. Aldenderfer A.D. Rienzo C. Warinner Long-term genetic stability and a high-altitude East Asian origin for the peoples of the high valleys of the Himalayan arc 2016 *Natl. Acad. Sci. U. S. A Proc* 10.1073/pnas.1520844113.

A. Kitchen C. Ehret S. Assefa C.J. Mulligan Bayesian phylogenetic analysis of Semitic languages identifies an Early Bronze Age origin of Semitic in the Near East 2009 *R. Soc. B Biol. Sci Proc* 10.1098/rspb.2009.0408.

Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanov, K., Sudmant, P.H., Schraiber, J.G., Castellano, S., Lipson, M., Berger, B., Economou, C., Bollongino, R., Fu, Q., Bos, K.I., Nordenfelt, S., Li, H., De Filippo, C., Prüfer, K., Sawyer, S., Posth, C., Haak, W., Hallgren, F., Fornander, E., Rohland, N., Delsate, D., Francken, M., Guinet, J.M., Wahl, J., Ayodo, G., Babiker, H.A., Bailliet, G., Balanovska, E., Balanovsky, O., Barrantes, R., Bedoya, G., Ben-Ami, H., Bene, J., Berrada, F., Bravi, C.M., Brisighelli, F., Busby, G.B.J., Cali, F., Churnosov, M., Cole, D.E.C., Corach, D., Damba, L., Van Driem, G., Dryomov, S., Dugoujon, J.M., Fedorova, S.A., Gallego Romero, I., Gubina, M., Hammer, M., Henn, B.M., Hervig, T., Hodoglugil, U., Jha, A.R., Karachanak-Yankova, S., Khusainova, R., Khusnutdinova, E., Kittles, R., Kivisild, T., Klitz, W., Kučinskas, V., Kushniarevich, A., Laredj, L., Litvinov, S., Loukidis, T., Mahley, R.W., Melegh, B., Metspalu, E., Molina, J., Mountain, J., Näkkäläjärvi, K., Nesheva, D., Nyambo, T., Osipova, L., Parik, J., Platonov, F., Posukh, O., Romano, V., Rothhammer, F., Rudan, I., Ruizbakiev, R., Sahakyan, H., Sajantila, A., Salas, A., Starikovskaya, E.B., Tarekgn, A., Toncheva, D., Turdikulova, S., Uktveryte, I., Utevska, O., Vasquez, R., Villena, M., Voevoda, M., Winkler, C.A., Yepiskoposyan, L., Zalloua, P., Zemanek, T., Cooper, A., Capelli, C., Thomas, M.G., Ruiz-Linares, A., Tishkoff, S.A., Singh, L., Thangaraj, K., Villems, R., Comas, D., Sukernik, R., Metspalu, M., Meyer, M., Eichler, E.E., Burger, J., Slatkin, M., Pääbo, S., Kelso, J., Reich, D., Krause, J., 2014. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*. <https://doi.org/10.1038/nature13673>.

Mathieson, I., Alpaslan-Roodenberg, S., Posth, C., Szécsényi-Nagy, A., Rohland, N., Mallick, S., Olalde, I., Broomandkoshbacht, N., Candilio, F., Cheronet, O., Fernandes, D., Ferry, M., Gamarra, B., Fortes, G.G., Haak, W., Harney, E., Jones, E., Keating, D., Krause-Kyora, B., Kucukkalipci, I., Michel, M., Mittnik, A., Nägele, K., Novak, M., Oppenheimer, J., Patterson, N., Pfrengle, S., Sirak, K., Stewardson, K., Vai, S., Alexandrov, S., Alt, K.W., Andrescu, R., Antonović, D., Ash, A., Atanassova, N., Bacvarov, K., Gusztáv, M.B., Bocherens, H., Bolus, M., Boroneanț, A., Boyadzhiev, Y., Budnik, A., Burmaz, J., Chohadzhiev, S., Conard, N.J., Cottiaux, R., Cúka, M., Cupillard, C., Drucker, D.G., Elenski, N., Francken, M.,

