

# Hindcasting global population densities reveals forces enabling the origin of agriculture

Patrick H. Kavanagh<sup>1,5\*</sup>, Bruno Vilela<sup>2,5</sup>, Hannah J. Haynie<sup>1</sup>, Ty Tuff<sup>2</sup>, Matheus Lima-Ribeiro<sup>3</sup>, Russell D. Gray<sup>4</sup>, Carlos A. Botero<sup>1,2</sup> and Michael C. Gavin<sup>1</sup>

**The development and spread of agriculture changed fundamental characteristics of human societies<sup>1-3</sup>. However, the degree to which environmental and social conditions enabled the origins of agriculture remains contested<sup>4-6</sup>. We test three hypothesized links between the environment, population density and the origins of plant and animal domestication, a prerequisite for agriculture: (1) domestication arose as environmental conditions improved and population densities increased<sup>7</sup> (surplus hypothesis); (2) populations needed domestication to overcome deteriorating environmental conditions (necessity hypothesis)<sup>8,9</sup>; (3) factors promoting domestication were distinct in each location<sup>10</sup> (regional uniqueness hypothesis). We overcome previous data limitations with a statistical model, in which environmental, geographic and cultural variables capture 77% of the variation in population density among 220 foraging societies worldwide. We use this model to hindcast potential population densities across the globe from 21,000 to 4,000 years before present. Despite the timing of domestication varying by thousands of years, we show that improving environmental conditions favoured higher local population densities during periods when domestication arose in every known agricultural origin centre. Our results uncover a common, global factor that facilitated one of humanity's most significant innovations and demonstrate that modelling ancestral demographic changes can illuminate major events deep in human history.**

The origins and spread of agriculture have been linked to important changes in the trajectory of human history<sup>2</sup>. However, longstanding debates persist regarding the degree to which different social and environmental conditions influenced the timing and locations of agricultural origins<sup>6,11</sup>. Agriculture began with a critical innovation: the domestication of plants and animals for food production<sup>11,12</sup>. Specifically, the pathway to agriculture started with low-level food production, including the cultivation of wild-type species. This cultivation continued for a number of generations, and in some cases thousands of years, before natural and artificial selection resulted in domesticated species used for food production<sup>6,11,12</sup>.

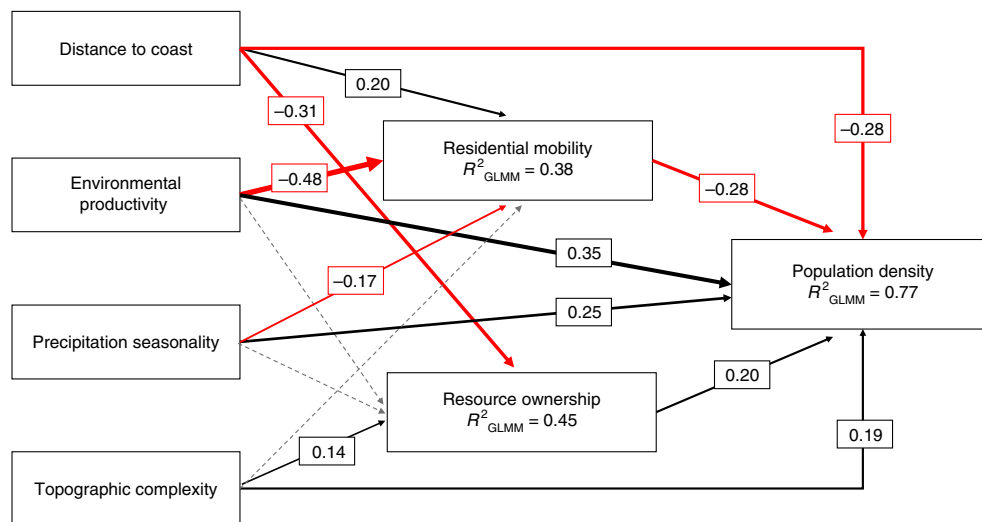
Three widely cited sets of hypotheses argue for different relationships between the innovation of domestication and changing environmental and demographic conditions in the late Pleistocene and early Holocene. One set of hypotheses, which we refer to as the surplus hypothesis, associates the innovation of domestication with improving environmental conditions, increased resource availability and growing human population densities. For example,

some scholars suggest that improving environmental conditions during the early Holocene resulted in a greater abundance of natural resources, which allowed the population densities of foraging societies to increase<sup>7</sup>. Others note that under such favourable conditions, feasting and economically based competition may have promoted a search for new ways to acquire food resources and produce further surpluses, resulting in plant and animal domestication<sup>5,6</sup>. Improving conditions and increasing population densities may also have facilitated the innovation of domestication because the likelihood of innovations, including domestication, scales positively with the number of potential innovators<sup>13,14</sup>. Recent research also suggests that rising temperatures and increases in atmospheric carbon during the early Holocene may have enhanced the productivity of wild progenitors of important domesticates<sup>15,16</sup>. The resulting increase in the availability of cultivated resources may have allowed population densities to increase as well as making dependence on cultivated food more viable<sup>15,16</sup>.

Others argue for what we term the necessity hypothesis, in which deteriorating environmental conditions may have led to the development of plant and animal domestication so as to secure better food supplies<sup>8,9,17,18</sup>. For example, climatic fluctuations during the Younger Dryas (a period of cold and dry conditions) may have led to an increased reliance on cultivated plants by societies in southwest Asia<sup>8,9</sup>. Finally, supporters of what we refer to as the regional uniqueness hypothesis suggest that distinct, local processes independently drove the different geographic origins of domestication<sup>10</sup>. These researchers argue that differences in the processes leading to domestication among origin centres are a result of local-scale phenomena and variable responses to macroscale factors<sup>10</sup>. Therefore, under the regional uniqueness hypothesis, we should not expect a general global pattern linking environment and population density with processes leading to the innovation of domestication.

A lack of reliable environmental and population density data has been a major constraint on previous tests of hypothesized drivers of the origins of domestication. Instead, previous studies have relied on narrative arguments or empirical tests with relatively small and geographically restricted samples<sup>6,17,19</sup>. Palaeoclimate simulations at a global scale and at fine spatial resolutions have only become available in recent years<sup>20,21</sup>. Additionally, previous estimates of historic population densities have been limited to coarse spatial scales, such as continents or countries, or do not cover time spans appropriate for investigating the origins of domestication (for example, do not span the late Pleistocene to middle Holocene)<sup>22,23</sup>. We overcome these constraints by identifying predictors of population density

<sup>1</sup>Department of Human Dimensions of Natural Resources, Colorado State University, Fort Collins, CO, USA. <sup>2</sup>Department of Biology, Washington University in St. Louis, St. Louis, MO, USA. <sup>3</sup>Universidade Federal de Goiás - Regional Jataí, Instituto de Biociências, São Paulo, Brazil. <sup>4</sup>Department of Linguistic and Cultural Evolution, Max Planck Institute for The Science of Human History, Jena, Germany. <sup>5</sup>These authors jointly supervised this work: Patrick H. Kavanagh, Bruno Vilela. \*e-mail: [patrick.kavanagh@colostate.edu](mailto:patrick.kavanagh@colostate.edu), [h.kavanagh@gmail.com](mailto:h.kavanagh@gmail.com)



**Fig. 1 | Path diagram for piecewise-SEM exploring the effects of environmental and cultural variables on population densities of foraging societies.**

Measured variables are represented by the large boxes and  $R^2_{GLMM}$  values (see Methods) are provided for response variables.  $n = 220$ . Red arrows depict negative relationships among variables, black arrows positive relationships, and dashed grey arrows depict non-significant paths ( $P \geq 0.05$ ). Standardized coefficients are presented for all paths (small boxes) and arrow widths are scaled to reflect the magnitude of path coefficients.

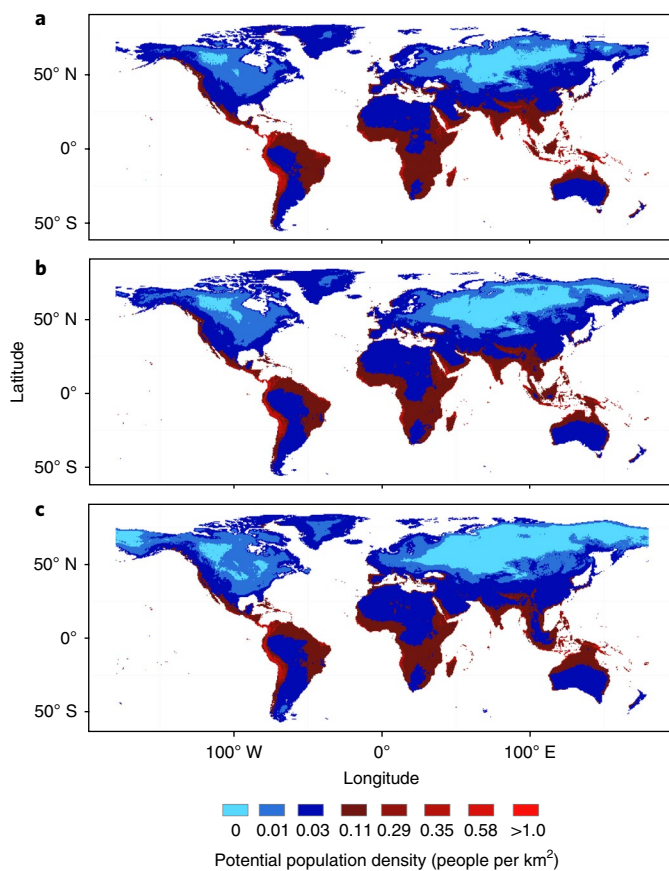
among 220 foraging societies of the recent past (Supplementary Fig. 1 and Supplementary Table 1) using piecewise structural equation modelling. Then, using this model of recent forager population densities, along with newly available climate reconstructions<sup>21</sup>, we hindcast population densities across the globe from the late Pleistocene (21,000 YBP) to the middle Holocene (4,000 YBP), a 17,000-year time span (see paragraphs below and Methods for more details). This time span encompasses the range of known domestication origin dates along with the amelioration of environmental conditions after the Last Glacial Maximum (LGM) that are suggested to have been necessary before agriculture was possible.

To improve understanding of the drivers of population density in foraging societies, we used piecewise structural equation modelling (piecewise-SEM)<sup>24</sup> to examine the correlation between the environment, cultural traits and population density (see Methods for more details). This approach allows both direct and indirect effects of predictors to be assessed, and builds on previous investigations, which have been limited by data and methodological constraints to smaller spatial scales, relatively small samples or univariate analyses<sup>19,25,26</sup>. For each society, we recorded the population density (people per km<sup>2</sup>) as well as two cultural traits hypothesized to scale with population density: residential mobility (that is, average distance travelled per residential move) and level of resource ownership (that is, whether resources are owned or not)<sup>19</sup>. We also examined the potential effects of distance to the coast, as access to marine resources may allow for increased population density<sup>27</sup>. Language family was included as a random effect in our SEM model to control for the non-independence of societies that share a common cultural background. We conducted principal components analysis (PCA) to avoid the potentially misleading effects of multicollinearity among environmental variables<sup>28</sup>. The PCA was used to summarize environmental variables into three axes, representing environmental productivity (PC1), topographic complexity (PC2) and precipitation seasonality (PC3) (see Methods and Supplementary Table 2).