- Galabova, Ganetsovski, G., Cély, B., Hajdu, T., Handzhyiska, V., Harvati, K., Higham, T., Iliev, S., Janković, I., Karavanić, I., Kennett, D.J., Komšo, D., Kozak, A., Labuda, D., Lari, M., Lazar, C., Leppek, M., Leshtakov, K., Vetro, D.L., Los, D., Lozanov, I., Malina, M., Martini, F., McSweeney, K., Meller, H., Mentušić, M., Mirea, P., Moiseyev, V., Petrova, V., Douglas Price, T., Simalcsik, A., Sineo, L., Šlaus, M., Slavchev, V., Stanev, P., Starović, A., Szeniczey, T., Talamo, S., Teschler-Nicola, M., Thevenet, C., Valchev, I., Valentin, F., Vasilyev, S., Veljanovska, F., Venelinova, S., Veselovskaya, E., Viola, B., Virag, C., Zaninović, J., Zäuner, S., Stockhammer, P.W., Catalano, G., Krauß, R., Caramelli, D., Zariņa, G., Gaydarska, B., Lillie, M., Nikitin, A.G., Potekhina, I., Papathanasiou, A., Borić, D., Bonsall, C., Krause, J., Pinhasi, R., Reich, D., 2016. The genomic history of southeastern Europe. *Nature*. <https://doi.org/10.1038/nature25778>.
- Mellars, P., 2006. Why did modern human populations disperse from Africa ca. 60,000 years ago? A new model. *Proc. Natl. Acad. Sci. U. S. A.* 10.1073/pnas.0510792103.
- Mielnik-Sikorska, M., Daca, P., Malyarchuk, B., Derenko, M., Skonieczna, K., Perkova, M., Dobosz, T., Grzybowski, T., 2013. The History of Slavs Inferred from Complete Mitochondrial Genome Sequences. *PLoS One*. <https://doi.org/10.1371/journal.pone.0054360>.
- Á. Montenegro R.T., Callaghan S.M., Fitzpatrick Using seafaring simulations and shortest-hop trajectories to model the prehistoric colonization of Remote Oceania 2016 *Natl. Acad. Sci. U. S. A Proc* 10.1073/pnas.1612426113.
- Moreno-Mayar, J.V., Potter, B.A., Vinner, L., Steinrücken, M., Rasmussen, S., Terhorst, J., Kamm, J.A., Albrechtsen, A., Malaspina, A.S., Sikora, M., Reuther, J.D., Irish, J., D., Malhi, R.S., Orlando, L., Song, Y.S., Nielsen, R., Meltzer, D.J., Willerslev, E., 2018. Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature*. <https://doi.org/10.1038/nature25173>.
- Nielsen, R., Akey, J.M., Jakobsson, M., Pritchard, J.K., Tishkoff, S., Willerslev, E., 2017. Tracing the peopling of the world through genomics. *Nature*. <https://doi.org/10.1038/nature21347>.
- NOAA, 2016. ETOPO1 Global Relief Model. *Natl. Ocean. Atmos. Adm.*
- Owens, K., King, M.C., 1999. Genomic views of human history. *Science* 80. <https://doi.org/10.1126/science.286.5439.451>.
- A. Piazza S. Rendine E. Minch P. Menozzi J. Mountain L.L. Cavalli-Sforza Genetics and the origin of European languages 1995 *Natl. Acad. Sci. U. S. A Proc* 10.1073/pnas.92.13.5836.
- Rabosky, D.L., Chang, J., Title, P.O., Cowman, P.F., Sallan, L., Friedman, M., Kaschner, K., Garilao, C., Near, T.J., Coll, M., Alfaro, M.E., 2018. An inverse latitudinal gradient in speciation rate for marine fishes. *Nature*. <https://doi.org/10.1038/s41586-018-0273-1>.
- Ryberg, W.A., Hill, M.T., Painter, C.W., Fitzgerald, L.A., 2013. Landscape Pattern Determines Neighborhood Size and Structure within a Lizard Population. *PLoS One*. <https://doi.org/10.1371/journal.pone.0056856>.
- M. Slatkin F. Racimo Ancient DNA and human history 2016 *Natl. Acad. Sci. U. S. A Proc* 10.1073/pnas.1524306113.
- V. Volpert S. Petrovskii Reaction-diffusion waves in biology 2009 *Life Rev Phys* 10.1016/j.plrev.2009.10.002.