The suite of environmental and cultural variables explored here explains the majority of variation in population density among foraging societies ( $R^2_{GLMM} = 0.77$ ; all variables have significant direct effects on population density; Fig. 1). Environmental productivity has the strongest effect on population density, consistent with earlier hypotheses that suggest greater environmental productivity

will increase resource availability per unit area, thereby permitting higher population densities<sup>25,29,30</sup> (but see ref. <sup>27</sup>). Similarly, the effects of precipitation seasonality, distance to coast and topographic complexity on population density may be due to their association with available resources. For example, species richness scales positively with habitat heterogeneity, topographic complexity and precipitation seasonality<sup>31–33</sup>. Furthermore, coastal populations of foragers often reach high densities due to the abundant, protein-rich, marine resources<sup>27</sup>. Environmental conditions also influence both cultural variables included in our model—residential mobility and resource ownership. However, the direct effects of environmental and cultural variables on population density are greater than the indirect environmental effects mediated by cultural traits (for example, direct effect of distance to coast on population density:  $\beta = -0.28$ ; indirect effect via residential mobility:  $0.2 \times -0.28 = -0.056$ ; see Fig. 1 and Methods). We acknowledge that multiple characteristics of foraging societies of the recent past may reflect contact with non-foraging societies<sup>26</sup>, but we note that our results remain unchanged when we exclude foraging societies known to have prior contact with non-foraging groups (see Methods). Specifically, we re-ran our piecewise-SEM excluding all societies known to have subsistence-based contact with non-foraging groups, as this may influence patterns of population density. The consistency among the results suggests that including societies that had contact with non-foraging groups did not influence the population density model (Supplementary Fig. 2).

We hindcasted population densities of foraging societies by fitting our population density model to simulations of palaeoclimate conditions and values for slope and elevation (adjusted for historical changes in sea level; see Methods for more details). Specifically, we estimated potential population densities of foraging societies (that is, the most likely population density estimated for a particular set of environmental conditions) across the globe at  $0.5 \times 0.5^\circ$  resolution at 1,000 year intervals, spanning the period of 21,000–4,000 YBP (see Methods for more details). Although residential mobility and resource ownership were important predictors in our model of recent population density, data for these variables do not exist for most societies in prehistory. However, recent genetic analyses suggest that the social structure (including levels of relatedness, residence patterns and individual mobility) of more contemporary foraging groups is similar to that of foraging groups from the



**Fig. 2 | Predictions of potential population density for foragers.**

**a–c.** Predicted population densities at 4,000 (**a**), 10,000 (**b**) and 21,000 (**c**) YBP. Blue hues depict potential population densities below the median population density of observed foraging societies, and red hues depict potential population densities above the median. The second red hue and above are greater than the mean population density of observed foraging societies. Note the increase in area, through time, with potential population densities greater than the mean of observed foraging societies (number of  $0.5^\circ \times 0.5^\circ$  cells: 21,000 YBP = 3,027; 4,000 YBP = 4,673). For example, a notable increase in the number of red cells in the Sudanic savannah and Ganges of East India (Northeast India) between panels **c** and **a**.

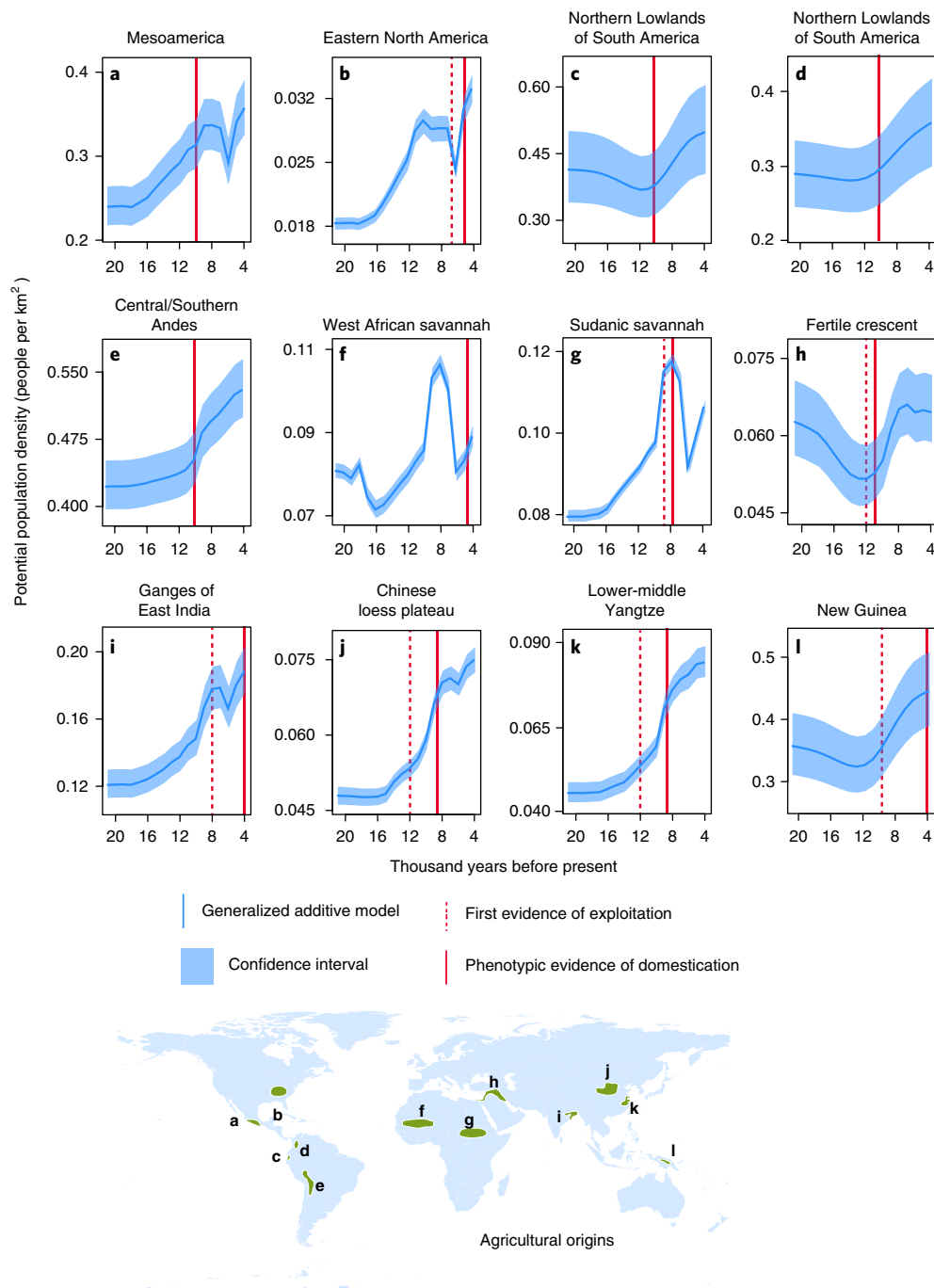
Upper Paleolithic ( $\sim 34$  YBP)<sup>34</sup>. To account for the uncertainty in residential mobility and resource ownership, we randomly assigned values for these traits (sampling stochastically from their corresponding natural range of values) to each map cell and then calculated a predicted population density value based on our model. This process was iterated 10,000 times and we then calculated the average predicted value for each  $0.5^\circ \times 0.5^\circ$  cell (Fig. 2 and Supplementary Fig. 3). We then used generalized additive models (GAMs) to investigate patterns of variation in potential population density over time at each agricultural centre of origin<sup>11</sup> (see Methods for more details). While domestication and agriculture are not one and the same<sup>12</sup>, the innovation of domestication is a key prerequisite for agriculture<sup>11</sup>. Furthermore, the pathway to agriculture may begin with an extended period (more than 2,000 years in the case of many cereal species) of cultivation before phenotypic evidence of domestication is observable<sup>11</sup>. Therefore, to provide a conservative estimate we defined the emergence of domestication as occurring between the first known evidence of exploitation of a subsequently domesticated species and phenotypic evidence indicative of domestication<sup>11</sup> (see Methods).

Despite variation in the emergence of domestication among the origin centres (11,000–5,000 YBP), along with significant climatic fluctuations in many regions, nearly all of the origin centres display a consistent positive slope of the potential population density index curve during the period when domestication arose (mean slope  $\pm$  standard error:  $6.85\text{E-}06 \pm 1.25\text{E-}06$ ; Fig. 3). We then tested whether this outcome was less likely than would be expected by chance. To do so, we randomly selected a focal year for each origin site, calculated the slope of the GAM curve at this point (see Methods for details), summed the slope values for all origin sites, and repeated this procedure 10,000 times. We then summed the slope values for the domestication origin times across the 12 origin sites and compared these to the distribution of slopes generated through the randomization procedure. The results of this test showed that it is very unlikely to obtain, at the same time and across all origin centres, high and positive observed slope values for the time at which domestication was innovated (as indicated by phenotypic evidence for domestication; Supplementary Fig. 4).

Only 2 of the 12 origin centres (eastern North America and Ganges of East India) display a brief decline before a sharp increase in potential population density during the period of domestication innovation. However, the exact timing of domestication events is not well resolved in these particular regions<sup>11</sup>, and both regions exhibit a net increase in potential population density during the domestication time window (Fig. 3). Furthermore, exploration of raw environmental variables confirms that the centres of agricultural origin did not experience unique climatic conditions during the studied time span when compared to other regions of the globe (Supplementary Fig. 5). These findings suggest that environmental conditions and patterns of population density are not unique among the domestication origin centres. While these results do not lend support to the regional uniqueness hypothesis, there are a range of other social and ecological conditions that may have created unique paths to domestication and agriculture in each centre of origin, and this deserves further attention.

Others suggest that human populations may have been driven to develop domestication due to deteriorating environmental conditions<sup>8,9,17,18</sup> (necessity hypothesis). Our results suggest that at the time of the first domestications, environmental conditions could support growing population densities in all centres of origin (that is, positive trend in potential population density curves; Fig. 3), a conclusion that is incompatible with the necessity hypothesis that links agriculture to environmental pressures. For example, it has been suggested that deteriorating environmental conditions may have contributed to the development of agriculture in southwest Asia<sup>8,9</sup>. However, our results suggest that during the time period in which domestication developed, not only in southwest Asia but also in all other known centres of origin, environmental conditions were improving and could support growing population densities (Fig. 3h).

In contrast, our results are consistent with the surplus hypothesis, which suggests that improving environmental conditions and the potential for increased population density may have facilitated the domestication of plants and animals in agricultural origin centres<sup>4,7</sup> (Fig. 3). Several factors may explain the links between environmental conditions, potential population density and the origin of domestication. For one, rates of innovation may scale positively with the number of potential innovators<sup>13,14</sup>. In turn, the likelihood of domestication innovations may have increased in environments that could support increasingly higher densities of foraging people. In addition, foraging societies may have become more sedentary to take advantage of locally abundant resources, some of which were later domesticated<sup>35</sup>. Our results indicate that residential mobility scales negatively with population density in foraging societies (Fig. 1). Therefore, increasingly sedentary lifestyles may have contributed further to increases in population density and the potential



**Fig. 3 | Potential population density trends and the emergence of domestication.** a–l, Index curves from GAMs (solid blue lines) for potential population density trends in the 12 independent centres of agricultural origins. Agricultural origins are estimated to have occurred between first evidence of exploitation (dashed red lines) and phenotypic evidence of domestication (solid red lines), or at phenotypic evidence of domestication (solid red line) where evidence of exploitation does not currently exist (see Methods for more details). Geographic locations of origin centres are represented by green polygons on the map and corresponding panel letters (a–l). Note, origin centres with broader confidence intervals (light blue bands) tend to encompass regions with greater amounts of environmental variation.

for innovation. Increases in the productivity of wild progenitors of important domesticates may have also facilitated growing population densities and the viability of cultivation for food production<sup>15,16</sup>.

It is also possible that improving environmental conditions may have resulted in a situation where necessity drove the origins of domestication. For example, population densities may have increased in foraging societies that occupied productive, coastal areas, causing an outflow of groups into regions with less ideal

conditions where the cultivation of plants and animals was required to secure adequate food resources<sup>6,17,18</sup>. Our results cannot support, or refute, the possible influence the outflow of people from hospitable locations to less ideal environments may have played. A detailed understanding of the movements of ancient populations is required for more rigorous testing of the role that forced habitation of marginal environments may have played in the origins of domestication at particular sites.



It has also been suggested that improving environmental conditions resulted in growing population densities that eventually exceeded environmental carrying capacities<sup>7</sup>. Given the absence of accurate data on ancient population sizes, we cannot conclusively rule out the possibility that population growth may have outpaced the amelioration of environmental conditions in some locations and created a scenario in which necessity drove the development of domestication. However, our results for the majority of agricultural origin centres do not display features that would indicate an increased likelihood of population growth exceeding carrying capacity. For example, plateaux or dips in potential population density curves may reflect an increased likelihood that populations may have temporarily surpassed carrying capacities during the period of interest due to environmental conditions becoming less favourable, or not being able to support growing population densities. The Ganges of East India does represent an interesting exception, where the potential population density decreases before the emergence of evidence for domestication (Fig. 3). However, the exact timing of domestication origins are uncertain at this location and there is a net increase in potential population density during the domestication time window (Fig. 3i). Regional palaeo-demographic work may eventually provide better evidence for evaluating whether populations continued to grow beyond environmental carrying capacities. Thus, we conclude that further research is needed to determine the precise links between improving environmental conditions, increasing population densities and domestication innovations in each origin centre. Nevertheless, our analysis uncovers patterns suggesting that the potential for increasing population density was an important enabling condition in domestication innovations across all origin sites.

The emergence of domestication, and subsequent innovation of agriculture, was probably linked to temporal and spatial variation in social and environmental conditions<sup>6,11</sup>. However, by investigating the drivers of population density in observed foraging societies we were able to estimate potential population densities in prehistory, providing a view of the deep past that has until now been obscured due to an absence of written records. This approach has allowed us to conduct a robust assessment of the role of demographic changes in the origins of agriculture and to conclude that the potential for increasing population density was an important enabling condition in all regions that developed an innovation that changed the course of human history. The approach we have demonstrated here lays a foundation for testing numerous other hypotheses that relate major developments in human history to changes in population density.

## Methods

Data on foraging societies (population density, resource ownership and residential mobility) were acquired from D-PLACE (Database of Places, Language, Culture and Environment<sup>19,36,37</sup>) and reflect data obtained from the Binford hunter-gatherer data set<sup>19,36</sup>. Residential mobility was quantified as the average distance per residential move (total distance travelled in a year/total number of residential moves) and resource ownership was coded as 'no ownership of resources' or 'resources are owned' (including ownership by local groups or elites). All climate data are based on the baseline historical (1900–1949) CCSM4.0 model obtained from the ecoClimate database<sup>21</sup>. The following climate variables were temporally interpolated from 21,000 YBP (LGM) through to 4,000 YBP, at 1,000 year intervals, using the globally stacked oxygen curve from ref. <sup>38</sup>: annual mean temperature, annual temperature variance, annual mean precipitation and annual precipitation coefficient of variation. Topographic variables (slope and elevation) are derived from the SRTM30\_plus Digital Elevation Model (DEM)<sup>39</sup>.

We adjusted coastlines to reflect global sea level predictions for each time period<sup>40</sup>. First, a raster was created from a polygon of contemporary coastlines with a cell size of  $0.5 \times 0.5^\circ$ . We then adjusted this raster utilizing bathymetry data contained in the SRTM30\_plus DEM and the predictions provided in a recent assessment of sea level and ice volumes from the LGM to the Holocene<sup>40</sup> (selecting the time period closest to our year of interest; maximum difference 64 years, mean difference = 24.6 years). To avoid issues of multicollinearity among environmental

variables, we ran PCA. First, data were extracted for all terrestrial raster cells at each time period, and variables were transformed to meet assumptions of normality (when needed), centred and scaled. The number of factors retained for analysis was informed by the Kaiser rule and parallel analysis<sup>41</sup>. The PCA produced three main composite variables: (1) 'environmental productivity' describes a gradient of increasing annual mean temperature, annual mean precipitation, and decreasing annual temperature variance; (2) 'topographic complexity' describes a gradient of increasing slope and elevation; (3) 'precipitation seasonality' describes a gradient of increasing annual precipitation variance (Supplementary Table 2). The structure of components was consistent across individual time periods and the pooled data set. Principal component scores were then assigned to associated raster cells in each time period. Distance to coast was also obtained for each raster cell. This was quantified as the distance (km) from the centre of each raster cell to the nearest marine cell.

Principal component scores and distance to coast values were extracted for the raster cell occupied by each of the 220 foraging societies (based on latitude–longitude points for each society from D-PLACE)<sup>36</sup>. To assess the relationship between population density of foraging societies and the various environmental and cultural predictors, we used a piecewise-SEM<sup>24</sup>. We included language family as a random effect in the structural equation model to account for the non-independence of societies that share a common cultural background<sup>42</sup>. The piecewise-SEM allowed both direct and indirect (that is, via the two cultural variables, residential mobility and resource ownership) environmental effects to be assessed. As our primary goal is to obtain the best possible prediction of population density, we assessed alternative path directions and selected the final model based on the predictive accuracy ( $R^2_{\text{GLMM}}$  value) for population density. The  $R^2_{\text{GLMM}}$  values are based on the variance of both the fixed and random effects<sup>43</sup>. We also repeated our piecewise-SEM analysis after excluding foraging societies that may have been influenced by relatively recent contact with non-foraging societies (see D-PLACE variable B005), and our results remained largely unchanged. The only significant change is that precipitation seasonality had a significant effect on resource ownership, whereas it did not in the original analysis. However, the direct effect of precipitation seasonality on population density was still greater than the indirect effect mediated by resource ownership (see final sentence of paragraph, below). Furthermore, the population density model was qualitatively identical to that in the original SEM. To assess whether environmental variables had greater direct or indirect (that is, via either of the two cultural variables) effects on population density, we first quantified indirect effects by multiplying the two path coefficients leading to population density. The magnitude of the indirect effect was then compared with the magnitude of the alternative direct path.

We then used the model fitted by the piecewise-SEM to predict potential population densities of all  $0.5 \times 0.5^\circ$  cells for historic time periods at 1,000 year intervals. These 1,000 year intervals are at a resolution appropriate for examining trends during the time period when agriculture arose due to current uncertainty in the timing of the development of agriculture at many origin centres<sup>11</sup>. Although humans did not occupy all regions of the globe throughout history (for example, the extent of ice sheets, surface water, other topographic features and climate may have put limits on the extent of human populations<sup>25</sup>), calculating predicted values for all cells provides an estimate of potential population density if foraging societies were occupying these areas. This increases the utility of the predictions for hypothesis testing, particularly in light of future archaeological evidence. First, we ran the mixed-effects model for population density (the population density model from the piecewise-SEM) and used this to derive predicted values. The Moran's I correlogram of model residuals by distance indicated that there is no evidence for unaccounted sources of spatial autocorrelation in our model (Supplementary Fig. 6). Precise residential mobility values and resource ownership practices of societies in prehistory are largely unknown. Therefore, residential mobility and resource ownership values were randomly assigned to each raster cell. Residential mobility was sampled from a truncated normal distribution of 100,000 values, with mean, standard deviation, minimum and maximum values set to those of the observed residential mobility distribution. Resource ownership was randomly assigned as either 0 (no ownership) or 1 (resources are owned). The random effect of language family was also omitted from predictions due to an absence of information on ancient language ranges. However, the random effect only accounted for a very small amount of the variance explained by our population density model (with random effect, conditional  $R^2_{\text{GLMM}} = 0.77$ ; without random effect, marginal  $R^2_{\text{GLMM}} = 0.74$ ). After residential mobility and resource ownership values were randomly assigned to cells, predicted population density values were calculated. This process was iterated 10,000 times, and average predicted population density values were calculated for each cell.

We used GAMs to track general trends in the potential population density of foragers between 21,000 YBP and 4,000 YBP at each  $0.5 \times 0.5^\circ$  cell<sup>44</sup>. The smooth term selected to generate the GAM for each cell was chosen using a jack-knife cross-validation procedure, also known as 'leave-one-out' (adapted from ref. <sup>45</sup>). A major advantage of this approach is that it reduces problems of overfitting while maintaining the ability to capture complex population trends. In the leave-one-out procedure, for each cell we ran  $n$  GAMs (where  $n$  is the number of estimated values for each cell;  $n = 18$ ) for each smoothing parameter value from 1 (which assumes a linear relationship between time and population density) to  $n$  (a complete

overfitting of the data). In each of the  $n$  GAMs generated per parameter value, one different observation was set aside to test the model. The model performance was assessed based on its overall capacity to predict the excluded observation. Following this, we selected the smoothing parameter that maximized the model performance in each cell.

The exact timing of domestication innovation in each origin centre is not currently known with a high degree of certainty. To provide the most conservative estimate for the timing of domestication innovation in each origin centre we utilized the supplementary material of ref. <sup>11</sup>. Specifically, we identified agricultural origins as occurring between the first evidence of exploitation of a subsequently domesticated species and first phenotypic evidence indicative of domestication. The time period between first evidence of exploitation and phenotypic evidence of domestication varied among origin centres, and in areas where evidence for first exploitation is currently not available, we use phenotypic evidence of domestication as a conservative estimate of the timing of agricultural origins. The spatial limits of the origin centres were approximated from the recent review in ref. <sup>11</sup>.

To test whether the slope of the GAM curves at the time of domestication origins were more or less likely than would be expected by chance, we randomly selected a focal year for each origin site, calculated the slope of the GAM curve at this point (be selecting the points 2,000, 3,000 and 5,000 years before this focal date; Supplementary Fig. 4), summed the slope values for the 12 origin sites, and repeated this procedure 10,000 times. We then valued the slope values for the domestication origin times across the 12 origin sites (following the same protocol of selecting points 2,000, 3,000 and 5,000 years earlier for slope calculation) and compared these to the distribution of slopes generated through the randomization procedure.

All analyses were conducted in the R environment for statistical computing<sup>46</sup>. The raster<sup>47</sup> and maptools<sup>48</sup> packages were used for the processing of GIS data, the piecewiseSEM<sup>24</sup> and lme4<sup>49</sup> packages were used for structural equation modelling and population density predictions, and the letsR<sup>50</sup> package was used to produce Moran's I correlogram.

**Reporting Summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

**Code availability.** All analyses were conducted using freely available packages in the R environment for statistical computing (see Methods) and no custom code was used. R scripts are available upon request.

**Data availability.** Data for population density and cultural traits of societies are available from D-PLACE (<https://d-place.org>) and climate data are available via the ecoClimate project (<http://ecoclimate.org/>). All other data are available upon request.

Received: 10 August 2017; Accepted: 3 May 2018;

Published online: 04 June 2018

## References

- Peoples, H. C. & Marlowe, F. W. Subsistence and the evolution of religion. *Hum. Nat.* **23**, 253–269 (2012).
- Diamond, J. & Bellwood, P. Farmers and their languages: the first expansions. *Science* **300**, 597–603 (2003).
- Currie, T. E. & Mace, R. Political complexity predicts the spread of ethnolinguistic groups. *Proc. Natl Acad. Sci. USA* **106**, 7339–7344 (2009).
- Richerson, P. J., Boyd, R. & Bettinger, R. L. Was agriculture impossible during the Pleistocene but mandatory during the Holocene? A climate change hypothesis. *Am. Antiq.* **66**, 387–411 (2001).
- Hayden, B. The proof is in the pudding: feasting and the origins of domestication. *Curr. Anthropol.* **50**, 597–601 (2009).
- Bellwood, P. *First Farmers: The Origin of Agricultural Societies* (Blackwell, Malden, MA, 2005).
- Cohen, M. N. *Food Crisis in Prehistory: Overpopulation and the Origins of Agriculture*. (Yale Univ. Press, New Haven, CT, 1977).
- Bar-Yosef, O. The Natufian culture in the Levant, threshold to the origins of agriculture. *Evol. Anthropol. Issues News Rev.* **6**, 159–177 (1998).
- Moore, A. M. T. & Hillman, G. C. The Pleistocene to Holocene transition and human economy in Southwest Asia: the impact of the Younger Dryas. *Am. Antiq.* **57**, 482–494 (1992).
- Zeder, M. A. & Smith, B. D. A conversation on agricultural origins. *Curr. Anthropol.* **50**, 681–690 (2009).
- Larson, G. et al. Current perspectives and the future of domestication studies. *Proc. Natl Acad. Sci. USA* **111**, 6139–6146 (2014).
- Smith, B. D. Low-level food production. *J. Archaeol. Res.* **9**, 1–43 (2001).
- Richerson, P. J., Boyd, R. & Bettinger, R. L. Cultural innovations and demographic change. *Hum. Biol.* **81**, 211–235 (2009).
- Boserup, E. *The Conditions of Agricultural Progress* (Aldine Publishing Company, Chicago, IL, 1965).
- Cunniff, J., Charles, M., Jones, G. & Osborne, C. P. Was low atmospheric CO<sub>2</sub> a limiting factor in the origin of agriculture? *Environ. Archaeol.* **15**, 113–123 (2010).
- Piperno, D. R., Holst, I., Winter, K. & McMillan, O. Teosinte before domestication: experimental study of growth and phenotypic variability in Late Pleistocene and Early Holocene environments. *Quat. Int.* **363**, 65–77 (2015).
- Binford, L. R. S. R. Binford & L. R. Binford. in *New Perspectives in Archeology* 313–341 (Aldine Publishing, Chicago, IL, 1968).
- Flannery, K. V., Ucko, P. J. & Dimbleby, G. W. *The Domestication and Exploitation of Plants and Animals* 73–100 (Duckworth, London, 1969).
- Binford, L. R. *Constructing Frames of Reference: An Analytical Method for Archaeological Theory Building Using Ethnographic and Environmental Data Sets*. (Univ. California Press, Berkeley, CA, 2001).
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).
- Lima-Ribeiro, M. S. et al. EcoClimate: a database of climate data from multiple models for past, present, and future for macroecologists and biogeographers. *Biodivers. Inform.* **10**, 1–21 (2015).
- Atkinson, Q. D., Gray, R. D. & Drummond, A. J. mtDNA variation predicts population size in humans and reveals a major southern Asian chapter in human prehistory. *Mol. Biol. Evol.* **25**, 468–474 (2008).
- Goldewijk, K. K., Beusen, A. & Janssen, P. Long-term dynamic modeling of global population and built-up area in a spatially explicit way: HYDE 3.1. *Holocene* **20**, 565–573 (2010).
- Lefcheck, J. S. piecewiseSEM: Piecewise structural equation modelling in R for ecology, evolution, and systematics. *Methods Ecol. Evol.* **7**, 573–579 (2016).
- Harcourt, A. H. *Human Biogeography* (Univ. California Press, Berkeley, CA, 2012).
- Kelly, R. L. *The Lifeways of Hunter-Gatherers: The Foraging Spectrum* (Cambridge Univ. Press, Cambridge, 2013).
- Hassan, F. A. S., *Polgar Population, Ecology, and Social Evolution*. 27–52 (Mouton: Berlin, 1975).
- Botero, C. A., Dor, R., McCain, C. M. & Safran, R. J. Environmental harshness is positively correlated with intraspecific divergence in mammals and birds. *Molec. Ecol.* **23**, 259–268 (2014).
- Marlowe, F. W. Hunter-gatherers and human evolution. *Evol. Anthropol. Issues News Rev.* **14**, 54–67 (2005).
- Belovsky, G. E. An optimal foraging-based model of hunter-gatherer population dynamics. *J. Anthropol. Archaeol.* **7**, 329–372 (1988).
- Kerr, J. T. & Packer, L. Habitat heterogeneity as a determinant of mammal species richness in high-energy regions. *Nature* **385**, 252–254 (1997).
- Jetz, W. & Rahbek, C. Geographic range size and determinants of avian species richness. *Science* **297**, 1548–1551 (2002).
- Kreft, H. & Jetz, W. Global patterns and determinants of vascular plant diversity. *Proc. Natl Acad. Sci. USA* **104**, 5925–5930 (2007).
- Sikora, M. et al. Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers. *Science* **358**, 659–662 (2017).
- Hayden, B. in *Last Hunters—First Farmers: New Perspectives on the Prehistoric Transition to Agriculture* (eds Price, T. D. & Gebauer, A. B.) Ch. 10 (School of American Research Press, Santa Fe, NM, 1995).
- Binford, L. R. & Johnson, A. L. *Documentation for Program for Calculating Environmental and Hunter-Gatherer Frames of Reference (ENVCALC2)* (2006).
- Kirby, K. R. et al. D-PLACE: a global database of cultural, linguistic and environmental diversity. *PLoS ONE* **11**, e0158391 (2016).
- Lisiecki, L. E. & Raymo, M. E. A Pliocene–Pleistocene stack of 57 globally distributed benthic δ<sup>18</sup>O records. *Paleoceanography* **20**, PA1003 (2005).
- Becker, J. J. et al. Global bathymetry and elevation data at 30 arc seconds resolution: SRTM30\_PLUS. *Mar. Geod.* **32**, 355–371 (2009).
- Lambeck, K., Rouby, H., Purcell, A., Sun, Y. & Sambridge, M. Sea level and global ice volumes from the Last Glacial Maximum to the Holocene. *Proc. Natl Acad. Sci. USA* **111**, 15296–15303 (2014).
- Raiche, G., Walls, T. A., Magis, D., Riepel, M. & Blais, J.-G. Non-graphical solutions for Cattell's scree test. *Methodol. Eur. J. Res. Methods Behav. Soc. Sci.* **9**, 23–29 (2013).
- Botero, C. A. et al. The ecology of religious beliefs. *Proc. Natl Acad. Sci. USA* **111**, 16784–16789 (2014).
- Nakagawa, S. & Schielzeth, H. A general and simple method for obtaining R<sup>2</sup> from generalized linear mixed-effects models. *Methods Ecol. Evol.* **4**, 133–142 (2013).
- Fewster, R. M., Buckland, S. T., Siriwardena, G. M., Baillie, S. R. & Wilson, J. D. Analysis of population trends for farmland birds using generalized additive models. *Ecology* **81**, 1970–1984 (2000).

45. Pearson, R. G., Raxworthy, C. J., Nakamura, M. & Townsend Peterson, A. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *J. Biogeogr.* **34**, 102–117 (2007).
46. R Core Team *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2016).
47. Hijmans, R. J. raster: Geographic Data Analysis and Modeling (2016); <https://CRAN.R-project.org/package=raster>
48. Bivand, R. maptools: Tools for Reading and Handling Spatial Objects (2017); <https://CRAN.R-project.org/package=maptools>
49. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
50. Vilela, B. & Villalobos, F. letsR: a new R package for data handling and analysis in macroecology. *Methods Ecol. Evol.* **6**, 1229–1234 (2015).

## Acknowledgements

This material is based on work supported by the National Science Foundation under grant no. 1519987. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Author contributions

M.C.G. secured funding. P.H.K., H.J.H., B.V., C.B. and M.C.G. contributed to the study design. M.L.-R. provided palaeoclimate data. P.H.K., B.V. and T.T. conducted analyses. P.H.K. wrote the initial draft of the manuscript and all authors contributed to revisions.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41562-018-0358-8>.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Correspondence and requests for materials** should be addressed to P.H.K.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated
- Clearly defined error bars  
*State explicitly what error bars represent (e.g. SD, SE, CI)*

*Our web collection on [statistics for biologists](#) may be useful.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

No software was used.

Data analysis

All analyses were conducted in the R environment for statistical computing (version 3.2.4). The following packages were used: raster, maptools, piecewiseSEM, lme4, and letsR. No custom code was used.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data for population density and cultural traits of societies are available from D-PLACE (<https://d-place.org>) and climate data are available via the eco-Climate project (<http://ecoclimate.org/>). All other data are available upon request.



## Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](https://www.nature.com/authors/policies/ReportingSummary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was chosen to be a large as possible and was determined by data availability.
Data exclusions	The only reason a society would have been omitted from the final dataset is due to an absence of required data (e.g. no information on variables of interest).
Replication	NA
Randomization	This is not specifically relevant to our study. However, a bootstrapping procedure was used when making predictions in an effort to remove bias.
Blinding	Not relevant to our study utilizing previously collected ethnographic data compiled in an online database.

## Reporting for specific materials, systems and methods

### Materials & experimental systems

- | n/a                                 | Involvement in the study                             |
|-------------------------------------|--|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Unique biological materials |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies                  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines       |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Human research participants |

### Methods

- | n/a                                 | Involvement in the study                        |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |