Title: Duration of agriculture and distance from the steppe predict the evolution of large-scale human societies in Afro-Eurasia

Online Supplementary Information

Authors: Thomas E. Currie[1*], Peter Turchin[2,3], Edward Turner[4], Sergey Gavrilets[5,6,7]

Affiliations:

[1] Human Behaviour and Cultural Evolution Group, Centre for Ecology & Conservation, Department of Biosciences, University of Exeter, Penryn Campus. UK

[2] Complexity Science Hub Vienna, Austria

[3] Department of Ecology & Evolution, University of Connecticut. USA

[4] Institute for Cognitive & Evolutionary Anthropology, University of Oxford, UK

[5] Department of Ecology & Evolutionary Biology & Department of Mathematics, University of Tennessee, Knoxville, USA

[6] National Institute for Mathematical and Biological Synthesis, University of Tennessee, Knoxville, USA

[7] Center for the Dynamics of Social Complexity, University of Tennessee, Knoxville, USA

* Corresponding author: T.Currie@exeter.ac.uk

Contents

List of Figures

List of Tables

Hypotheses

In this paper we use cultural evolutionary theory to develop, organize and synthesize a number of different hypotheses about the evolution of the largest-scale societies in human history. We developed 3 primary hypotheses relating to the socio-ecological conditions that might have caused the observed patterns in the occurrence of large-scale societies over time. We also identified other factors that relate to null-hypotheses about what might explain the distribution of such societies. These different hypotheses outlined in the main text make different predictions about the relationship between imperial density and different ecological and socio-cultural variables. In table S1 we summarize the factors we examined in this paper relating to these hypotheses, the variables we use to proxy these factors, and their predicted relationships with imperial density. We also developed two further hypotheses relating to the ways in which ecological or cultural similarity may affect the spread of cultural traits supporting large-scale societies. However, these were not straightforward to incorporate into the same statistical framework and therefore dealt with separately in an exploratory fashion (see Supplementary results below).

Table S1 *Potential factors important in the evolution of the largest-scale societies in human history. We identify proxy measures of these factors and outline the expected relationships with our measure of the historical duration of large-scale societies in different regions (imperial density).*

Factor		**Proxy**	**Prediction** *Imperial density…*
Main factors	*Warfare intensity*	*Distance from the Eurasian steppe*	*…decreases with increasing distance from the steppe*
	Duration of agriculture	*Estimated time since adoption of agriculture*	*…increases with time since adoption of agriculture*
	Agricultural productivity	*Estimated potential productivity of traditional agriculture under present-day conditions*	*…increases with potential productivity*
Control factors	*Terrain*	*Standard deviation of altitude*	*…decreases in more uneven regions*
	Proximity to earliest empires	*Distance from large-scale societies at 1500BCE*	*…decreases with increasing distance from location of first empires*

Methods

Data

Data were organized in grid cells created in ArcGIS v.10 under an equal-area projection so as to maintain a constant cell size of 10,000km^2 (approximately 100km by 100km). Analyses were conducted using cells in which agriculture was practiced by 1500 CE, and cells in which potential agricultural production was possible (i.e. potential productivity>0).

Imperial density: Data on the global and historical distribution of large-scale societies were collated from historical atlases and related sources to create hand-drawn reference maps of the geographic distribution of large, well-attested polities at 100-year time-steps (from 1500 BCE to 1500 CE)(Appendix I). Polygons of the societies at each time step were created from these maps in ArcGIS and areas were calculated under an equal-area projection. For the main analysis we focus on polygons >100,000km^2, but we also examine the effects of altering this threshold in confirmatory analyses (see below). Maps of the cells inhabited by large-scale societies at each 100-year time step are shown in Appendix I.

Duration of agriculture: We created new maps that give an estimate of the time since agriculture was practiced in various parts of the world by assembling data of the dates of earliest evidence of reliance on agriculture in different parts of the world. These dates were entered into ArcGIS as points. In order to gain full global coverage, these data were interpolated using the Spline algorithm, which creates a smooth surface that goes through the input data points. In order to incorporate uncertainty about these estimated dates, we propose a "best" estimate as well as minimum and maximum estimates (see Appendix II). Details of the sources and explanations of dates are provided as an appendix to this document (see Appendix II).

Agricultural productivity: The *agricultural productivity* hypothesis reflects the natural endowment of conditions conducive to productive agriculture (e.g. climate, broad available crop type, soil) rather than achieved agricultural production that may rely on technological or cultural innovations that raise productivity (cf. (*1*)). Data on estimated potential crop productivities were taken from the Food and Agriculture Organization of the UN's Global Agro-Ecological Zones (FAO GAEZ) methodology v.3 (*2*) . Different crop types were used in different regions to reflect the kinds of crops that were grown historically. We make the simplifying assumption that overall potential agricultural productivity can be reasonably proxied by the maximum potential yield from the main carbohydrate staple crop. Wheat (representing temperate cereals in general) was used as the main crop for Europe, North Africa, and western to central Eurasia (as far as India), rice was the main crop for East and Southeast Asia and India, Sorghum (representing tropical cereals) is used in sub-Saharan Africa, whilst the potential productivities of Yam were estimated in sub-Saharan Africa and Southeast Asia and New Guinea/Melanesia (see Fig S2). Here we use estimates from the agro-ecological suitability and productivity model which estimates crop productivity based on soil fertility, crop management input levels, water supply, and climatic conditions. In most places we use estimates based on rainfed production. However, in some important places such as the middle-East and Egypt large rivers run through arid regions in which rainfed estimates would be zero. Therefore in such regions we take the estimated productivity for gravity irrigation. As rice (unlike wheat) does well in water-logged soils and is commonly grown in wet fields we also take the estimated productivities for rice cultivation under gravity

irrigation where this is greater than under rainfed conditions. Medium input levels were used for both rainfed and irrigated regions to hold constant the effects of labour or crop management practices on productivity. Climatic effects were based on the baseline average climatic conditions from 1961 to 1990. Crop productivity estimates from the FAO GAEZ model are given in terms of weight (kg/ha). To account for differences in calorific content between crop types we multiply values by the number of kcals per kg based on information from USDA Food Composition Databases (http://ndb.nal.usda.gov/ndb/foods/)(Wheat: 3290, Rice: 3650, Sorghum: 3290, Yam: 1180). We calculated the measure of potential productivity per cell as the sum of the 100 raster tiles per grid cell (adjusting proportionally for cells that are not completely land). The measure takes into account the fact that some areas of land may not be productive but does not penalize cells that have less land (to get an average measure of kcal per hectare the values presented in our dataset can be divided by 100). In some places (i.e. India, SE Asia, sub-Saharan Africa) it can be seen that more than one crop could potentially be used to estimate productivities. In these places for each cell we use the crop that had the highest estimated productivity for that cell.
Distance measures: Great Circle distances (in km) from the Steppe, and distance from the first empires were calculated in R using the package *geosphere* (Version 1.5-7).
Steppe: The steppe region was defined according to the WWF terrestrial ecoregions of the world map and includes all the contiguous regions of steppe that extend from Europe to the Far East (*3*). To assess the sensitivity of our assumptions about the origins of the spread of horse-based warfare we also created a more minimal steppe classification that did not extend below the Caucasus Mountains in Eastern Europe or into the Tarim Basin in Central Asia (Fig S1). We used the maximum extent of the steppe in our main analyses and explored the effect of assuming the more minimal extent in supplementary analyses (see below).
First Empires: The first empires are defined as those regions that had empires >100,000 km^2 at 1500BC (adjusted variables were created for the analyses where we explored adjusting this threshold – see below).
Elevation: Elevation data (in metres) was taken from the GTOPO30 digital elevation model of the world (USGS)(resolution: 30 arc-seconds) (*4*). The measure of the unevenness of terrain was calculated as the standard deviation of altitude across each grid cell.
Geographical distributions of predictor variables and their relationships with imperial density are shown in Fig S3.

Data availability
Data, R code, and sources used in these analyses were included with the submission and are openly available at Harvard Dataverse https://doi.org/10.7910/DVN/8TP2S7 and currielab.com.

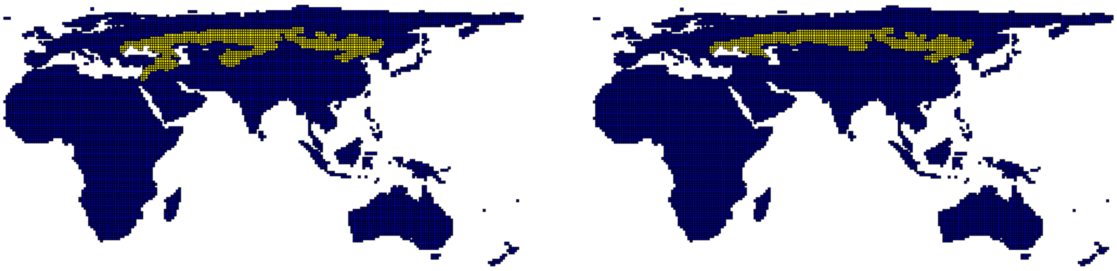


Figure S1. *Different assumptions about the extent of the Eurasian steppe used to calculate our proxy for warfare intensity (distance from the steppe). Left: Maximum extent of the steppe, Right: Minimum extent.*

Figure S2. *Distributions of main crops used to calculate potential agricultural productivity based on historical presence. In regions that have more than one candidate crop, the crop with the highest estimated productivity was used.*

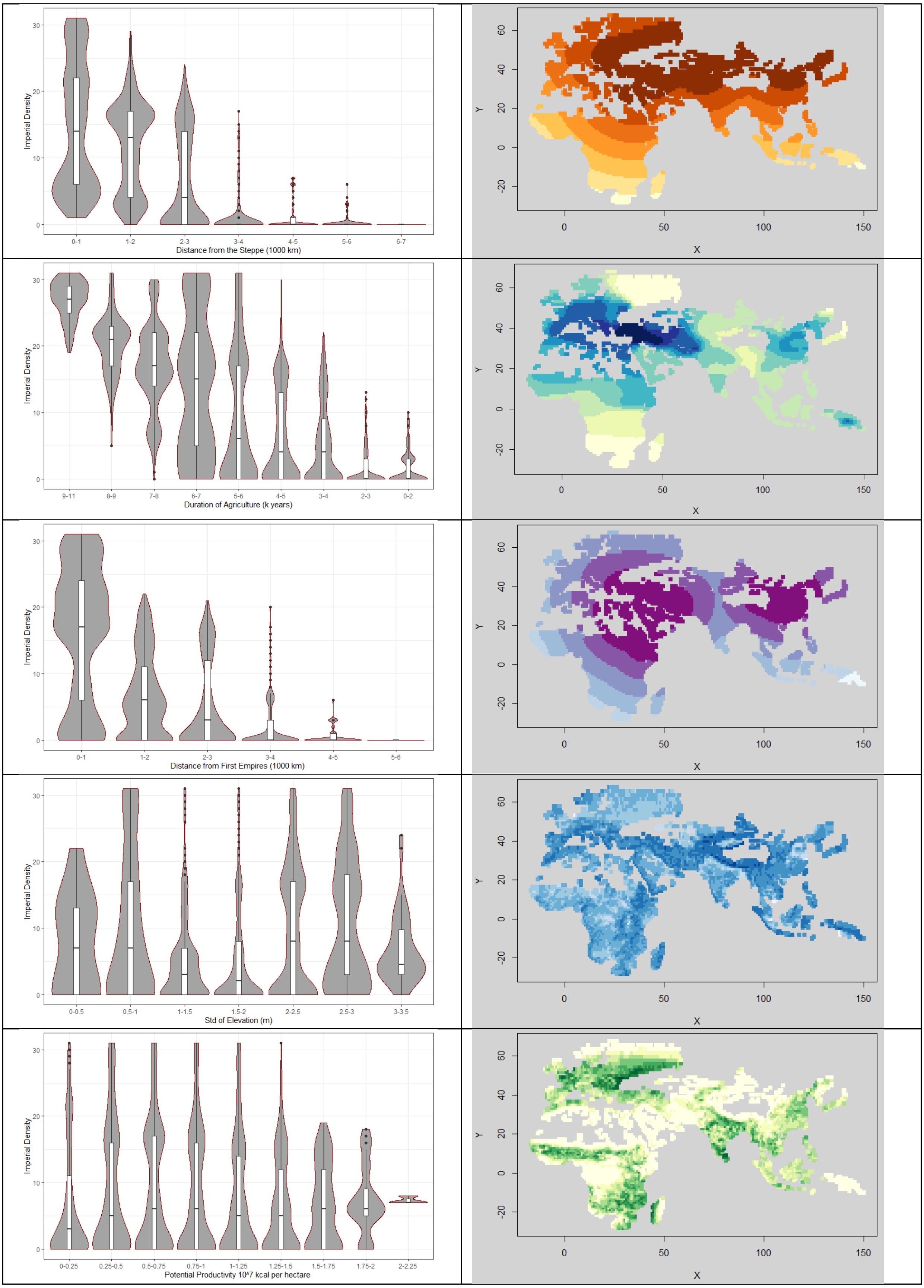


Figure S3. *Geographical distributions of main predictor variables (right) and their relationship with imperial density (left)*

Statistical analyses

Main analyses were carried out using the `nlme` (non-linear mixed-effects) package (*5*) in R (*6*), using longitude and latitude as random control variables (following (*7*)) and with an adjustment so that distances were calculated based on great-circle distances https://github.com/toph-allen/nlmehaversine). Ordinary least squares (OLS) regression assumes that errors are independent. Our spatially sampled grid cells may violate this assumption as points in space are not independent and may have similar values. This can lead to inaccurate parameter estimates and inflated assessments of statistical significance. Generalized least squares (GLS) regression allows us to model the expected co-variance between our cells and control for this spatial autocorrelation in the residuals of the statistical models we specify. Whilst OLS regression can be easily run on our data, including the spatial structure of the data in `nlme` using GLS is extremely memory intensive, and it was not possible to run GLS on the full dataset. Given that a large number of modelling assumptions need to be tested and a large number of alternative models need to be run in order to test alternative hypotheses (see below) we ran analyses on random sub-samples of the data involving 1000 cells. For the main analyses we ran analyses over a number of such samples in order to ensure that results were not dependent on a particular sample (see below).

All dependent and predictor variables were scaled using the `scale()` command in R, in order to produce standardized parameter estimates. Some variables were reflected so that all predicted relationships would be positive to ease interpretation of results.

Detecting and controlling for spatial autocorrelation

Unsurprisingly, given the gridded nature of the data, exploratory analyses indicate the presence of spatial autocorrelation in the residuals of an ordinary least squares analyses. This can be demonstrated through examining the spatial relationships between the residuals of a model in which distance from the steppe (max), duration of agriculture (Best), Standard deviation of Altitude, and potential productivity are used to predict imperial density (100km^2 threshold). The spatial autocorrelation between the OLS residuals can be seen in the variogram below (Fig S4), which shows that the variance in the difference between cells is lower at closer spatial distances.

Here we use an adjusted version of the R package "nlme" to fit our GLS models. The package allows various spatial correlation structures to be fitted. Here we examined four different spatial correlation structures (Spherical, Ratio, Gaussian, Exponential) using model selection via AIC (see table S2) and examination of variograms and mapped residuals (Fig S4) Analyses were run with starting values for the range and nugget (where appropriate) parameters but we allowed the model to search for values for these parameters (and the parameters for the predictor variables) that maximized the likelihood. Analyses indicated that the rational quadratic spatial correlation structure (corRatio) would be most appropriate. corRatio models the correlation between two observations a distance "d" apart is $(1-n)/(1+(d/r)^2)$ (where "r" is the range parameter). Including a nugget effect "n" (i.e. $n>0$) appeared to substantially improve the fit of the correlation structure.

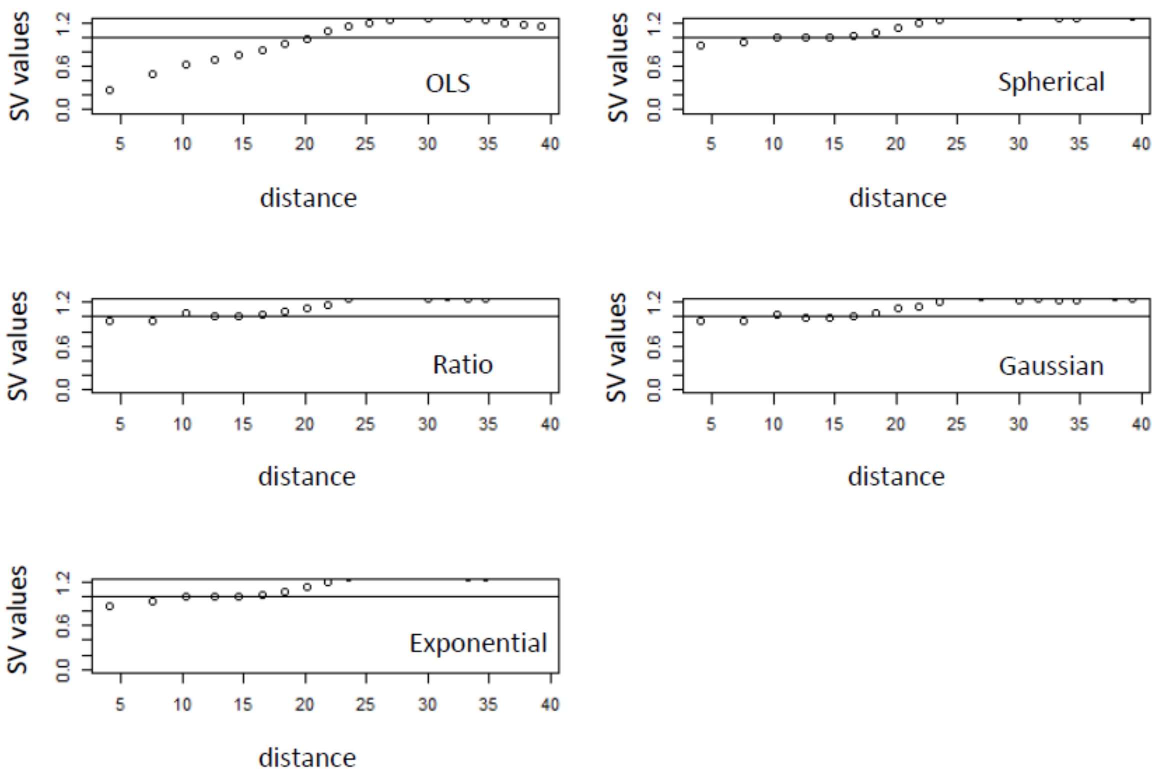


Figure S4 *Assessing and controlling for spatial autocorrelation. Semivariograms of the normalized residuals from an OLS model (here imperial density is regressed on the variables mentioned in the text)(top left), and those of GLS models with the same variables but with longitude and latitude include as random variables to control for spatial autocorrelation (others). Residuals from the OLS show less variance between cells at closer distances which is indicative of spatial autocorrelation. The GLS models vary as to how they model the spatial correlation structure but all models show that they have controlled for the spatial autocorrelation as the variance values do not increase with increasing distance.*

Table S2 *Different GLS models described in the text with different spatial correlation structures. The corRatio model with a "nugget" effect produced the best fit to the data in this example. Including spatial error structure shows large and substantial improvements in fit compared to OLS, which indicates the importance of controlling for spatial autocorrelation.*

Spatial correlation structure	**nugget**	**df**	**AIC**
Ratio	Yes	9	-111.36
Spherical	No	8	-44.14
Spherical	Yes	9	-42.14
Exponential	No	8	-34.45
Exponential	Yes	9	-32.45
Gaussian	Yes	9	-18.95
Ratio	No	8	29.90
Gaussian	No	8	447.25
none (OLS)	-	7	1969.85

Testing alternative models

In order to test between the alternative hypotheses we specified different statistical models containing different predictors of imperial density. We ran main models containing the predictor variables of duration of agriculture, potential productivity of agriculture, terrain ruggedness, and distance from the steppe. Running these models is computationally expensive and so such judgments need to be made for practical purposes. We also assessed to what extent model fit and parameter estimates were affected by including distance from first empires as a variable in these models.

In order to assess whether the effects of duration of agriculture and warfare intensity are amplified or inhibited by each other we included interaction terms in the main models. This allows us to see whether the effect of warfare intensity is stronger when places have been practicing agriculture for longer.

We compared different models in a model selection framework (*8*) based on Akaike Information Criteria (AIC)(*9*). The AIC is calculated as $-2*LLH + 2k$, where LLH is the log likelihood of the model and k is the number of parameters in the model.
Each model was run on 20 random samples of 1000 cells to assess variation in the parameter estimates. As likelihoods are not directly comparable across samples we use change in AIC (ΔAIC, delta AIC) to guide model comparability across samples.

Models over time

In order to assess if there were any changes in the strength of our different main factors as predictors of imperial density over time we ran analyses over a sliding time frame of 1000 years. In other words we first analyse a model containing duration of agriculture, terrain ruggedness, productivity of agriculture, and distance from the steppe, and the interaction between duration of agriculture and distance from the steppe as predictors of imperial density for the period 1500BCE-600BCE. We then analyse the period 1400BCE-500CE etc. Within these time frames we only include areas to which agriculture had spread by the end of the time window. Each model was run on 10 random samples of 1000 cells to assess variation in the parameter estimates.

Results

Table S3. *GLS spatial models with different predictor variables ranked according to how well they fit the data based on mean AIC values from across the 20 sub-samples of data. Standardized coefficients (β) are presented to indicate the relative strength of each predictor. Models are presented here both with and without the control variable "distance from first empires" to assess the effect including it has on other parameter estimates.*

Model no.	***Intercept***	***Duration of agriculture***	***Agricultural productivity***	***Elevation (Std)***	***Distance from Steppe***	***Distance from First Empires***	***Steppe* Agriculture***	***Mean AIC***
6	*-0.15 (0.01)*	*0.24 (0.04)*			*0.47 (0.02)*		*0.15 (0.02)*	*-76.12*
15	*-0.15 (0.01)*	*0.24 (0.04)*	*0.00 (0.01)*	*0.01 (0.01)*	*0.46 (0.02)*		*0.15 (0.02)*	*-74.51*
9	*-0.09 (0.01)*	*0.32 (0.04)*			*0.38 (0.02)*			*-69.80*
16	*-0.09 (0.01)*	*0.32 (0.04)*	*0.00 (0.01)*	*0.01 (0.01)*	*0.38 (0.02)*			*-68.26*
1	*-0.19 (0.02)*	*0.37 (0.05)*						*-50.28*
4	*-0.13 (0.02)*				*0.49 (0.02)*			*-42.60*
14	*-0.28 (0.02)*							*-17.58*
3	*-0.28 (0.02)*			*0.01 (0.01)*				*-17.14*
2	*-0.28 (0.02)*		*0.00 (0.01)*					*-16.51*
8	*-0.09 (0.02)*	*0.21 (0.04)*			*0.16 (0.06)*	*0.34 (0.06)*	*0.13 (0.02)*	*-84.29*
7	*-0.09 (0.02)*	*0.22 (0.04)*	*0.00 (0.01)*	*0.01 (0.01)*	*0.16 (0.06)*	*0.34 (0.06)*	*0.13 (0.02)*	*-82.70*
12	*-0.03 (0.02)*	*0.28 (0.04)*				*0.42 (0.03)*		*-81.93*
10	*-0.03 (0.02)*	*0.28 (0.04)*			*0.05 (0.05)*	*0.39 (0.05)*		*-80.27*
11	*-0.03 (0.02)*	*0.28 (0.04)*	*0.00 (0.01)*	*0.01 (0.01)*	*0.05 (0.05)*	*0.39 (0.06)*		*-78.74*
5	*-0.03 (0.02)*					*0.55 (0.02)*		*-62.51*
13	*-0.03 (0.02)*				*0.02 (0.06)*	*0.53 (0.05)*		*-60.70*

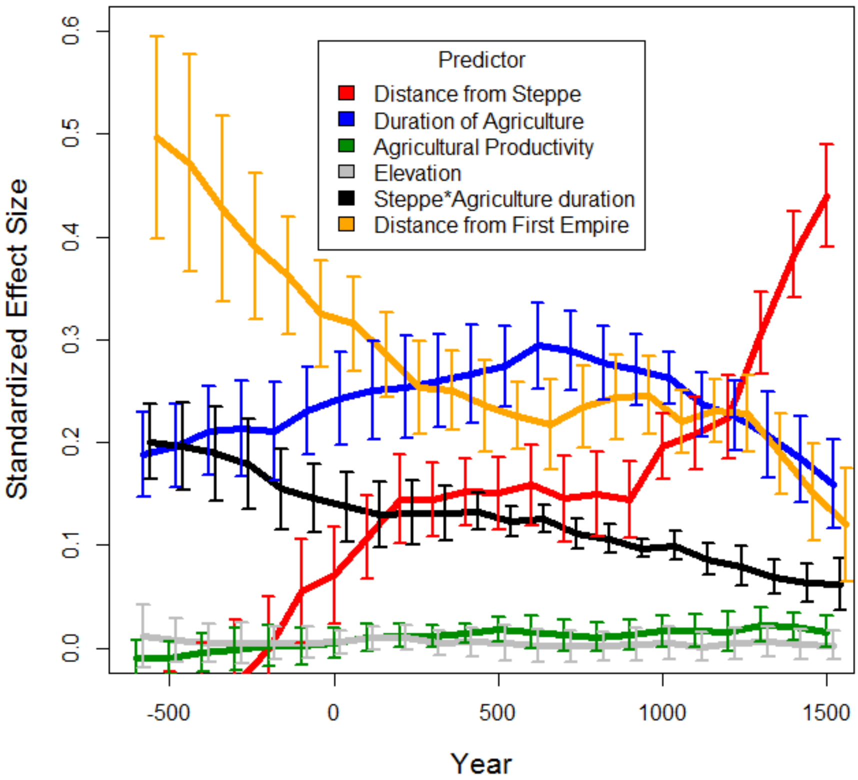


Figure S5. *Variation in predictive strength of variables from the full model including distance from first empires over the sampled period. Graphs show the standardized parameter estimates of variables (+/- SD) as predictors of imperial density over the 1000 years prior to the date indicated on the X-axis (negative values on this axis indicate BCE, positive values indicate CE).*

Model selection

For each of the 20 samples we calculated AIC for each of the main models, and ΔAIC for each model within that sample. Models were ranked according to their AIC (lower values indicating better fits to the data). For each sample we selected the models that were within 2 AIC units of the best-fitting model for that sample. While there was some expected variation in ΔAIC and rankings between samples a consistent pattern emerged across samples. The model that included only the effects of duration of agriculture and distance from the steppe and the interaction between these two was the best-fitting in model in 17 out of 20 sub-samples. In the remaining 3 sub-samples the best-fitting model included these terms as well as standard deviation of elevation and agricultural productivity. In no other sub-sample was a model other these two less than 2 AIC units of the best model (Table S4).

Table S4. *Summary of fit of models across sub-samples. Within each sub-sample models were ranked according to their AIC, and differences in AIC from the best model in that sub-sample were calculated. The distribution of ΔAIC is summarized as the minimum (min), maximum (max), median (Med), and 25th (25Q) and 75th (75Q) percentiles. Only two models (Ag+St+Ag*St and Ag+St+Ag*St+Y+E) are ever the best-fitting models (min ΔAIC =0)*

			Distribution of relative likelihoods (ΔAIC) across samples				
Model	**mean rank**	**mean ΔAIC**	**Min**	**25Q**	**Med**	**75Q**	**Max**
Ag+St+Ag*St	1.15	0.23	0.00	0.00	0.00	0.00	3.57
Ag+St+Ag*St+Y+E	1.95	1.85	0.00	0.41	1.74	3.24	4.00
Ag+St	3.1	6.56	2.00	4.83	6.65	8.01	11.80
Ag+St+Y+E	3.8	8.10	3.78	6.39	8.06	10.07	13.21
Ag	5.2	26.08	16.57	24.17	26.36	28.74	38.01
St	5.8	33.76	15.01	26.01	33.91	39.44	49.35
Constant only	7.45	58.78	32.77	50.82	61.21	65.10	74.52
E	7.95	59.22	31.52	50.98	62.47	66.72	75.84
Y	8.6	59.85	32.97	52.33	61.43	66.66	75.25

Ag: Duration of agriculture, St: Distance from the Steppe, Y: potential productivity, E: terrain ruggedness

CONFIRMATORY ANALYSES

Assessing uncertainty in variables

The construction of the variables we have used to examine to different hypotheses in this paper rest on certain assumptions and rely on the information used to construct them. As such there is a degree of uncertainty in the measurement of these variables. In order to assess whether this uncertainty affects the overall findings we created different measures for the variables of interest.

Predictor variables

Specifically for the variables that our main analyses revealed to be the strongest predictors of imperial density (distance from the steppe, and duration of agriculture) we created different measures based on different assumptions. For distance from the steppe we created a variable based on a narrower definition of the steppe that leads to two major changes: the minimal definition does not extend a) below the Caucasus Mountains into the Middle East, and b) into the Tarim Basin in Central Asia (see Fig S1). For duration of agriculture, dates from the literature on the beginnings of agriculture in different regions are often presented as ranges. In addition to our "best" estimate we therefore also created a minimum and maximum estimate based on the information available (see Appendix II).

As Table 1 in the main text shows, these variables show very similar patterns in terms of their raw correlations with imperial density. The "best" duration of agriculture estimate shows a correlation coefficient of 0.57, while it slightly higher for the minimum estimate (0.61) and slightly smaller for the maximum estimate (0.54). For distance for the steppe the correlation coefficients are very similar for both the maximum (0.71) and minimum (0.72) measurements. However, these estimates do not take into account spatial autocorrelation, nor the effects of including other variables. We therefore ran confirmatory GLS analyses using `nlme` that examined combinations of the different duration of agriculture and distance from the steppe measures. The results were extremely consistent across these different parameter combinations (Table S5) demonstrating that our main findings are robust to different assumptions about the precise values of these variables.

Table S5. *Confirmatory GLS analyses examining the effect of using different measures of the duration of agriculture and distance from the steppe variables. Standardized parameter estimates are extremely similar across similar models (i.e. those with or without the interaction between the duration of agriculture and distance from the steppe variables). For comparison the parameter estimates in the main analyses were: duration of agriculture (best) = 0.24, distance from the steppe (max) = 0.46, interaction = 0.15, and duration of agriculture (best) = 0.32, distance from the steppe (max) = 0.38 when the interaction term was not included. Analyses were run on 5 sub-samples of the data and the mean parameter estimates across these run are shown here (values in parentheses are the standard deviation of the parameter estimates across these 5 samples).*

		Duration of agriculture			Distance from the Steppe		Distance from First Empires	interaction
Model	intercept	Best	Min	Max	Min	Max		
1	-0.11 (0.04)	0.36 (0.07)			0.33 (0.04)			
2	-0.06 (0.04)		0.39 (0.06)			0.36 (0.03)		
3	-0.11 (0.04)		0.39 (0.06)		0.31 (0.04)			
4	-0.06 (0.05)			0.31 (0.07)		0.41 (0.03)		
5	-0.11 (0.04)			0.31 (0.07)	0.37 (0.04)			
6	-0.16 (0.03)	0.28 (0.06)			0.44 (0.04)			0.15 (0.02)
7	-0.12 (0.04)		0.29 (0.04)			0.44 (0.03)		0.15 (0.04)
8	-0.16 (0.03)		0.31 (0.06)		0.40 (0.04)			0.13 (0.02)
9	-0.11 (0.05)			0.25 (0.05)		0.48 (0.02)		0.17 (0.04)
10	-0.16 (0.04)			0.24 (0.05)	0.47 (0.03)			0.17 (0.03)
11	-0.02 (0.04)	0.32 (0.06)			0.06 (0.04)		0.35 (0.02)	
12	-0.07 (0.04)	0.25 (0.06)			0.16 (0.04)		0.32 (0.02)	0.12 (0.02)
13	-0.02 (0.04)		0.36 (0.06)			0.11(0.04)	0.29 (0.03)	
14	-0.07 (0.03)		0.27 (0.04)			0.22 (0.05)	0.25 (0.03)	0.13 (0.04)
15	-0.02 (0.04)		0.36 (0.06)		0.03 (0.04)		0.35 (0.02)	
16	-0.07 (0.04)		0.29 (0.06)		0.12 (0.05)		0.32 (0.02)	0.10 (0.02)
17	-0.01 (0.05)			0.28 (0.07)		0.15 (0.04)	0.30 (0.05)	
18	-0.07 (0.05)			0.23 (0.06)		0.28 (0.05)	0.23 (0.06)	0.15 (0.04)
19	-0.02 (0.05)			0.28 (0.07)		0.07 (0.04)	0.37 (0.03)	
20	-0.07 (0.04)			0.22 (0.06)		0.20 (0.05)	0.31 (0.04)	0.14 (0.03)

Outcome variable

To assess uncertainty in our measure of imperial density we assessed the effect of changing the threshold for inclusion in our analyses. Our main analyses are based on the assumption that only polities greater than 100,000 km^2 should be included in the analyses. We therefore ran confirmatory GLS models that changed the threshold for inclusion by 20% both up (120,000 km^2) and down (80,000 km^2). The parameter estimates from these models (Table S6) were extremely similar to the main analyses indicating our main findings are likely to be robust to sources of error or uncertainty in our measurement of the distribution of large scale societies in human history.

Table S6. *Confirmatory analyses examining the effect of using different size inclusion thresholds for the imperial density variable. Standardized parameter estimates are extremely similar to the equivalent models from the main analyses. Analyses were run on 10 sub-samples of the data and the mean parameter estimates across these runs are shown here (values in parentheses are the standard deviation of the parameter estimates across these 10 samples).*

Imperial density threshold	**Intercept**	**Duration of agriculture (Best)**	**Distance from the Steppe (max)**	**Potential productivity**	**Ruggedness**	**FE**	**interaction**
80,000km^2	-0.13 (0.02)	0.23 (0.04)	0.47 (0.04)	0.007 (0.01)	0.009 (0.01)		0.15 (0.03)
120,000km^2	-0.14 (0.02)	0.24 (0.03)	0.47 (0.04)	0.009 (0.01)	0.009 (0.01)		0.15 (0.02)
80,000km^2	-0.07 (0.02)	0.21 (0.04)	0.21 (0.06)	0.007 (0.01)	0.009 (0.01)	0.29 (0.06)	0.13 (0.03)
120,000km^2	-0.08 (0.02)	0.21 (0.03)	0.15 (0.07)	0.009 (0.01)	0.009 (0.01)	0.36 (0.06)	0.14 (0.02)

Alternative GLS implementation

To assess whether the results were robust to different implementations of the GLS analyses we conducted confirmatory analyses using the program SAM (Spatial Analysis in Macroecology - v4.0)(*10*). We ran a GLS model of total imperial density (threshold 100,000 km^2) as predicted by distance from the Steppe (Max), duration of Agriculture (Best), Standard deviation of altitude, and Potential productivity (interaction terms are not implemented in SAM). Spatial autocorrelation was modelled using a Spherical model for the error structure with input parameters for the error structure being $C_0 = 0$, C_1, = 45, $\alpha = 4000$. It was possible to run the analysis on the full dataset so sub-sampling was not required. Fig S6 shows that these analyses control for spatial autocorrelation effectively. The analysis produces results that are quantitatively and qualitatively similar to the *nlme* analyses in R (Table S7), and the predictor variables in this model explain around 52% of the variance in total imperial density (r^2=0.516).

Table S7. *Parameter estimates from a GLS model implemented in the program SAM with a spherical error structure. Parameter estimates are very similar to the analyses in the nlme package in R, and the p values indicate the same findings from our model selection analyses that distance from the steppe and duration of agriculture predict imperial density well, but productivity and ruggedness do not.*

Variable	**Standarized parameter estimate**	**p**
Intercept	2.9	0.009
Distance from the Steppe (max)	-0.49	<0.001
Duration of agriculture (Best)	-0.25	<0.001
Potential productivity	-0.002	0.775
Ruggedness	-0.005	0.248

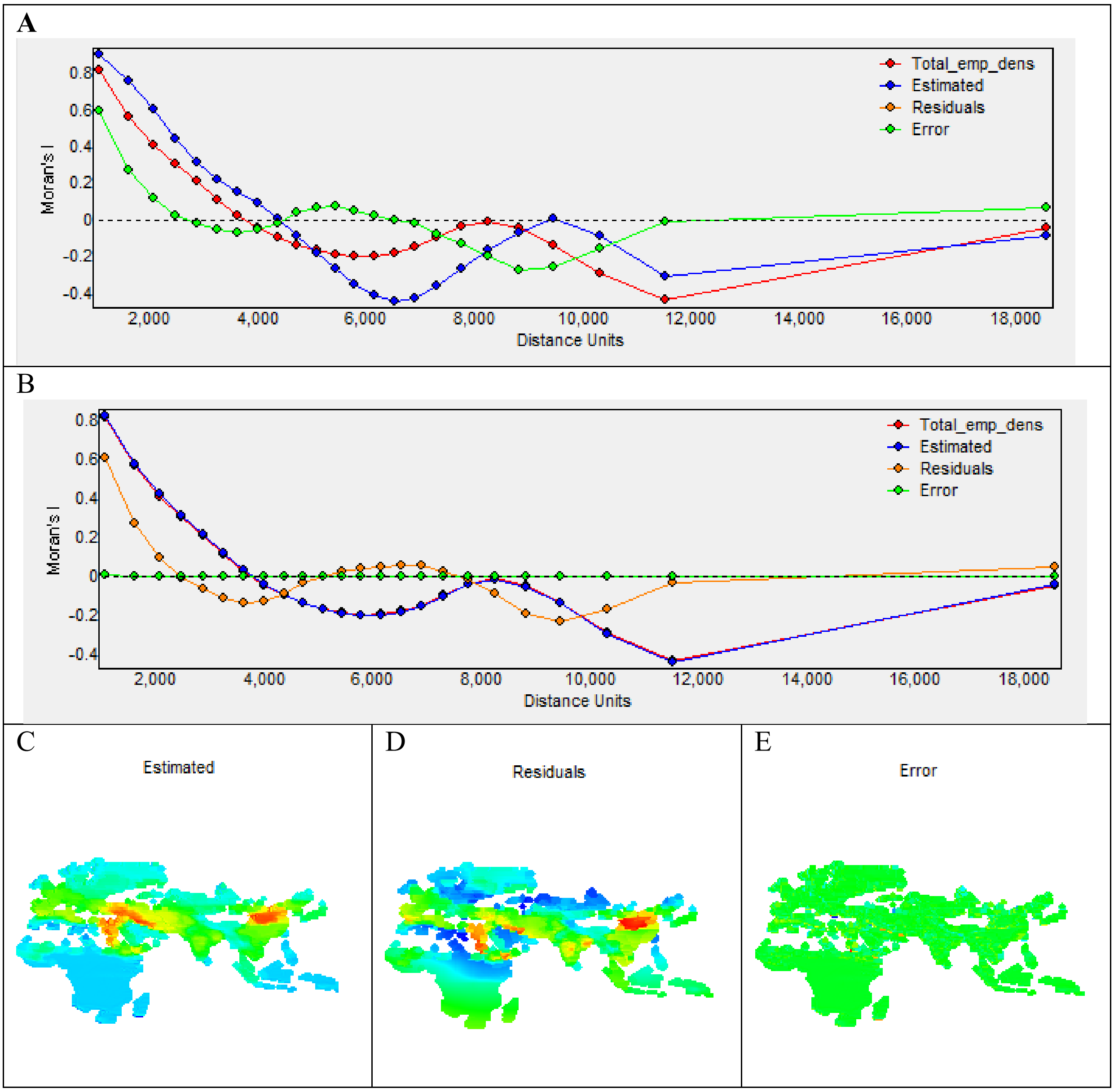


Figure S6. *SAM analysis output showing that the GLS model has effectively controlled for spatial autocorrelation. Moran's I values (a measure of spatial autocorrelation) for the outcome variable, model estimated values, residuals, and errors (A & B). Here we see a comparison between a GLS model with an error structure that does not model spatial autocorrelation effectively for these data (a pure nugget model [C_0 = 45] that assumes spatial autocorrelation does not change with distance)(A),and a GLS model with a spherical error structure (see text for parameter inputs)(B). The errors in A show the same spatial pattern as the residuals (indeed they are obscuring them in this figure) - high Moran's I values at shorter distances between points. This spatial pattern has been removed in B however, and Moran's I values are low (near 0) and consistent, indicating spatial autocorrelation has been controlled for. The maps show the estimated values from the GLS model with the Spherical error structure (C), the residuals (D) and the error (E) from this analysis. While the spatial pattern remains in the residuals, the errors are uniform and show no spatial pattern which is the aim for such analyses. Colour bar values go from low (dark blue) through green (middle values), to red (high values).*

A preliminary test of the ecological and cultural similarity hypotheses

As discussed in the main manuscript, traits that enable societies to become spread over larger regions may be more likely to be transmitted between societies that are similar ecologically and culturally, or human groups may find it easier to spread and expand their control over regions that are similar ecologically or culturally. A prediction that follows from these ideas is that regions where cultural or ecological similarity extends over larger areas will be home to greater imperial density because i) the rate at which societies can acquire traits that enable larger scale societies is greater, ii) the opportunity to extend control over a larger area is greater. As a preliminary test of these predictions we systematically classify the old world into regions of ecological and cultural similarity and assess whether imperial density is greater where these regions of similarity are larger. In order to examine the logic of the original idea by Diamond about the latitudinal extent of the continents shaping relative patterns of social development we also assess whether zones that extend further in the East-West axis have higher imperial density.

For the ecological similarity hypothesis, we classified ecological zones according to the zoogeographic zones identified by Holt et al. (*11*) based on the distributions and phylogenetic relationships of a large number of vertebrate species. These zones therefore reflect regions in which similar animal species are found (Fig S7). The area covered by each zoogeographic zone was calculated under an equal area projection. We also calculated the approximate distance between the furthest point West and the furthest point East for each zone. We calculated these measures taking into account the fact that the distance between points of longitude varies at different latitudes. The difference in latitude at the East and West points was generally small so we calculated distances assuming the latitude of the furthest point West for that zone.

For the cultural similarity hypothesis, we classified cultural zones (i.e. regions of greater cultural similarity and shared cultural descent) based on language families. We created a map based on figure 2 of Diamond & Bellwood (*12*). We use language family distributions here as they are thought to reflect deep cultural or demographic expansions that took place prior to the time period considered here. Societies belonging to a particular language family are thought to have descended from a shared common ancestral society and may therefore share more features in common with each other than with societies from other language families. As this is a preliminary test this map was simplified in a few ways to capture the broadest patterns of cultural similarity and to reflect the fact that the language family affiliation in some regions has changed over the course of the time span being considered here, e.g. the regions occupied in the present day by Hungarian, which is a Uralic language, and Turkish, which an Altaic language, are classified as Indo-European in our map, reflecting the relatively late arrival of these group into these regions. As with the zoogeographic zones, the area covered by each language family was calculated under an equal area projection. The differences in latitude between East and West points tended to be greater for language families than was the case for the zoogeographic zones. We therefore calculated East-West distances in two ways – one based on the lowest value for latitude at either East point or West point, and one based on the maximum value for latitude at either point. In R we overlaid the cultural and ecological zone polygon maps on a raster map of imperial density, and extracted values for each zone.

The results show that there is substantial variation between ecological and cultural zones in terms of imperial density (Figs S8 & S9, Tables S8 & S9). However there is not an apparent and clear relationship between imperial density and the area covered by these zones, nor is there a clear relationship between imperial density and the latitudinal extent of these zones. Therefore the predictions of the ecological and cultural similarity hypotheses as outlined above do not appear to be well supported.

The preliminary nature of these analyses should be re-emphasized and there are a number of reasons that these hypotheses may not have received support from these analyses. Importantly, while our categorizations of ecological and cultural zones have the advantage of being suitably coarse grain for simple tests, they may not be the most appropriate classifications for testing these ideas. For example, we have chosen a broad, simplified language family distribution that reflects the period around the end of our analysis period, yet we know that there are some issues here such as it doesn't capture variation in similarity within language families (e.g. in sub-Saharan Africa Bantu languages are more closely related to each other than they are to other branches of the Niger-Congo family (*13*)), New Guinea is treated as whole despite there being many language families here (*14*), and some classifications, like Altaic, are controversial (*15, 16*). Similarly, the ecological zones classification we used was chosen because it reflected well the logic of the original argument about latitudinal axes (*17*) but these could of course be drawn in different ways by making different assumptions (*18*). An important point is that both regions of cultural and ecological similarity may change over time, even over the 3000 year period covered in our data. Climate change and other environmental phenomena can alter the distribution of regions of ecological similarity. Cultural similarity will be affected by movements of populations or even potential large-scale horizontal transfer of traits that are not well captured by the simple language family classifications employed here. Quantitative methods are increasing our ability to infer changes in culture over time (e.g. (*19*)), and it may be possible to infer historical changes in ecological zones too. Future work could potentially create time sensitive maps of cultural and ecological similarity, or could examine other means of assessing these hypotheses such as examining alternative classifications based on other criteria (e.g. plant-based ecozones (*3*), or shared religion), or using more quantitative means of assessing similarity. Another way of assessing these hypotheses might be to examine more closely the connectivity between different areas that is produced by ecological or cultural similarity. This may be difficult to operationalize within the statistical framework used in this paper, but could perhaps be tackled using computer simulations. Despite the limitations of the current preliminary analyses they provide an important initial, broad-scale assessment of these hypotheses and also help to demonstrate some of the challenges involved in operationalizing and testing what initially appears to be a straightforward hypothesis.

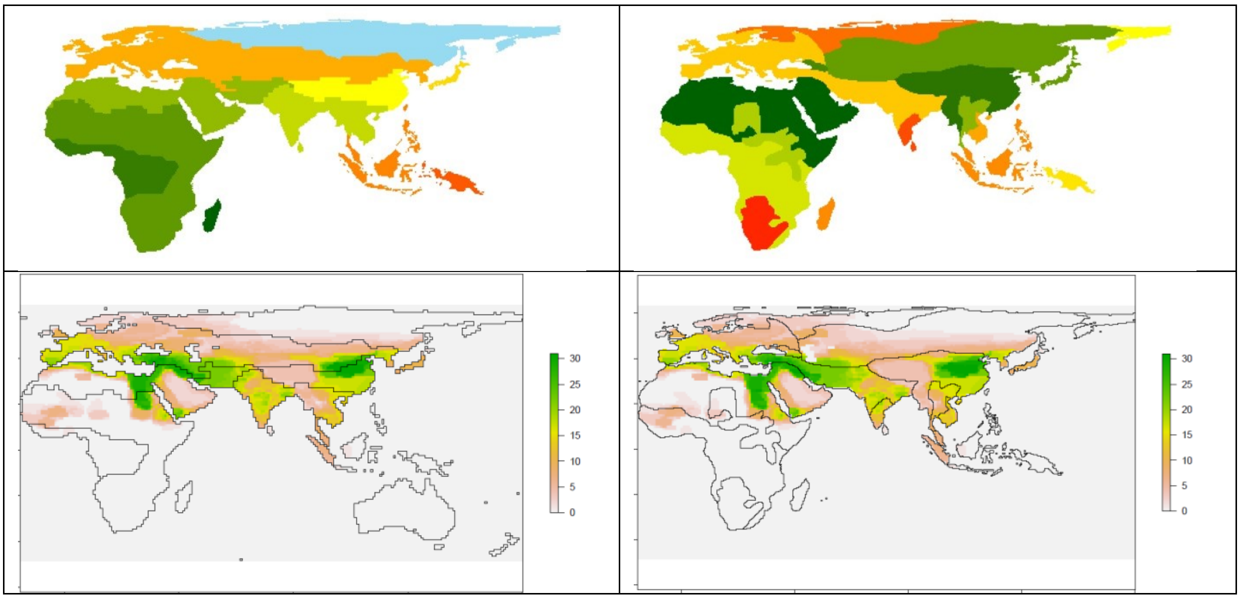


Figure S7. *Regions of ecological (top left) and cultural (top right) similarity. The polygons are overlaid on maps of imperial density (bottom). It can be seen that some zones contain regions of high imperial density, while others do not.*

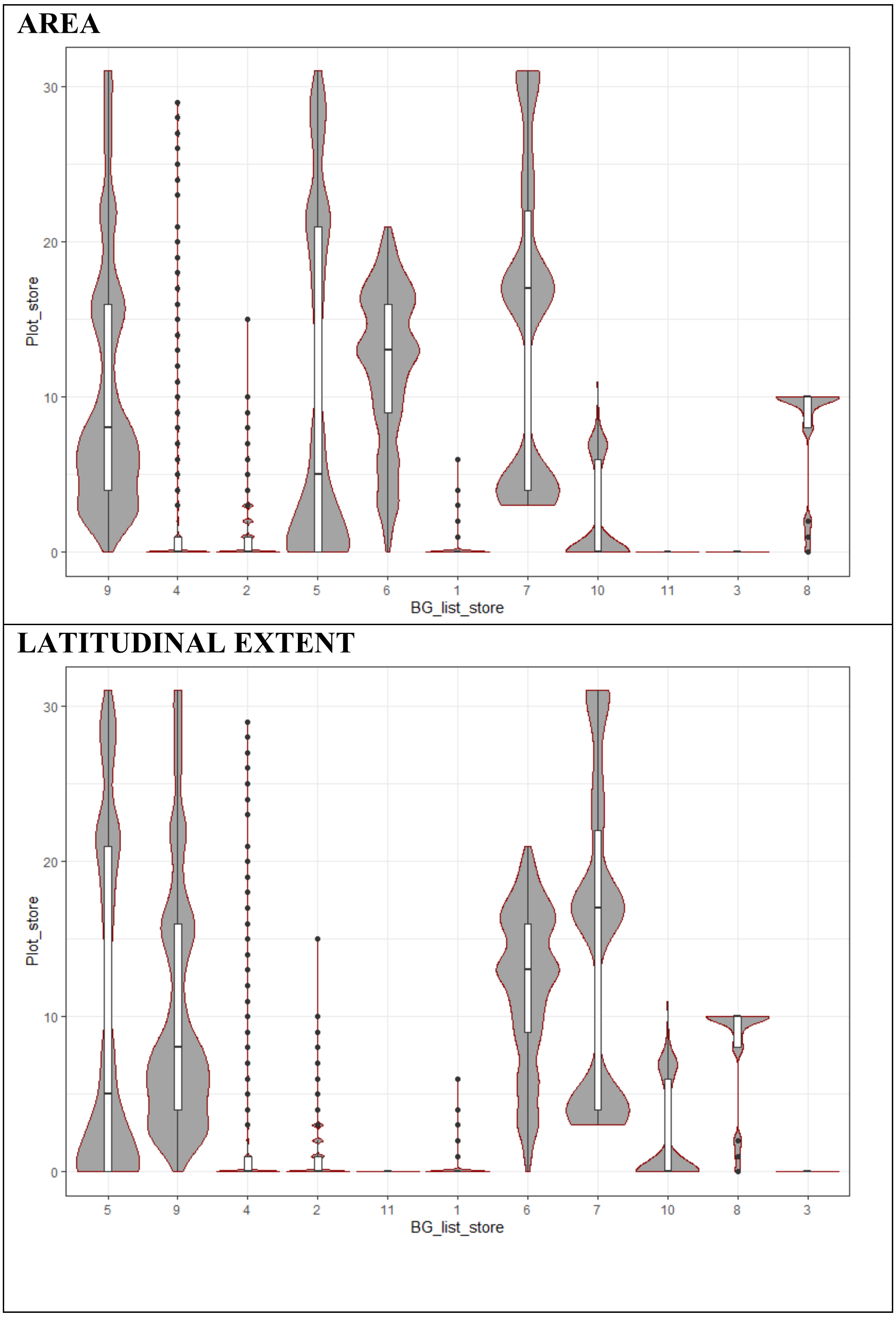


Figure S8. *Violin plots showing distribution of imperial density scores within Zoogeographic zones. Overlaid box plots show medians and 25th and 75th percentiles. Zones are ordered (left to right) by decreasing area (top) and decreasing latitudinal extent (bottom). There is no obvious relationship between these variables and imperial density.*

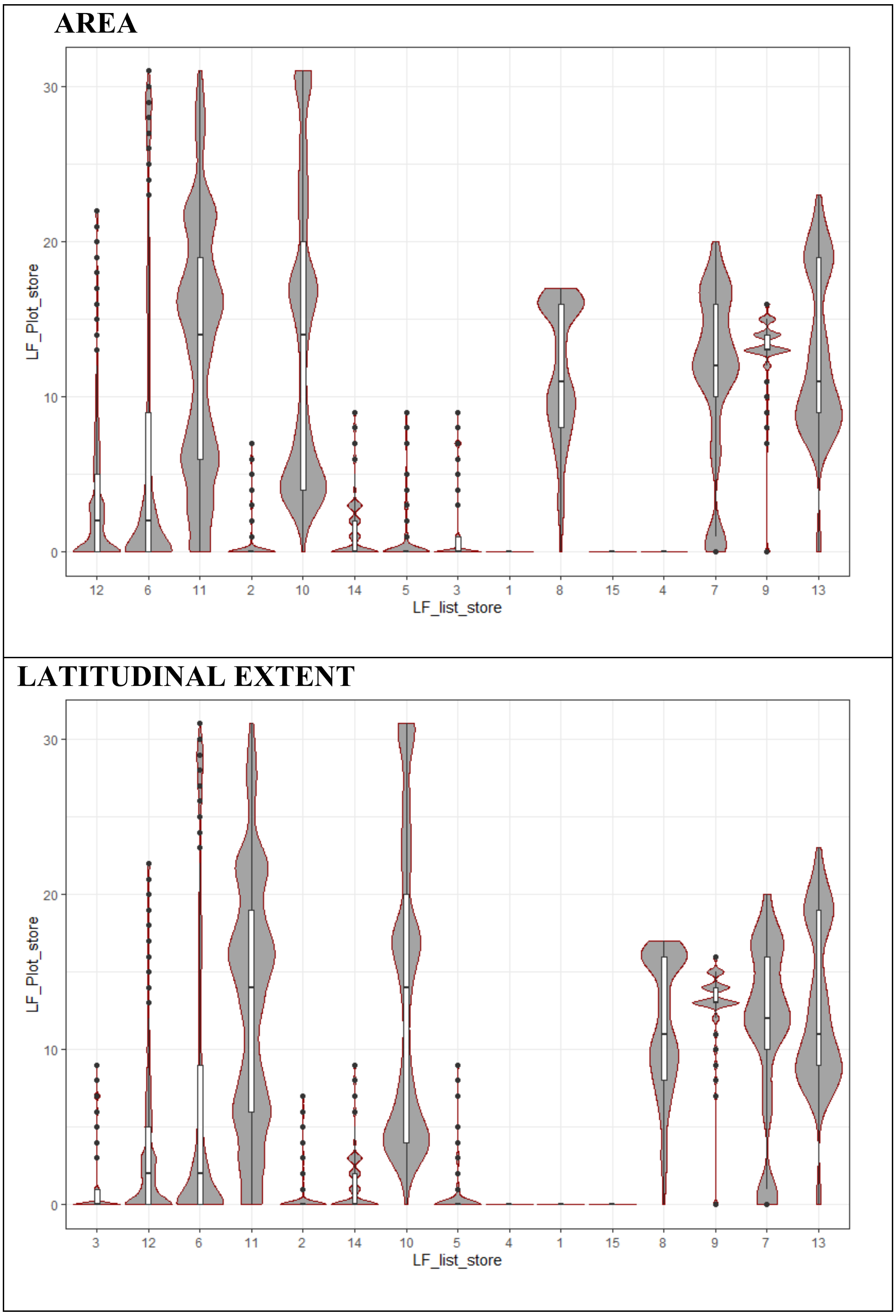


Figure S9. *Violin plots showing distribution of imperial density scores within language families. Overlaid box plots show medians and 25th and 75th percentiles. Families are ordered (left to right) by decreasing area (top) and decreasing latitudinal extent (bottom). There is no obvious relationship between these variables and imperial density.*

Table 8. *Summary data for different zoogeographic zones - Latitudinal extent, area, and mean imperial density*

Zoogeographic zone	Latitudinal extent (*km*)	Area (km^2)	Mean Imperial Density
African Rainforest	5117.3	4550000	0.12
Palearctic North	7169.4	16100000	0.87
Madagascan	651.2	590000	0.00
Afrotropical	7227.4	19510000	1.85
Saharo-Arabian	9520.8	10420000	10.27
Oriental	4713.7	5880000	12.06
Sino	4202.4	4440000	14.57
Japanese	1427.4	370000	8.14
Palearctic South	9397.5	19650000	10.34
SE Asia	3481.1	2160000	2.30
Oceanian	5487.1	890000	0.00

Table S9. *Summary data for different language families - Latitudinal extent, area, and mean imperial density*

Cultural zone (Language family)	Latitudinal extent (*km*)		Area (km^2)	Mean Imperial Density
	Min	Max		
Khoi-San	1937.5	1879.8	2965000	0.00
Niger-Congo	6424.3	6641.9	11122500	0.54
Austronesian	9021.1	9720.7	3050000	1.21
Papuan	2297.4	2266.3	902500	0.00
Nilo-Saharan	3088.0	3187.9	3732500	0.59
Afro-Asiatic	7819.9	7926.0	15350000	6.63
Dravidian	972.0	948.6	757500	11.47
Tai-Kadai	1321.6	1261.5	1317500	11.54
Austro-Asiatic	1021.5	987.5	572500	12.57
Sino-Tibetan	4415.7	4498.5	7445000	13.67
Indo-European	6912.4	10178.0	12940000	13.02
Altaic	8470.6	5232.0	19467500	3.61
Caucasian	777.6	829.7	302500	12.44
Uralic	4519.3	2310.3	5787500	1.46
Chukotko-Kamchatkan	1431.6	1016.0	1315000	0.00

Supplementary discussion

The Darwinian cultural evolutionary approach adopted in this paper helps to highlight that whilst being part of larger and more cohesive societies potentially provides benefits to individuals or groups, there are also costs associated with increased group size, including the time and energy required to communicate and coordinate actions over large distances, and the increased probability of free-riding in larger groups (*20, 21*). A particular issue is that of collective action problems which arise when individuals pay certain costs privately in order to produce a group-beneficial outcome (*22*). As individuals can benefit without contributing such situations provide an incentive to reduce effort or contribution to the public good. If too many individuals fail to contribute then the public good will not be produced or will be less effective and everybody will suffer (*20*). Cultural norms and institutions can provide solutions to such collective action problems (*23-25*). However, such institutions may also have their own costs. For example, members of society not involved in food production have to be supported through payment of tribute or taxes, whilst individuals established in positions of political authority can exploit their position and disproportionately benefit themselves.

In order for institutions to be favoured by selection they must on balance provided greater benefits than costs. These benefits may come from solving some ecological adaptive problem (games of "us-vs-nature") (*20*), e.g. pooling risk of crop failure, or returns to scale in the construction of irrigation systems create direct benefits. Benefits may also be evident due to competition with other groups (games of "us-vs-them"), where being larger and more cohesive can help groups win conflicts, representing a cultural form of multilevel selection (*26-28*). In this paper we have focussed on one case of us-vs-them explanations; the role of warfare intensity in providing a selective force that favoured the emergence of institutions that enabled groups to function on a massive scale. It seems unlikely that us-vs-nature explanations are plausible for the scale of society we are dealing with as solving problems of agricultural production, or ensuring access to resources typically involve smaller distances and can be solved at more local levels (*29, 30*). Although on a contemporary note the scale of certain ecological challenges we face today such as climate change or energy production may well be characterized as "games" of us-vs-nature. We also note that institutions can be established by elites that benefit themselves even at the expense of the majority of the population. It could be possible that elites sought to increase social scale in the face of hostility from other groups to protect their own interests and create situations in which they and their relatives and offspring benefited even if this was at a net cost to the rest of the population. While we cannot rule out this explanation with the current data, we note that large inequalities within societies and more extractive institutions are thought to be detrimental to overall group success for a number of reasons. First, they may not make the most of the talents of the majority of the population so may not perform as well economically and politically (*31*), which could lead to lower population growth or even decline as there might be fewer resources. Secondly, inequality may reduce social cohesion and make societies less effective in between group competition (*32*), or cause dissatisfied members of the population to leave or join other groups. In future work, we will use systematic historical data to assess between-group competition and the changing levels of inequality over time (*33*).

Here we have considered that variation in the onset of agriculture has enabled larger-scale societies to develop more in some regions because they have had longer to generate the kinds of institutions that such organization requires. Socio-political organization, like other aspects of human culture, is likely to proceed in a cumulative fashion (*34, 35*), with larger-scale societies building upon and elaborating from existing norms or institutions (*36, 37*). Related to this, another way that agriculture might affect socio-political evolution is that it created higher population densities, which is hypothesized to influence the rate at which cultural variants are created and retained within a population (*38-40*). However, institutions are a group-level property (*41*) and many may be established or imposed by ruling elites who make up a relatively small proportion of the population. Cumulative evolution of institutions may therefore be less affected by overall population size or density due to such a mechanism, although the distribution of populations in space could have important consequences for the kinds of institutions that need to be established in order to organize societies effectively. Another reason that duration of agriculture could be linked to the occurrence of large-scale societies is that the intensity of between-group competition may have been stronger in places in which groups with higher population densities were established earlier. However, this possibility is controlled for by including distance from first empires; in these confirmatory models duration of agriculture was still a substantial and reliable predictor of imperial density.

Our analyses do not show support for the potential productivity hypothesis. One potential limitation of the approach we have taken is that we have based our estimate on the FAO GAEZ crop model which is based on present-day conditions and inputs. However, potential productivity in the period we examine in this paper may have been affected by historical climate changes or other exogenous factors. At the current moment it is not possible to look at varying our measure of potential productivity over time in a meaningful way. It is not straightforward to include historical changes in climate into the estimates of potential productivity as scenarios with paleoclimatic data or other historical inputs are not available in the FAO GAEZ crop models. It is important to re-emphasize that the hypothesis we are testing relates to “potential” productivity (or a productivity endowment) not actual (or “achieved” productivity), and our measure attempts to hold constant factors like labour inputs and agricultural practices (which might be driven by social complexity). Attempting to use historical or archaeological data to show changes in productivity runs in to the problem that changes may well be due to climate or other exogenous factors (which would be relevant) but they may also be due to changes in cultivars, or agricultural practices and technologies, or myriad other things that would need to be assessed and held constant. Some of us are in the process of developing a method for a separate project (*42*) that attempts to create long time-depth estimates of productivity in past societies based on climate, crop type, and agricultural practices. However this approach currently only considers a very limited number of small geographically bounded regions so is not relevant to the scale of the current study. While it would be desirable in future to have estimates that take into account temporal changes we think that the data we are using at the moment is a fairly good proxy. Figure S3 shows the geographical distribution of the productivity variable used in this analysis. For the pattern in the productivity variable to be changed in a way that would make it a substantial predictor of imperial density would require quite substantial and geographically specific changes in climate or other exogenous variables.

Finally another consideration about the role of agriculture is that our warfare hypothesis assumes that pastoralists would be raiding agricultural regions that had a certain level of organization or achieved productivity. For example, one issue with our distance from the steppe variable is that all regions bordering the steppe may not have been targets for raiding pastoralists. Large societies did not start forming in Eastern Europe until later in the time span considered possibly due to the adoption of plough technologies that increased actual agricultural productivity (*28*). This might explain why the strength of the distance from steppe variable increases in the later periods in our time-varying analyses. Further tests that incorporated actual productivity and political development over time might prove insightful, although this would require careful consideration so as to avoid circular logic and spurious correlations. Such non-linear processes might be usefully assessed within a computational, agent-based modelling framework that incorporates these kinds of dynamic feedbacks (*28, 43-45*).

In this study we have assembled data from a variety of sources in order to test between various competing hypotheses. In going through this process various decisions and assumptions have to be made and it is important to recognize the limitations of the data and assess what affect these decisions might have on our findings. Our dependent variable of imperial density is made up of historical maps of the estimated extent of large societies. We have focussed only on the largest societies as these are more likely to have had their presence recorded in the archaeological and historical record. This is done to reduce the bias that exists in our knowledge about different parts of the world. One source of error with such data is that their course-grain nature does not allow us to distinguish between different forms of organization within these societies, or the extent to which different parts of these territories were important or directly controlled. This could create errors; for example, map territories may cover regions of high elevation or rough terrain even if those areas were peripheral to the main concerns and functions of the polity. This could explain why our elevation control variable does not show a strong relationship with imperial density. However, the issue does not appear to have generated any false relationships between imperial density and our main predictor variables. The fact that the main variables identified as either being good or not good predictors of imperial density remain qualitatively similar in the time series analyses (despite some interesting trends and quantitative changes in strength), and the fact that our main results are not sensitive to changes in the size threshold for inclusion in our measure of imperial density, suggest that our findings are reasonably robust to uncertainty in the measure of imperial density.

While there are sources of error in the data used in this study, as there will be in any comparative analysis, there does not appear to be reasons to suspect that these errors are due to systematic biases that may have favoured some of our main hypotheses over others. Even with these limitations it is important to attempt to systematically address hypotheses in this way, and highlight where future data collection efforts should be targeted. By going through such a process we help to uncover some of the hidden assumptions in ideas that have often only been presented as informal verbal arguments.

Supplementary References

1. T. E. Currie *et al.*, Agricultural productivity in past societies: Toward an empirically informed model for testing cultural evolutionary hypotheses. *Cliodynamics: The Journal of Quantitative History and Cultural Evolution* **6**, (2015).
2. FAO. (2012).
3. D. M. Olson *et al.*, Terrestrial ecoregions of the worlds: A new map of life on Earth. *Bioscience* **51**, 933 (2001).
4. USGS. (Data available from the U.S. Geological Survey, 1993).
5. J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, R. C. Team. (R package, 2015).
6. R Core Team. (Vienna, Austria, 2015).
7. J. C. Pinheiro, D. Bates, *Mixed-Effects Models in S and S-PLUS*. (Springer, 2009).
8. K. P. Burnham, D. R. Anderson, *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach*. (Springer, 2002).
9. H. Akaike, New Look at Statistical-Model Identification. *Ieee Transactions on Automatic Control* **Ac19**, 716 (1974).
10. T. F. Rangel, J. A. F. Diniz, L. M. Bini, SAM: a comprehensive application for Spatial Analysis in Macroecology. *Ecography* **33**, 46 (Feb, 2010).
11. B. G. Holt *et al.*, An Update of Wallace's Zoogeographic Regions of the World. *Science* **339**, 74 (January 4, 2013, 2013).
12. J. Diamond, P. Bellwood, Farmers and their languages: The first expansions. *Science* **300**, 597 (2003).
13. T. E. Currie, A. Meade, M. Guillon, R. Mace, Cultural phylogeography of the Bantu Languages of sub-Saharan Africa. *Proceedings of the Royal Society B: Biological Sciences* **280**, (July 7, 2013, 2013).
14. W. A. Foley, The languages of New Guinea. *Annual Review of Anthropology* **29**, 357 (2000).
15. A. Vovin, Japanese, Korean, and Other 'Non-Altaic' Languages. *Central Asiatic Journal* **53**, 105 (2009).
16. R. A. Miller, I. f. s. kulturforskning, *Languages and History: Japanese, Korean, and Altaic*. (Institute for Comparative Research in Human Culture, 1996).
17. J. Diamond, *Guns, germs and Steel*. (Vintage, London, 1997).
18. H. Kreft, W. Jetz, Comment on "An Update of Wallace's Zoogeographic Regions of the World". *Science* **341**, 343 (2013-07-26 00:00:00, 2013).
19. R. Bouckaert *et al.*, Mapping the Origins and Expansion of the Indo-European Language Family. *Science* **337**, 957 (August 24, 2012, 2012).
20. S. Gavrilets, Collective action problem in heterogeneous groups. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **370**, (2015-12-05 00:00:00, 2015).
21. P. Turchin, in *Cultural Evolution,* P. J. Richerson, Morten H. Christiansen, Eds. (MIT Press, Cambridge, MA, 2013), vol. 12, pp. 61-73.
22. M. Olson, *The Logic of Collective Action: Public Goods and the Theory of Groups*. (Harvard University Press, Cambridge, MA, 1965).
23. P. Richerson, J. Henrich, Tribal Social Instincts and the Cultural Evolution of Institutions to Solve Collective Action Problems. *Cliodynamics* **3**, 38 (2012).
24. T. E. Currie *et al.*, in *Complexity and Evolution: A New Synthesis for Economics,* D. S. Wilson, A. Kirkman, E. Beinhoffer, Eds. (MIT Press, Ernst Strüngmann Forum, Cambridge, 2016).
25. S. T. Powers, C. P. van Schaik, L. Lehmann, How institutions shaped the last major evolutionary transition to large-scale human societies. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **371**, (2016).
26. M. R. Zefferman, S. Mathew, An evolutionary theory of large-scale human warfare: Group-structured cultural selection. *Evolutionary Anthropology: Issues, News, and Reviews* **24**, 50 (2015).

27. T. E. Currie, R. Mace, Political complexity predicts the spread of ethnolinguistic groups. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 7339 (May 5, 2009).
28. P. Turchin, T. E. Currie, E. A. L. Turner, S. Gavrilets, War, space, and the evolution of Old World complex societies. *Proceedings of the National Academy of Sciences* **110**, 16384 (October 8, 2013, 2013).
29. T. Earle, *Economic and social organization of a complex chiefdom : the Halelea District, Kauai, Hawaii*. (Museum of Anthropology, University of Michigan, Ann Arbor, 1978).
30. P. V. Kirch, *The Evolution of the Polynesian Chiefdoms*. New Studies in Archaeology (Cambridge University Press, Cambridge, 1984).
31. D. Acemoglu, J. A. Robinson, *Why Nations Fail: The Origins of Power, Proseprity, and Poverty*. (Crown Publishing Group, New York, 2012).
32. P. Turchin, *Historical dynamics: why states rise and fall*. (Princeton University Press, Princeton, NJ, 2003).
33. P. Turchin *et al.*, Seshat: The Global History Databank. *Cliodynamics: The Journal of Quantitative History and Cultural Evolution* **6**, (2015).
34. M. Kempe, S. J. Lycett, A. Mesoudi, From cultural traditions to cumulative culture: parameterizing the differences between human and nonhuman culture. *Journal of Theoretical Biology* **359**, 29 (10/21/, 2014).
35. E. A. Smith, Endless forms: human behavioural diversity and evolved universals. *Philosophical Transactions of the Royal Society B: Biological Sciences* **366**, 325 (February 12, 2011, 2011).
36. T. E. Currie, R. Mace, Mode and tempo in the evolution of socio-political organization: reconciling "Darwinian" and "Spencerian" evolutionary approaches in anthropology. *Philosophical Transactions of the Royal Society B: Biological Sciences* **366**, 1108 (April 12, 2011, 2011).
37. T. E. Currie, S. J. Greenhill, R. D. Gray, T. Hasegawa, R. Mace, Rise and fall of political complexity in island South-East Asia and the Pacific. *Nature* **467**, 801 (2010).
38. J. Henrich, Demography and cultural evolution: How adaptive cultural processes can produce maladaptive losses - The Tasmanian case. *American Antiquity* **69**, 197 (Apr, 2004).
39. A. Powell, S. Shennan, M. G. Thomas, Late Pleistocene Demography and the Appearance of Modern Human Behavior. *Science* **324**, 1298 (Jun 5, 2009).
40. P. E. Smaldino, P. J. Richerson, in *Handbook of Human Computation,* P. Michelucci, Ed. (Springer New York, New York, NY, 2013), pp. 979-992.
41. P. E. Smaldino, The cultural evolution of emergent group-level traits. *Behavioral and Brain Sciences* **37**, 243 (2014).
42. P. Turchin *et al.*, Crop Productivity Estimates for Past Societies in the World Sample-30 of Seshat: Global History Databank. *SocArXiv*, (2019).
43. P. Gerbault *et al.*, Storytelling and story testing in domestication. *Proceedings of the National Academy of Sciences* **111**, 6159 (April 29, 2014, 2014).
44. A. Griffin, C. Stanish, An agent-based model of prehistoric settlement patterns and political consolidation in the Lake Titicaca basin of Peru and Bolivia. *Structure and Dynamics* **2**, (2007).
45. C. Cioffi-Revilla, *Introduction to Computational Social Science: Principles and Applications*. (Springer London, 2013).

APPENDIX I

Polities greater than 100,000 km^2 used in creating the imperial density score listed by century time slice.

Century	Polity Name	Area km^2	Latitude	Longitude
1500 BCE	Babylon (Kassite)	136614	31.72	47.29
1500 BCE	Egypt (New Kingdom)	1525448	25.03	31.74
1500 BCE	Elam	125602	29.97	52.23
1500 BCE	Hittite (Old)	230343	38.43	35.44
1500 BCE	Mittani	114580	36.55	42.33
1500 BCE	Punt	112683	12.46	41.11
1500 BCE	Shang	672578	34.58	110.67
Century	**Polity Name**	**Area km^2**	**Latitude**	**Longitude**
1400 BCE	Babylon (Kassite)	194979	31.71	46.58
1400 BCE	Egypt (New Kingdom)	1502371	25.09	31.74
1400 BCE	Elam (Old)	130495	30.03	52.30
1400 BCE	Hittite (New)	135304	39.55	35.30
1400 BCE	Mittani	432811	36.80	40.90
1400 BCE	Punt	113086	12.59	41.12
1400 BCE	Shang	699923	34.62	110.66
Century	**Polity Name**	**Area km^2**	**Latitude**	**Longitude**
1300 BCE	Babylon (Kassite)	247822	32.04	46.50
1300 BCE	Egypt (New Kingdom)	1478626	24.96	31.64
1300 BCE	Elamite (Middle, Dur-Untash)	126849	30.02	52.32
1300 BCE	Hittite (New)	586141	37.74	37.13
1300 BCE	Punt	104682	12.52	41.11
1300 BCE	Shang	674357	34.59	110.77
Century	**Polity Name**	**Area km^2**	**Latitude**	**Longitude**
1200 BCE	Assyria	229406	35.40	44.94
1200 BCE	Babylon (Kassite)	221376	31.63	46.60
1200 BCE	Egypt (New Kingdom)	1475089	24.83	31.65
1200 BCE	Elamite (Middle, Dur-Untash)	116324	30.02	52.24
1200 BCE	Hittite (New)	541223	37.93	36.48
1200 BCE	Punt	121503	12.50	41.06
1200 BCE	Shang	702464	34.60	111.01
Century	**Polity Name**	**Area km^2**	**Latitude**	**Longitude**
1100 BCE	Assyria	244280	36.24	43.16
1100 BCE	Babylon - Aramaean and Chaldean	221796	31.84	46.53
1100 BCE	Egypt (New Kingdom)	1471171	24.82	31.59
1100 BCE	Elamite (Middle, Dur-Untash)	122962	30.05	52.20

1100 BCE	Hittite (Neo)	147666	37.32	36.93
1100 BCE	Saba	101056	14.89	44.05
1100 BCE	Western Zhou	965058	35.17	112.23
Century	**Polity Name**	**Area km^2**	**Latitude**	**Longitude**
1000 BCE	Babylon - Aramaean and Chaldean	194962	31.63	46.60
1000 BCE	Hittite (Neo)	148330	37.38	36.95
1000 BCE	Kush	463656	20.05	31.22
1000 BCE	Saba	105174	14.96	44.19
1000 BCE	Western Zhou	981610	35.20	112.16
Century	**Polity Name**	**Area km^2**	**Latitude**	**Longitude**
900 BCE	Assyria (Neo)	206106	35.17	43.76
900 BCE	Ba	158561	31.70	104.44
900 BCE	Babylon - Aramaean and Chaldean	125148	32.07	46.43
900 BCE	Egypt	321581	27.73	29.61
900 BCE	Hittite (Neo)	140939	37.37	36.81
900 BCE	Kush	454227	20.05	31.25
900 BCE	Phyrgia	131706	39.74	33.22
900 BCE	Saba	100934	14.87	44.13
900 BCE	Western Zhou	998493	35.22	112.14
Century	**Polity Name**	**Area km^2**	**Latitude**	**Longitude**
800 BCE	Assyria (Neo)	386439	34.90	44.07
800 BCE	Ba	166673	31.84	104.42
800 BCE	Elamite (Neo)	122060	30.01	52.12
800 BCE	Hittite (Neo)	145219	37.41	36.79
800 BCE	Kush	476838	19.98	31.26
800 BCE	Phyrgia	167752	39.47	33.17
800 BCE	Saba	101839	14.82	44.09
800 BCE	Urartu	102006	39.16	42.70
800 BCE	Western Zhou	993463	35.22	112.17
Century	**Polity Name**	**Area km^2**	**Latitude**	**Longitude**
700 BCE	Assyria (Neo)	702715	33.63	43.17
700 BCE	Ba	185725	31.90	104.29
700 BCE	Eastern Zhou	1490577	35.89	113.42
700 BCE	Elamite (Neo)	116061	29.98	52.20
700 BCE	Hittite (Neo)	130190	37.29	37.01
700 BCE	Kush (25th, Kushite)	823936	23.32	30.48
700 BCE	Median	128745	34.16	51.40
700 BCE	Phyrgia	180409	39.51	33.45
700 BCE	Saba	106074	14.95	44.02
700 BCE	Urartu (Lake Van)	252258	39.06	43.08

Century	Polity Name	Area km^2	Latitude	Longitude
600 BCE	Ba	183418	31.86	104.30
600 BCE	Babylonian (Neo)	654601	34.37	42.26
600 BCE	Eastern Zhou	1497757	35.89	113.48
600 BCE	Egypt	626022	27.00	29.77
600 BCE	Lydian (Mermnadae)	268569	38.91	30.31
600 BCE	Median	2618859	33.82	60.63
600 BCE	Napata - Kush	358958	19.30	31.19
600 BCE	Saba	108806	15.06	44.22
600 BCE	Urartu (Lake Van)	182492	39.43	42.49

Century	Polity Name	Area km^2	Latitude	Longitude
500 BCE	Ba	165177	31.92	104.27
500 BCE	Carthagian	174571	37.53	11.17
500 BCE	Eastern Zhou	1494083	35.85	113.43
500 BCE	Garamantes	121221	30.32	7.37
500 BCE	Magadha	310988	22.77	86.02
500 BCE	Meroe/Kush	348352	19.21	31.32
500 BCE	Persian	6184635	34.89	35.99
500 BCE	Qeder	121413	29.65	35.66
500 BCE	Saba	243232	14.54	42.09
500 BCE	Tartessian	172395	36.04	-3.38

Century	Polity Name	Area km^2	Latitude	Longitude
400 BCE	Ba	173259	31.90	104.27
400 BCE	Carthagian	166617	37.60	11.12
400 BCE	Eastern Zhou	1469842	35.85	113.46
400 BCE	Egypt - 28th	698801	26.49	30.36
400 BCE	Garamantes	124825	30.23	7.43
400 BCE	Kalinga	108850	19.16	83.42
400 BCE	Magadha	706026	22.49	82.23
400 BCE	Meroe/Kush	363101	19.33	31.28
400 BCE	Persian	5268224	35.16	43.96
400 BCE	Qeder	116302	29.63	35.58
400 BCE	Saba	232718	14.53	42.13
400 BCE	Tartessian	158987	36.05	-3.37

Century	Polity Name	Area km^2	Latitude	Longitude
300 BCE	Armenia (Orontids)	161485	39.96	41.63
300 BCE	Carthagian	269765	38.45	7.05
300 BCE	Eastern Zhou	1650736	35.52	112.39
300 BCE	Garamantes	115997	30.22	7.37
300 BCE	Kalinga	111454	19.08	83.44

300 BCE	Lysimachus	469913	40.67	29.12
300 BCE	Mauryan	2655315	27.31	74.39
300 BCE	Meroe	336613	19.35	31.18
300 BCE	Min-Yue	205611	26.03	118.09
300 BCE	Nabatea	119330	29.66	35.54
300 BCE	Ptolemaic	953782	27.55	28.91
300 BCE	Saba	255638	15.10	45.99
300 BCE	Seleucid	3155275	34.91	43.46

Century	**Polity Name**	**Area km^2**	**Latitude**	**Longitude**
200 BCE	Armenia	148777	39.93	41.51
200 BCE	Attalid	204334	38.54	29.71
200 BCE	Bactria (Hellenistic)	754599	36.38	67.26
200 BCE	Early Han	3129366	33.33	111.23
200 BCE	Garamantes	123241	30.25	7.55
200 BCE	Mauryan	2769903	22.39	77.11
200 BCE	Meroe	349143	19.31	31.29
200 BCE	Min-Yue	197834	26.04	118.19
200 BCE	Nabatea	118373	29.65	35.65
200 BCE	Numantia (Berber)	208530	35.56	3.81
200 BCE	Parthia	192487	37.99	57.78
200 BCE	Ptolemaic	943740	27.61	28.85
200 BCE	Rome	373531	39.99	7.93
200 BCE	Saba	276889	15.07	45.92
200 BCE	Seleucid	2474354	34.34	41.72
200 BCE	Xiongnu (Hsiung-Nu)	990846	43.65	110.30

Century	**Polity Name**	**Area km^2**	**Latitude**	**Longitude**
100 BCE	Andhra (Satavahana/Satakani)	395580	18.08	78.72
100 BCE	Armenia	200741	39.85	42.98
100 BCE	Early Han	4265049	31.58	109.85
100 BCE	Garamantes	119805	30.21	7.46
100 BCE	Indo-Greek	418533	32.20	71.14
100 BCE	Indo-Scythian (Sakas)	909697	28.89	65.14
100 BCE	Maha-Megha-Vahana (Kalinga)	111313	19.28	83.93
100 BCE	Mauretania	153348	35.18	-0.47
100 BCE	Meroe	346901	19.27	31.28
100 BCE	Min-Yue	174847	26.27	118.26
100 BCE	Numidia	245886	33.67	12.74
100 BCE	Parthia	1814838	33.46	53.69
100 BCE	Petra (Nabatea)	314923	29.91	37.03
100 BCE	Pontus	172483	44.33	37.61
100 BCE	Ptolemaic	773463	30.88	31.57
100 BCE	Rome	1158984	39.27	10.28

100 BCE	Saba	120374	15.36	44.53
100 BCE	Seleucid	111090	36.11	36.56
100 BCE	Shunga (Maghada)	485953	23.04	84.83
100 BCE	Xiongnu (Hsiung-Nu)	4151037	45.80	100.02
Century	**Polity Name**	**Area km^2**	**Latitude**	**Longitude**
0 CE	Andhra (Satavahana/Satakani)	850366	17.88	76.82
0 CE	Axum	208819	14.13	38.18
0 CE	Han	4714747	31.92	107.63
0 CE	Indo-Parthia	1193069	30.08	67.64
0 CE	Koguryo	148387	41.34	127.35
0 CE	Maha-Megha-Vahana (Kalinga)	538783	19.43	83.61
0 CE	Meroe	344792	19.31	31.12
0 CE	Parthia	2379607	33.95	54.38
0 CE	Petra (Nabatea)	286356	29.55	36.75
0 CE	Rome	4437227	38.49	16.86
Century	**Polity Name**	**Area km^2**	**Latitude**	**Longitude**
100 CE	Andhra (Satavahana/Satakani)	965373	17.55	77.90
100 CE	Axum	279347	14.70	37.87
100 CE	Dacia	101853	44.52	24.33
100 CE	Funan	207090	11.67	104.59
100 CE	Han	4698961	25.82	108.56
100 CE	Indo-Scythian	520742	24.24	74.31
100 CE	Koguryo	147219	41.34	127.29
100 CE	Kushans	2794507	33.38	69.87
100 CE	Maha-Megha-Vahana (Kalinga)	175692	20.86	85.18
100 CE	Meroe	330618	19.29	30.97
100 CE	Parthia	2174577	33.26	53.63
100 CE	Petra - Nabatea	289999	29.69	36.85
100 CE	Qataban	105704	14.90	47.58
100 CE	Rome	4821880	39.92	15.01
Century	**Polity Name**	**Area km^2**	**Latitude**	**Longitude**
200 CE	Axum	262769	14.75	37.82
200 CE	Champa	122528	12.75	108.16
200 CE	Funan	211217	11.61	104.56
200 CE	Han	4074265	25.07	110.03
200 CE	Koguryo	153689	41.40	127.37
200 CE	Kushans	2288774	34.62	67.70
200 CE	Maha-Megha-Vahana (Kalinga)	167108	20.78	85.15
200 CE	Meroe	320640	19.31	31.03
200 CE	Parthia	2033336	33.20	54.07
200 CE	Rome	5138113	39.93	15.13

200 CE	Satavahanihara (Andhra)	963961	17.59	77.91
200 CE	Western Satraps	314068	23.18	73.08
Century	**Polity Name**	**Area km^2**	**Latitude**	**Longitude**
300 CE	Axum	301493	14.80	37.92
300 CE	Champa	113904	14.84	107.86
300 CE	Funan	362743	12.40	104.29
300 CE	Himyarite	296205	15.21	46.43
300 CE	Jin	4357531	25.43	109.41
300 CE	Koguryo	158645	41.50	127.06
300 CE	Maha-Megha-Vahana (Kalinga)	172438	20.93	84.99
300 CE	Meroe	321511	19.39	30.74
300 CE	Rome	5068209	39.99	14.93
300 CE	Salankayana of Vengi	153318	14.20	79.05
300 CE	Sassanians	2826885	32.30	57.48
300 CE	Vakataka	566469	20.19	78.75
300 CE	Western Satraps	297054	23.47	73.00
Century	**Polity Name**	**Area km^2**	**Latitude**	**Longitude**
400 CE	Axum	297497	14.77	37.96
400 CE	Chen-La / Funan	404564	12.68	103.78
400 CE	Gupta	632465	24.43	83.46
400 CE	Himyarite	305578	15.12	46.65
400 CE	Hunnic	144116	48.64	34.60
400 CE	Jin	3151290	23.97	109.77
400 CE	Koguryo	284396	41.02	127.47
400 CE	Nobatia	129652	21.07	29.72
400 CE	Rome	4838186	39.82	14.91
400 CE	Sassanians	3079341	32.83	56.80
400 CE	Vakataka	531982	19.40	78.27
Century	**Polity Name**	**Area km^2**	**Latitude**	**Longitude**
500 CE	Axum	314478	14.90	37.92
500 CE	Byzantine Empire	2598437	36.64	27.38
500 CE	Chen-La	529913	13.13	104.09
500 CE	Gupta	627182	24.50	83.54
500 CE	Himyarite	568446	16.29	46.98
500 CE	Kalabhras (Tamil)	132814	10.34	78.12
500 CE	Koguryo	360285	40.66	126.64
500 CE	Maitraka	166174	22.49	72.48
500 CE	Merovingian	458082	49.03	5.95
500 CE	Nobatia	160076	20.94	29.80
500 CE	Ostrogoth	452413	41.10	11.80
500 CE	Rai	257318	26.73	68.16

500 CE	Sassanians	2821227	32.76	54.81
500 CE	Toba Wei	1619624	36.91	108.33
500 CE	Vandal	229974	37.44	5.61
500 CE	Vishnukundins	753482	19.08	77.94
500 CE	Visigoth	730537	41.73	-1.42
Century	**Polity Name**	**Area km^2**	**Latitude**	**Longitude**
600 CE	Avars	873194	47.96	25.44
600 CE	Axum	340662	14.75	38.13
600 CE	Byzantine Empire	3401431	38.78	15.44
600 CE	Chalukyas (Vatapi)	246809	17.11	76.38
600 CE	Chen-La	381999	12.96	105.61
600 CE	Davarati	128305	14.52	100.46
600 CE	Gupta	146722	23.97	80.02
600 CE	Kabul Shahi	135858	32.68	71.78
600 CE	Khazars	248606	48.11	46.05
600 CE	Koguryo	315028	41.34	126.19
600 CE	Lombard	144626	45.60	10.84
600 CE	Maitraka	159860	22.57	72.35
600 CE	Makkura	251070	18.66	29.98
600 CE	Maukhari	401274	24.49	84.35
600 CE	Meovingian	745824	47.75	4.83
600 CE	Nan Chao	555075	23.67	103.39
600 CE	Nobatia	152796	21.94	28.58
600 CE	Rai	236934	26.74	68.13
600 CE	Sassanians	2447848	33.01	55.13
600 CE	Sui	3959502	25.51	109.65
600 CE	Visigoth	513238	40.29	-0.35
600 CE	Yamato	187763	34.66	134.95
Century	**Polity Name**	**Area km^2**	**Latitude**	**Longitude**
700 CE	Alwa	358983	14.56	30.71
700 CE	Aquitaine	148767	45.11	1.27
700 CE	Avars	120004	47.07	21.06
700 CE	Axum	364955	14.59	38.13
700 CE	Bulgar (Turkic)	181513	44.46	25.70
700 CE	Byzantine	826089	38.89	17.75
700 CE	Chalukyas	506776	18.92	76.41
700 CE	Chen-La	460544	13.42	105.00
700 CE	Ghana	278689	14.95	-3.44
700 CE	Gupta	322539	23.83	86.88
700 CE	Kabul Shahi	128288	33.01	72.27
700 CE	Khazars	785402	46.55	43.60
700 CE	Lombard	182700	45.37	10.94

700 CE	Maitraka	208621	22.94	73.33
700 CE	Makkura	561066	20.40	28.70
700 CE	Merovingian	606331	48.66	5.97
700 CE	Nan Chao	418582	23.98	103.69
700 CE	Parhae	294118	41.57	126.31
700 CE	Silla	133337	36.70	127.73
700 CE	Srivijaya	455625	1.91	101.97
700 CE	Tang	4358544	26.03	108.46
700 CE	Tibet	2704938	32.78	89.01
700 CE	Umayyad	7160599	29.08	45.39
700 CE	Visigoth	599111	40.37	-3.79
700 CE	Volga Bulgars	361952	53.65	51.63
700 CE	Yamato	207470	34.88	135.22
Century	**Polity Name**	**Area km^2**	**Latitude**	**Longitude**
800 CE	Abbasid	8368358	30.07	48.00
800 CE	Alwa	347602	14.73	30.76
800 CE	Asturias	126850	42.57	-5.61
800 CE	Axum	397354	14.46	38.13
800 CE	Bulgar (Turkic)	207157	44.86	26.08
800 CE	Byzantine	828957	38.13	23.03
800 CE	Carolingian	1093376	44.73	7.90
800 CE	Champa	103918	14.20	108.22
800 CE	Chen-La	241139	15.39	104.30
800 CE	Fujiwara	249132	34.14	133.81
800 CE	Ghana	286553	14.99	-3.37
800 CE	Gurjara-Pratiharas	409428	25.79	73.85
800 CE	Kabul Shahi	357703	33.09	71.80
800 CE	Khazars	861030	46.85	44.19
800 CE	Makkura	513690	20.36	28.63
800 CE	Nan Chao	508705	23.93	101.21
800 CE	Pagan	110638	22.54	94.54
800 CE	Palas	523034	24.79	84.21
800 CE	Parhae	292545	41.81	126.68
800 CE	Rashtrakuta	1051033	18.88	77.70
800 CE	Rustamid	145653	35.04	4.38
800 CE	Silla	141244	36.96	127.71
800 CE	Srivijaya	602953	-0.01	104.59
800 CE	Tang	3648413	25.19	110.25
800 CE	Tibet	3879657	34.00	87.22
800 CE	Umayyad (Cordoba)	460079	39.63	-3.74
800 CE	Volga bulgars	344540	53.64	51.52
Century	**Polity Name**	**Area km^2**	**Latitude**	**Longitude**

900 CE	Abbasid	1028769	36.41	45.66
900 CE	Aghlabid	106966	35.64	9.24
900 CE	Alwa	377707	14.76	31.02
900 CE	Asturias	131577	42.51	-5.62
900 CE	Axum	124211	14.66	38.61
900 CE	Bulgar (Turkic)	264073	43.94	25.00
900 CE	Byzantine	778988	38.64	21.36
900 CE	Carolingian	891625	48.16	5.75
900 CE	Fujiwara	265285	34.09	133.80
900 CE	Ghana	276388	14.92	-3.32
900 CE	Gurjara-Pratiharas	302924	23.75	73.56
900 CE	Hindu Shahi	281138	30.80	74.79
900 CE	Idriscid	128367	35.14	-0.67
900 CE	Khazars	805496	46.67	43.96
900 CE	Khmer	477521	13.47	105.05
900 CE	Makkura	504735	20.38	28.53
900 CE	Nan Chao	500631	23.93	101.14
900 CE	Pagan	113506	22.54	94.57
900 CE	Palas	513995	24.82	84.53
900 CE	Parhae	533808	43.42	129.76
900 CE	Rashtrakuta	764788	18.88	77.12
900 CE	Rustamid	377955	33.66	7.82
900 CE	Saffarid	2471262	32.42	60.88
900 CE	Samanid	645407	41.11	66.83
900 CE	Srivijaya	642992	0.60	103.62
900 CE	Tang	3651559	25.21	110.26
900 CE	Tulunid	1394687	28.11	33.74
900 CE	Umayyad (Cordoba)	436937	39.55	-0.35
900 CE	Volga Bulgars	343548	53.70	51.50
900 CE	Zaydite	116715	14.08	45.51
Century	**Polity Name**	**Area km^2**	**Latitude**	**Longitude**
1000 CE	Abbasid (Buwayhids)	1793726	28.30	53.56
1000 CE	Alwa	309956	14.74	29.75
1000 CE	Axum	251654	13.97	37.75
1000 CE	Bulgar (Turkic)	355101	43.60	23.61
1000 CE	Byzantine	810970	38.09	23.02
1000 CE	Capetian	446124	46.70	1.89
1000 CE	Chalukyas (Western, Later)	640066	18.34	76.53
1000 CE	Cholas (Tamil)	273800	10.29	79.45
1000 CE	Dai Viet	117383	20.25	105.80
1000 CE	Davarati	154953	13.27	100.11
1000 CE	Eastern Ganga (Kalinga)	123961	19.77	84.61
1000 CE	England	120855	52.50	-1.28

1000 CE	Fatimid	1429892	27.68	32.32
1000 CE	Fujiwara	252570	34.12	133.73
1000 CE	Ghana	280146	15.01	-3.33
1000 CE	Ghaznavid (Afghan)	1306966	33.24	64.12
1000 CE	Gotaland	112095	58.43	15.64
1000 CE	Gurjara-Pratiharas	381577	27.36	80.36
1000 CE	Hindu Shahi	168849	31.30	74.32
1000 CE	Holy Roman	730890	45.31	9.99
1000 CE	Khazars	393857	46.32	44.93
1000 CE	Khmer	571667	14.34	104.82
1000 CE	Kievan Rus	1036870	54.04	31.04
1000 CE	Koryo	138818	36.85	127.75
1000 CE	Leon	115165	42.36	-5.93
1000 CE	Liao	4455351	45.67	106.06
1000 CE	Magyars	243451	47.20	20.86
1000 CE	Makkura	480946	20.44	29.36
1000 CE	Mali	144672	10.81	-7.10
1000 CE	Nan Chao	575807	23.26	102.53
1000 CE	Pagan	118265	21.24	95.36
1000 CE	Paramaras	126258	24.51	76.85
1000 CE	Poland (Piast)	322735	51.69	18.54
1000 CE	Saffarid	260919	26.92	63.50
1000 CE	Samanid	265195	39.70	58.40
1000 CE	Song (Northern)	3061308	24.77	110.53
1000 CE	Srivijaya	792453	-0.58	105.17
1000 CE	Umayyad (Cordoba)	560818	38.02	-0.74
1000 CE	Volga Bulgars	334892	53.66	51.64
1000 CE	Xixia	635484	38.06	99.53
1000 CE	Zaydite	326261	15.65	45.06
1000 CE	Zirid	440008	33.83	10.12
Century	**Polity Name**	**Area km^2**	**Latitude**	**Longitude**
1100 CE	Almoravid	453358	36.80	-4.11
1100 CE	Alwa	454817	15.10	30.10
1100 CE	Byzantine	865582	37.99	27.23
1100 CE	Capetian	426718	46.84	1.82
1100 CE	Castille	148735	42.32	-5.56
1100 CE	Chalukyas (Western, Later)	653231	18.27	76.54
1100 CE	Cholas (Tamil)	195138	12.76	78.69
1100 CE	Danishment	107062	39.28	37.03
1100 CE	Eastern Ganga (Kalinga)	116579	19.81	84.72
1100 CE	England	143325	52.47	-1.61
1100 CE	Fatimid	1173325	26.74	33.80
1100 CE	Fujiwara	236815	34.10	133.74

1100 CE	Ghaznavid (Afghan)	1592174	29.59	69.43
1100 CE	Gotaland	112332	58.47	15.61
1100 CE	Great Seljuk	3352648	36.44	58.51
1100 CE	Gurjara-Pratiharas	194885	23.71	73.24
1100 CE	Holy Roman	824497	48.35	10.31
1100 CE	Khmer	790763	13.97	103.44
1100 CE	Kievan Rus	1022775	54.01	30.87
1100 CE	Koryo	148683	37.33	127.43
1100 CE	Liao	2454089	43.21	115.31
1100 CE	Magyars	333359	46.66	20.11
1100 CE	Makkura	567061	20.83	29.84
1100 CE	Mali	135238	10.78	-7.02
1100 CE	Nan Chao	681232	23.75	102.39
1100 CE	Orissa	268060	23.76	88.73
1100 CE	Pagan	215358	21.93	95.91
1100 CE	Poland (Piast)	225544	51.41	18.76
1100 CE	Song (Northern)	2799854	24.66	110.82
1100 CE	Srivijaya	385771	1.98	102.14
1100 CE	Volga Bulgars	304995	53.72	51.54
1100 CE	Xixia	633853	38.02	99.78
1100 CE	Zaydite	130324	14.30	45.96
Century	**Polity Name**	**Area km^2**	**Latitude**	**Longitude**
1200 CE	Abbasid (Buwayhids)	148732	31.55	46.76
1200 CE	Almohad	552759	36.43	-1.76
1200 CE	Alwa	383600	15.19	30.21
1200 CE	Ayyubid	1894077	30.12	33.22
1200 CE	Bulgaria	251855	43.69	24.76
1200 CE	Byzantine	330717	39.31	25.10
1200 CE	Capetian	341121	46.63	2.12
1200 CE	Castille	151130	40.88	-3.04
1200 CE	Champa	182432	17.43	107.14
1200 CE	Choe	152321	37.27	127.55
1200 CE	Dai Viet	182432	17.43	107.14
1200 CE	Denmark	155626	55.34	12.44
1200 CE	Eastern Ganga (Kalinga)	121725	19.84	84.58
1200 CE	Georgia	132450	41.93	43.75
1200 CE	Ghana	124617	15.92	-2.69
1200 CE	Ghurid	2577000	30.07	69.41
1200 CE	Holy Roman	789283	48.43	10.27
1200 CE	Hoysalas	158703	13.81	77.66
1200 CE	Jurchen / Jin	2761619	40.73	116.35
1200 CE	Kamakura	274458	34.23	133.87
1200 CE	Kara-Khitai	995366	43.74	70.32

1200 CE	Khmer	773743	14.01	103.58
1200 CE	Khwarizm	1311657	36.80	57.14
1200 CE	Kievan Rus	939619	53.48	30.86
1200 CE	Leon	122289	41.98	-6.49
1200 CE	Magyars	334053	46.66	20.08
1200 CE	Makkura	554087	20.93	29.72
1200 CE	Mali	156357	10.95	-6.88
1200 CE	Nan Chao	717450	24.09	102.05
1200 CE	Orissa	133058	23.35	87.75
1200 CE	Pagan	448493	20.47	95.70
1200 CE	Pandyas (Tamil)	107964	10.51	78.06
1200 CE	Plantagenet	220344	50.38	-0.35
1200 CE	Poland (Piast)	179517	51.26	18.27
1200 CE	Seljuk	387003	38.94	34.49
1200 CE	Song (Southern)	2165495	23.85	110.78
1200 CE	Songhai	116155	15.17	4.80
1200 CE	Srivijaya	454082	0.13	104.14
1200 CE	Sweden	112050	58.06	17.24
1200 CE	Vladimir-Suzdal	148079	58.81	30.69
1200 CE	Volga Bulgars	360286	53.68	51.70
1200 CE	Xixia	394144	39.08	96.28
1200 CE	Yadava (Seuna)	430106	19.59	77.02
1200 CE	Zagwe	228754	12.51	38.06
1200 CE	Zaidi	449439	16.10	47.61

Century	**Polity Name**	**Area km^2**	**Latitude**	**Longitude**
1300 CE	Alwa	374925	15.39	30.29
1300 CE	Aragon	129973	40.77	0.12
1300 CE	Byzantine	367662	38.57	24.89
1300 CE	Capetian	459898	46.76	1.81
1300 CE	Castille	330520	40.60	-4.73
1300 CE	Chagatai	2949831	40.63	74.04
1300 CE	Delhi	1181612	27.54	73.21
1300 CE	Eastern Ganga (Kalinga)	129996	19.79	84.49
1300 CE	Golden Horde	4579141	51.24	53.53
1300 CE	Hafsid (Berber)	240521	34.41	10.59
1300 CE	Hoysalas	144820	12.95	77.16
1300 CE	Il-Khanate	3070562	34.22	53.70
1300 CE	Kakityas	212764	16.49	79.39
1300 CE	Kamakura	277839	34.20	133.78
1300 CE	Kanem-Bornu	387969	15.77	12.79
1300 CE	Khmer	405577	13.15	105.41
1300 CE	Lithuania	141302	54.73	24.52
1300 CE	Magyars	414839	46.30	21.07

1300 CE	Makkura	531197	20.96	29.78
1300 CE	Mali	1531681	14.40	-7.61
1300 CE	Mamluke	1381615	27.77	33.21
1300 CE	Marinid	152023	33.55	-5.99
1300 CE	Moscow	148664	56.05	31.51
1300 CE	Novgorod	716009	61.96	36.77
1300 CE	Orissa	165702	23.19	87.61
1300 CE	Pandyas (Tamil)	144169	10.92	78.41
1300 CE	Plantagenet	148179	52.45	-1.67
1300 CE	Poland (Piast)	176147	51.33	18.81
1300 CE	Rasulid	453620	16.03	47.43
1300 CE	Rum	203292	39.05	35.55
1300 CE	Solomonid	204510	12.72	37.87
1300 CE	Sukhothai	409704	14.94	100.56
1300 CE	Sweden	112874	58.56	15.58
1300 CE	Yadava (Seuna)	606403	19.35	77.05
1300 CE	Yuan	13210835	29.30	107.46
1300 CE	Zayyanid/Tlemcen	142479	35.50	1.93
Century	**Polity Name**	**Area km^2**	**Latitude**	**Longitude**
1400 CE	Alwa	368484	15.29	30.01
1400 CE	Aragon	107589	40.57	1.92
1400 CE	Ashikaga	324315	36.34	135.77
1400 CE	Ayutthaya	349014	13.71	100.37
1400 CE	Bahmani	378236	18.55	76.51
1400 CE	Bohemia	163191	50.78	14.39
1400 CE	Byzantine	729454	38.83	27.49
1400 CE	Castille	359754	40.30	-4.46
1400 CE	Chagatai	869548	40.33	75.86
1400 CE	Dai Viet	110501	19.94	106.28
1400 CE	Delhi	1018560	26.27	74.49
1400 CE	France	453698	46.75	2.16
1400 CE	Golden Horde	3729082	52.22	54.52
1400 CE	Habsburg	155581	47.73	12.65
1400 CE	Hafsid (Berber)	167791	35.14	9.50
1400 CE	Hanseatic league	212751	54.72	10.70
1400 CE	Kalmar	845387	62.16	16.14
1400 CE	Kanem-Bornu	160443	13.97	13.45
1400 CE	Khmer	347798	12.54	105.57
1400 CE	Magyars	408301	46.01	20.33
1400 CE	Mali	639075	11.55	-9.15
1400 CE	Mamluke	1102634	29.27	32.75
1400 CE	Marinid	258319	32.70	-6.11
1400 CE	Ming	6508790	27.47	110.31

1400 CE	Moldavia	101218	47.03	26.94
1400 CE	Moscow	355998	57.79	37.87
1400 CE	Novgorod	998893	63.02	42.35
1400 CE	Orissa	278261	23.92	86.86
1400 CE	Plantagenet	177042	52.89	-4.22
1400 CE	Poland-Lithuania	1032701	51.84	27.54
1400 CE	Rasulid	455888	16.06	47.35
1400 CE	Solomonid	266759	13.35	37.62
1400 CE	Teutonic Order	154422	55.94	22.47
1400 CE	Timur	4790109	35.69	58.22
1400 CE	Vijayanagara	397478	12.66	77.28
1400 CE	Yi	306682	39.29	127.58
Century	**Polity Name**	**Area km^2**	**Latitude**	**Longitude**
1500 CE	Adal	274718	10.04	41.02
1500 CE	Alwa	303344	15.69	30.19
1500 CE	Ashikaga	297033	36.30	135.61
1500 CE	Astrakhan	108796	46.40	47.64
1500 CE	Ayutthaya	393877	14.06	100.33
1500 CE	Benin	105614	7.22	1.69
1500 CE	Chagatai	1040446	42.05	74.94
1500 CE	Crimean Khanate	271620	46.28	38.64
1500 CE	Dai Viet	163630	18.99	106.36
1500 CE	Delhi	1100143	29.03	76.08
1500 CE	France	452265	46.48	2.14
1500 CE	Gajapathi	197251	18.41	82.91
1500 CE	Golden Horde	651865	50.43	44.12
1500 CE	Gondwana	250934	21.07	81.51
1500 CE	Hafsid (Berber)	228700	34.46	10.51
1500 CE	Holy Roman	905414	53.42	11.09
1500 CE	Kalmar	869175	62.47	16.38
1500 CE	Kanem-Bornu	217767	13.85	13.37
1500 CE	Magyars	400492	46.11	20.26
1500 CE	Malaccca	246426	2.47	101.94
1500 CE	Mali	573315	11.60	-9.24
1500 CE	Malwa	115237	23.73	76.20
1500 CE	Mamluke	1141641	29.01	33.78
1500 CE	Ming	4990533	26.04	108.81
1500 CE	Moscow	2386531	62.68	46.12
1500 CE	Orissa	299942	23.42	87.97
1500 CE	Ottoman	984909	37.44	27.27
1500 CE	Pegu (Mon)	108262	17.56	96.23
1500 CE	Poland-Lithuania	1179851	52.02	27.40
1500 CE	Saffavid empire	138604	41.24	47.82

1500 CE	Shaibani / Shaybanids	1752134	44.09	61.66
1500 CE	Solomonid	241663	13.63	36.63
1500 CE	Songhai	1077765	16.88	-0.77
1500 CE	Spain	502910	40.32	-3.47
1500 CE	Tibet	2760019	33.24	89.01
1500 CE	Timurids	1208195	32.33	61.56
1500 CE	Tudor	153506	52.47	-1.68
1500 CE	Vijayanagara	363056	12.35	77.63
1500 CE	Wattasid	252361	32.85	-5.89
1500 CE	White Sheep Turks	1371855	33.88	49.13
1500 CE	Yi	252070	38.65	127.52
1500 CE	Zaidi	234284	15.10	45.13
1500 CE	Zayyanid/Tlemcen	123037	35.39	1.15

Sources

Barnes, I. and Hudson, R. (1998) Historical Atlas of Asia: From the Word's Oldest Civilizations to Emerging Superpower. Wiley.

Barraclough, G. ed. (1994) The Times Atlas of World History (Fourth Edition). BCA.

Black, J. ed. (2008) World History Atlas (Third Edition). Dorling Kindersley.

Bregel, Y. (2003) An Historical Atlas of Central Asia. Brill Academic.

Cussans, T., Parker, P., Winkleman, B., and Oliver, C. ed. (1994) The Times Atlas of European History. Times Books.

Davidson, P. (2011) Atlas of Empires. New Holland Publishers (UK) Ltd.

Encyclopedia Britannica (2011) Encyclopedia Britannica (http://www.britannica.co.uk). Encyclopedia Britannica UK Ltd

Euratlas-Nüssli (2011) Euratlas (euratlas.com). Euratlas-Nüssli.

Hammond World Atlas Corporation (2007) Historical World Atlas (Revised Edition). Hammond World Atlas Corporation.

Haywood, J. (2000) The Cassell Atlas of World History: The Ancient & Classical Worlds (Volume One). Cassell

Haywood, J. (2000) The Cassell Atlas of World History: The Medieval & Early Modern Worlds (Volume 177 Two). Cassell

Haywood, J. (2011) The New Atlas of World History. Thames & Hudson Ltd.

Kinder, H., and Hilgemann, W. (1974) The Anchor Atlas of World History From the Stone Age to the Eve of the French Revolution (Volume I). Anchor Books.

Lessman, T. (2011) Talessman's Atlas of World History. Thomas Lessman. downloaded 25/3/2011 - free to download from worldhistorymaps.info.

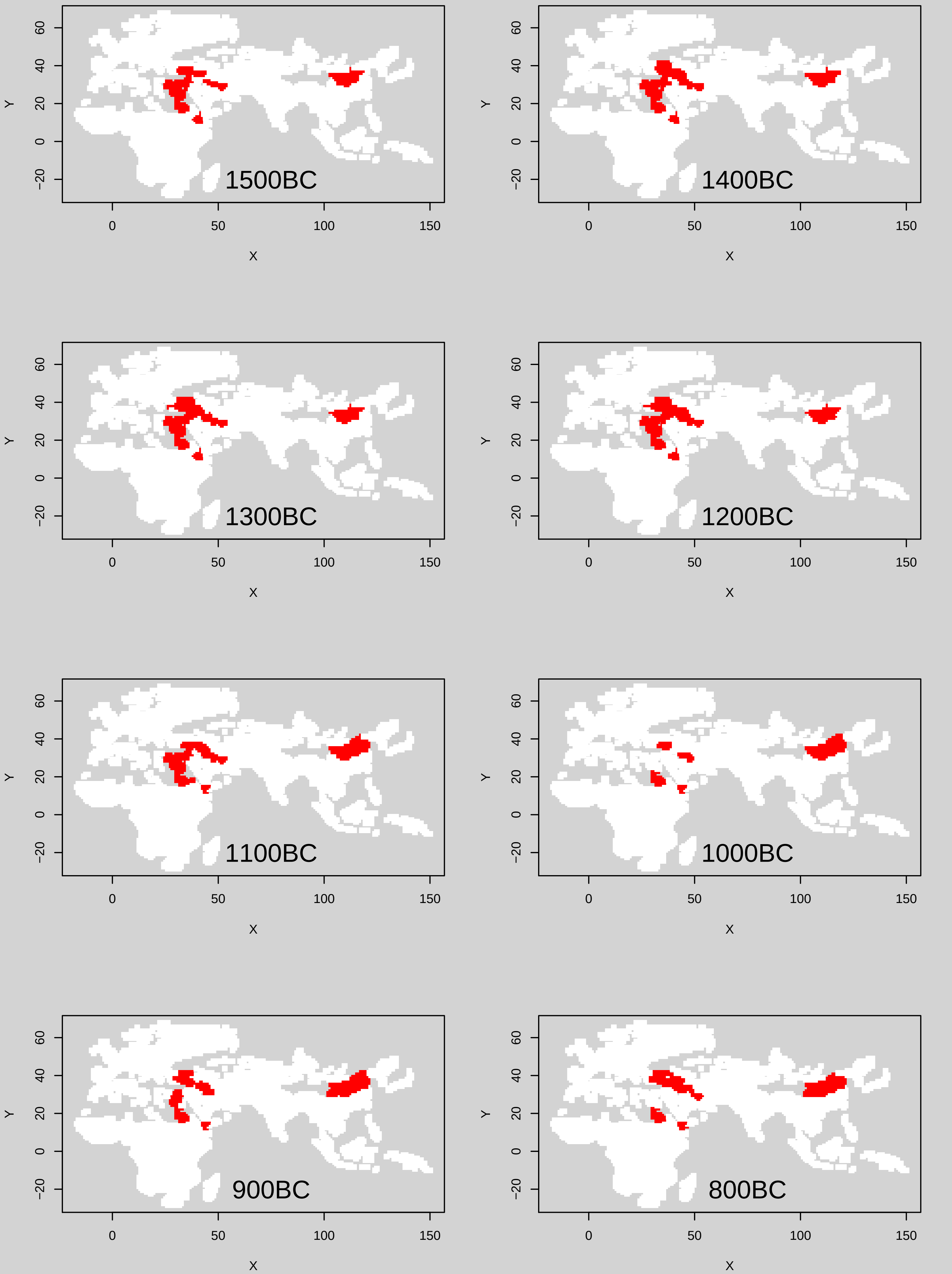
Nelson, J. (2010) World History Maps Interactive Historical Atlas: One Map for Every Year The Old World Since 500 BCE (worldhistorymaps.com). World History Maps Inc.

O'Brien, P. ed. (2007) Atlas of World History (Concise Edition). Oxford University Press.

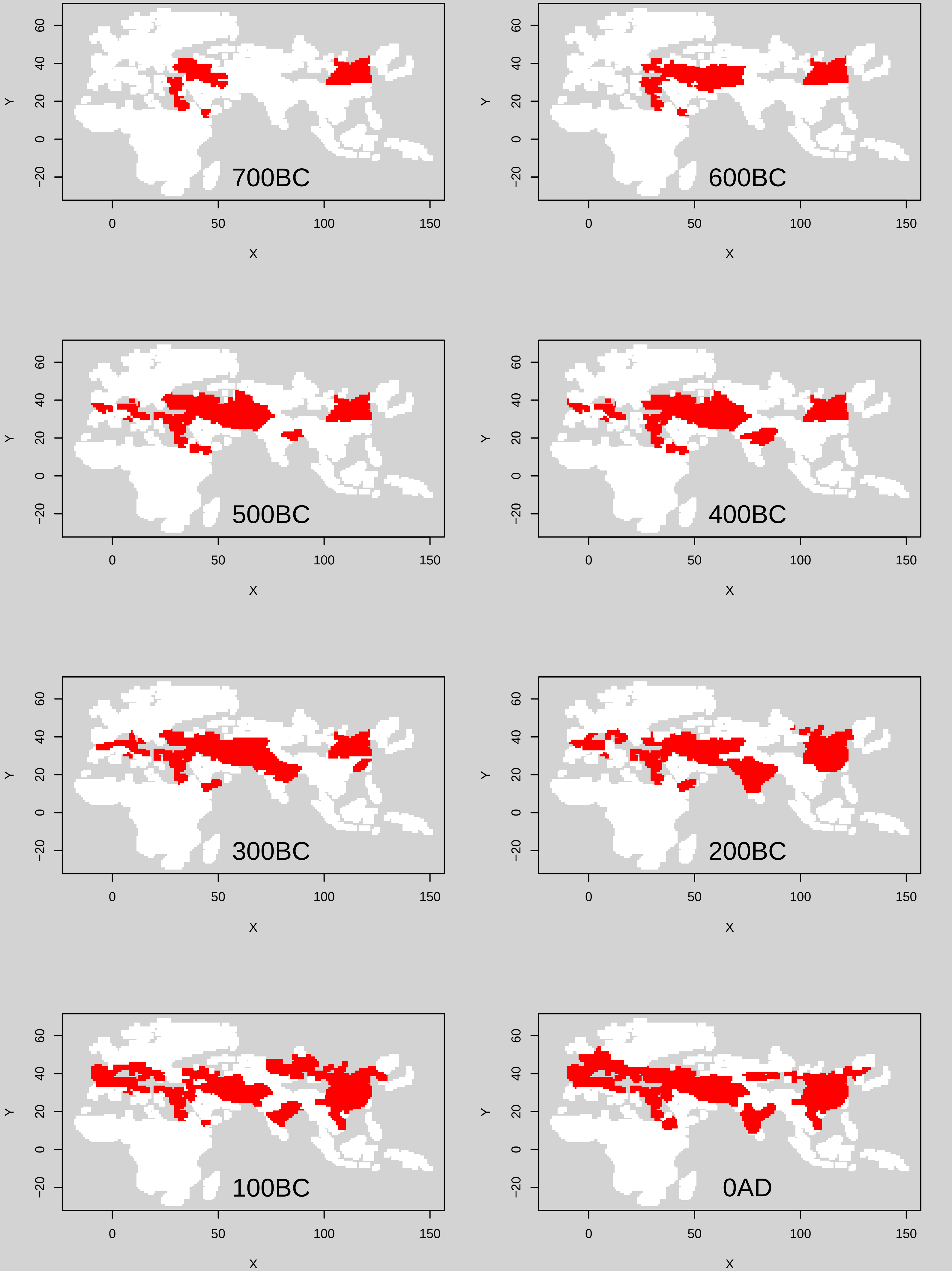
Shepherd, W. (1911) Historical Atlas. Henry Holt and Company.

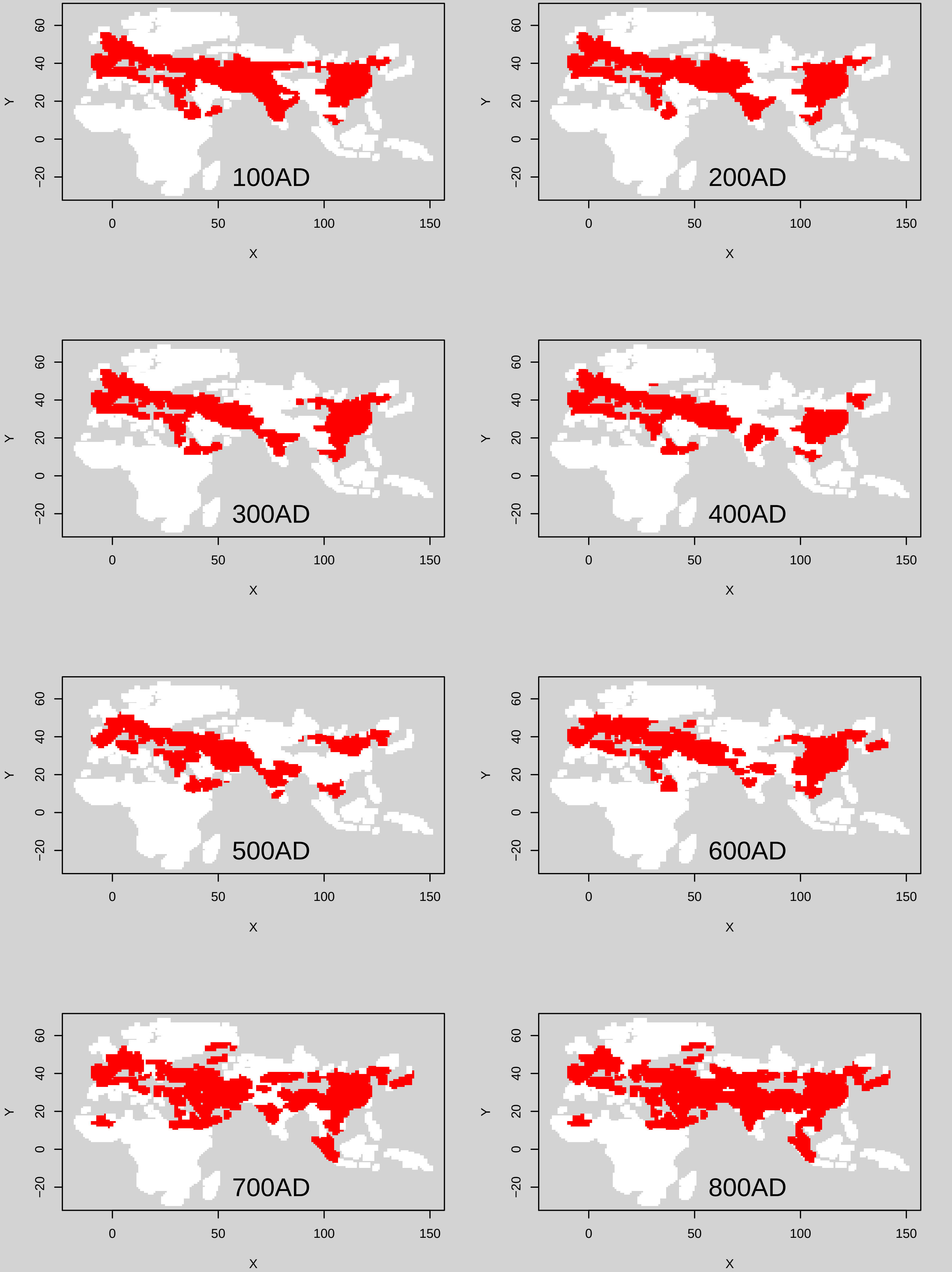
Shepherd, W. (1923) Historical Atlas. Henry Holt and Company.

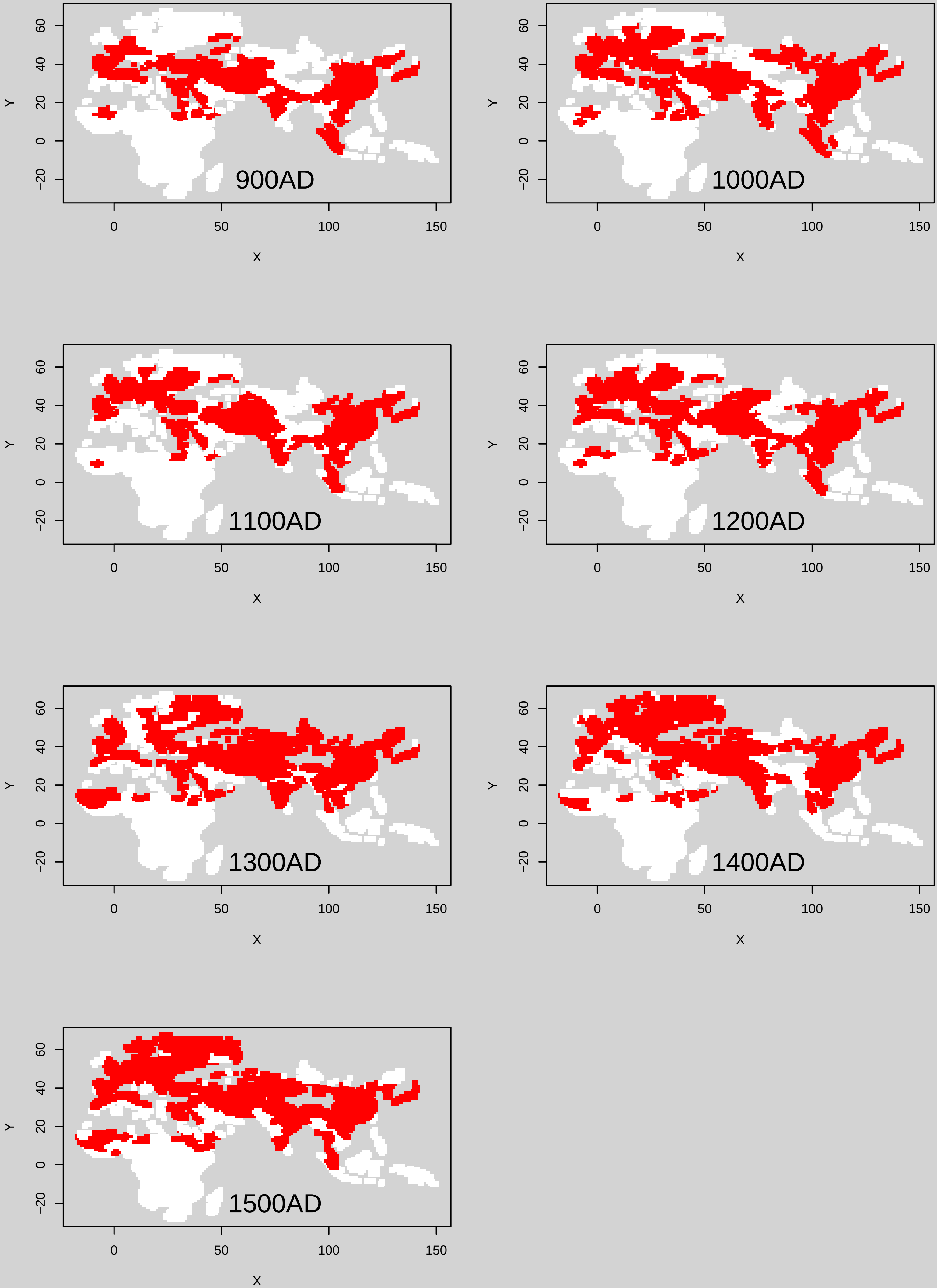
Wawro, G. ed. (2008) Historical Atlas A Cartographic History of Our Lives and History from the Beginning of Recorded Time to the Present Day. Millennium House Pty Ltd.



1500BC
1400BC
1300BC
1200BC
1100BC
1000BC
900BC
800BC
X
Y
0
50
100
150
−20
0
20
40
60


700BC
600BC
500BC
400BC
300BC
200BC
100BC
0AD
Y
X
60
40
20
0
−20
0
50
100
150

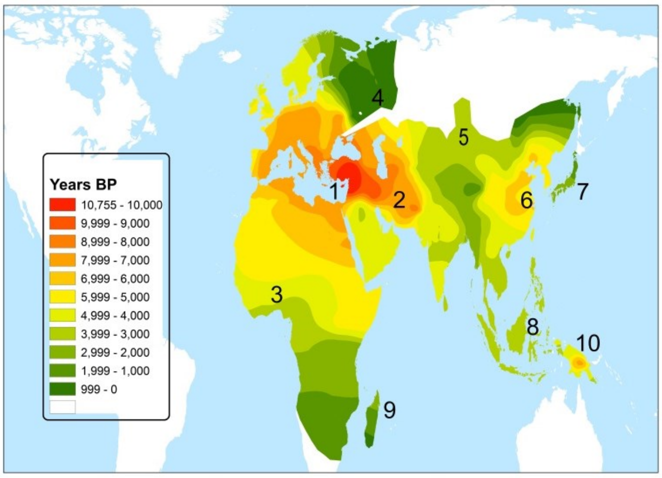

100AD
200AD
300AD
400AD
500AD
600AD
700AD
800AD
X
Y
0
50
100
150
−20
0
20
40
60


900AD
1000AD
1100AD
1200AD
1300AD
1400AD
1500AD
X
Y
0
50
100
150
−20
0
20
40
60

APPENDIX II: Creating maps of the spread of agriculture

In order to create our duration of agriculture variables we needed spatially-explicit dates of the beginnings of agricultural production in different parts of the Old world and New Guinea. In order to do this we needed a map. However, a reliable, up-to-date map that covers the whole of our area of interest does not exist. We therefore created such a map by joining together information from different regions, assigning dates to points and lines in a GIS and interpolating between these points (see Methods above). Here we describe the sources used and the decisions made for the different regions of the world. The sources below are primarily syntheses of primary sources including some maps, but in some cases we also consulted primary sources for certain radiocarbon dates of archaeological materials. In addition to the regionally specific sources consulted below we also made use of some more general sources. Larson et al. (*1*) provide a recent summary of likely dates of the beginning of agriculture in different regions of the world based on knowledge of the primary literature from a number of co-authors, while Mazoyer & Roudart (*2*) provide an approximate map of the spread of agriculture which was useful for assessing broad, global-scale patterns but was not used to determine to dates.

The aim here is not to provide the definitive statement on the dates of the beginnings of agriculture in different parts of the world but rather to provide a reasonable, global-scale summary of the main patterns of the spread of agricultural forms of production. In order to produce data on the scale we need for our analyses we necessarily have to make estimates and infer dates for regions that are not well covered by the archaeological and historical records. This is done systematically by interpolating between points where information has been collected or where previous scholars have placed estimates. Particular estimates or judgements can always be updated or challenged based on new evidence or different interpretations of evidence. In order to incorporate the uncertainty inherent in such an enterprise we include a "best" estimate as well as a minimum and maximum estimate. Where possible this is based on ranges presented in the literature. Our resulting maps are necessarily painted with broad brush strokes, however this is suitable for the scale of the analyses we are conducting in this paper. We feel they capture the main global-scale patterns sufficiently in a manner that is not biased to providing support for the duration of agriculture hypotheses for the emergence large scale-societies. Our best, maximum, and minimum interpolated estimates are presented below after describing the information used for each region.



Appendix II Fig 1. *Evidence for the emergence and spread of agriculture is described for the regions numbered 1-10*

1. SOUTHWEST ASIA, EUROPE, & NORTH AFRIA

Data on the spread of southwest Asian Neolithic crop assemblage were taken from Zohary et al. (*3*) who present a map (Plate 6, Map 2) of the earliest archaeological sites in which domesticated grain crops have been found.

2. WEST ASIA & SOUTH ASIA

The eastward spread of wheat and barley (key parts of the southwest Asian Neolithic crop assemblage) through west Asia into the Indian subcontinent have been mapped by Stevens et al. (*4*) using the AsCAD database. For this region the timing and extent were taken from figure 2 of this publication. There is also evidence for an independent southern Indian domestication of crops, and an independent origin of rice in the Ganges – the geographical distribution of these events was incorporated using maps and information from Fuller (*5, 6*)

3. SUB-SAHARAN AFRICA

Bellwood (*7*) has recently summarized archaeological and various other lines of evidence for the spread of livestock with an origin in Southwest Asia into sub-Saharan Africa, the domestication of various local species of plant (names) in Sub-Saharan Africa, and the southward spread of Bantu agriculturalists from equatorial Africa (*8*). Fuller (*9*) also provides dates for earliest evidence of cattle in archaeological sites in this region.

4. EASTERN EUROPE & EUROPEAN RUSSIA

The archaeological record for the spread of agriculture in this region is not as detailed as in other parts of Europe. Bellwood (*7*) (Figure 7.2, pg. 146) provides dates for eastern Europe – the movement of Neolithic cultures to regions east of the Dnieper river (7200 BP) and the Corded Ware archaeological culture (4800 BP in approximately modern-day Belarus, and 2500BP for regions further north). Later expansions of agriculture further east are unclear but we indicated in the "best" estimates that agriculture had spread as far as the eastern extent of the Kievan Rus principality at 1000AD, and as far as the Urals by 1300 AD based on the evidence from an historical atlas (http://www.worldhistorymaps.info/maps.html) that Russian principalities had extended this far by that point.

5. NORTHERN CENTRAL ASIA

Spengler (*10*) has summarized and synthesized the recent literature in this region, concluding that agriculture was present in northern Central Asia by the mid-3rd millennium BC, with a later spread north into the Altai mountain region.

6. CHINA & MAINLAND SOUTHEAST ASIA

Recent archaeological evidence and archaeobotanical data point to two major regions of substantial early cultivation in China: rice farming in wet fields around the Yangtze River, and dry-land cultivation of millets to the North (*11, 12*). The subsequent spread of agriculture involving these crops to the rest of this region were estimated using figures and information derived data from geo-referenced data on archaeological sites presented in papers by Stevens et al. (*4*) and by Fuller and colleagues (*5, 9*).

7. **JAPAN**

In Japan the spread of agriculture in the form of reliance on wet-rice cultivation is associated with the beginning of the Yayoi historical period (*13*), and is generally considered to have begun around 2300 BP. The earliest Yayoi sites are on the island of Kyushu in western Japan. According to Kidder (*14*) expansion to the Kanto (around modern-day Tokyo) and Tohoku (northern) regions continued from the middle to Late Yayoi. Agriculture was practiced in northern Tohuku before the end of the Yayoi period (1700BP).

8. ISLAND SOUTHEAST ASIA

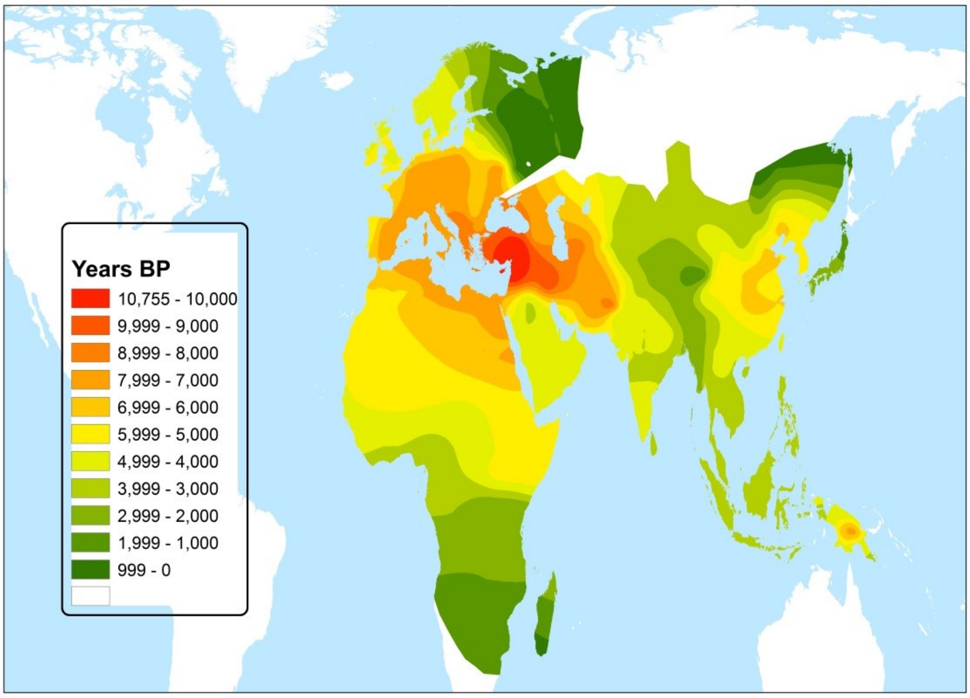
The spread of agriculture through island southeast Asia is associated with the spread of Austronesian-speaking societies from the island of Formosa (modern-day Taiwan)(*15-17*). Here this was modelled using archaeological dates from Bellwood (*18*), indicating a first movement into the Philippines 4000-4500BP, and a rapid spread to the rest of the region by around 3000-3500 BP.

9. MADAGASCAR

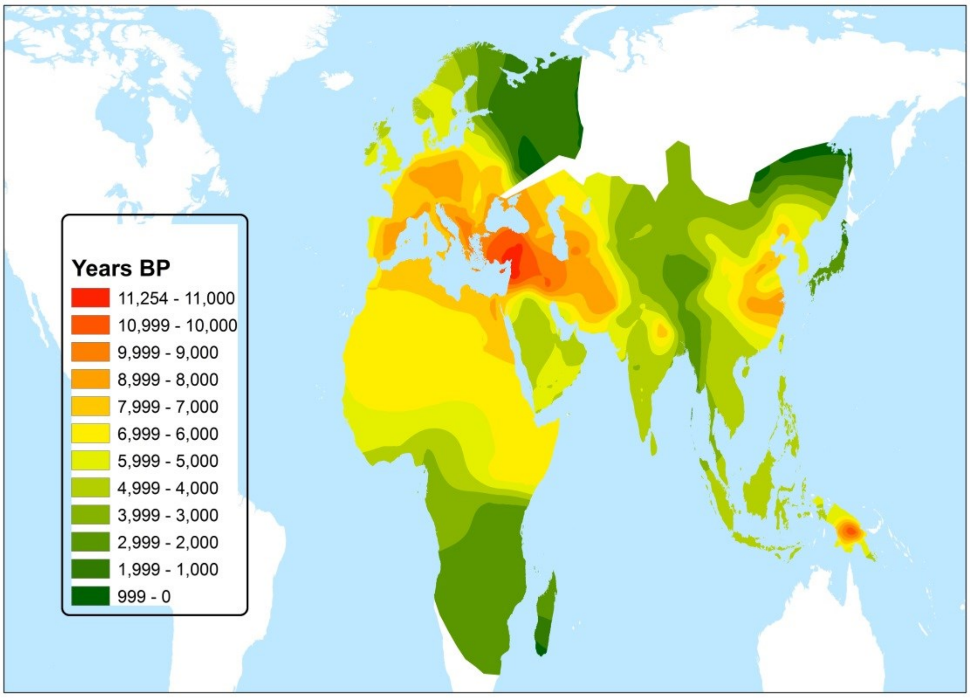
Modern populations in Madagascar speak Austronesian languages pointing to settlement from island southeast Asia as part of the expansion outlined above. Crowley (*19*) has summarized the evidence of human occupation: sites in northern Madagascar date to around 2000 BP, with a slightly later movement into the central highlands, and sites in the south suggest occupation was later still at around 600 BP.

10. NEW GUINEA

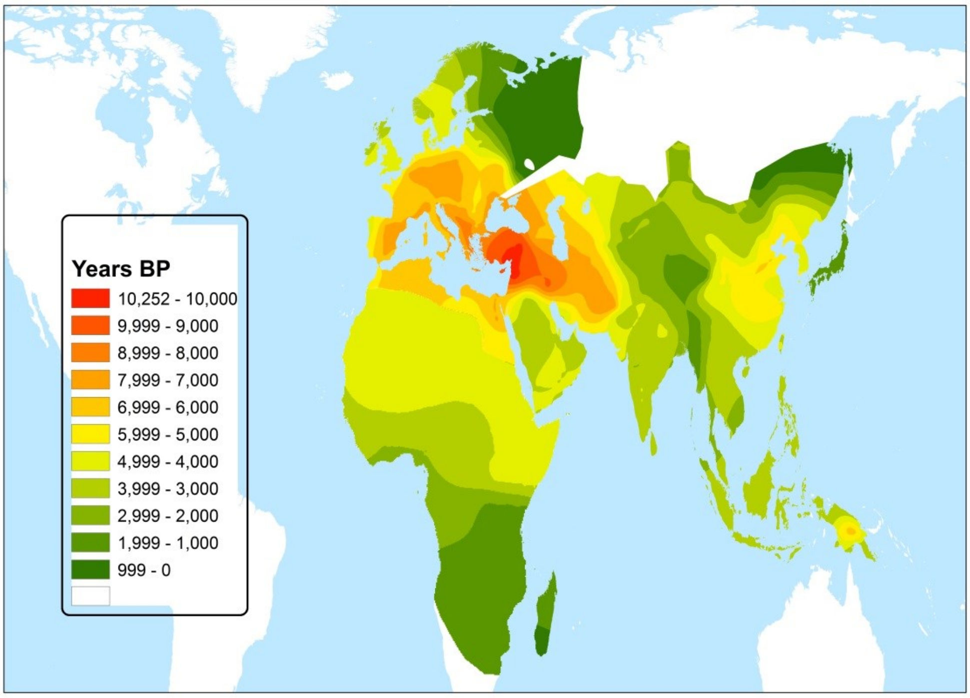
Archaeological excavations at Kuk swamp in the Highlands of New Guinea point this region as early and independent origin of agriculture, with evidence of cultivation of Taro and Yam dating to between 7000 and 4000 BP (*20, 21*). The timing and extent of the spread of agriculture within the island of New Guinea before Austronesians arrived on the northern coast around 3500BP (*22*) is unclear, however, Denham (*20*) argues that evidence from Banana and sugarcane in island southeast Asia suggests that there could have been contacts before 4000BP. Here we have incorporated this by modelling agriculture as spreading throughout New Guinea by 3000-5000BP.



Appendix II Figure 2. "Best" estimates for dates for the emergence of agriculture in Afro-Euraisa



Appendix II Figure 3. Maximum estimates for dates for the emergence of agriculture in Afro-Euraisa



Appendix II Figure 4. Minimum estimates for dates for the emergence of agriculture in Afro-Euraisa

Duration of agriculture references

1. G. Larson *et al.*, Current perspectives and the future of domestication studies. *Proceedings of the National Academy of Sciences* **111**, 6139 (April 29, 2014, 2014).
2. M. Mazoyer, L. Roudart, *A history of world agriculture: from the neolithic age to the current crisis*. (Earthscan/James & James, 2006).
3. D. Zohary, M. Hopf, E. Weiss, *Domestication of Plants in the Old World: The Origin and Spread of Domesticated Plants in Southwest Asia, Europe, and the Mediterranean Basin*. (OUP Oxford, 2012).
4. C. J. Stevens *et al.*, Between China and South Asia: A Middle Asian corridor of crop dispersal and agricultural innovation in the Bronze Age. *The Holocene* **26**, 1541 (2016/10/01, 2016).
5. D. Q. Fuller, Pathways to Asian Civilizations: Tracing the Origins and Spread of Rice and Rice Cultures. *Rice* **4**, 78 (2011).
6. D. Q. Fuller, in *Examining the Farming/language Dispersal Hypothesis,* P. Bellwood, C. Renfrew, Eds. (McDonald Institute for Archaeological Research, University of Cambridge, Cambridge, 2002).
7. P. Bellwood, *First Migrants: Ancient Migration in Global Perspective*. (Wiley, 2014).

8. T. E. Currie, A. Meade, M. Guillon, R. Mace, Cultural phylogeography of the Bantu Languages of sub-Saharan Africa. *Proceedings of the Royal Society B: Biological Sciences* **280**, (July 7, 2013, 2013).
9. D. Q. Fuller *et al.*, The contribution of rice agriculture and livestock pastoralism to prehistoric methane levels. *The Holocene* **21**, 743 (2011/08/01, 2011).
10. R. N. Spengler, N. Ryabogina, P. E. Tarasov, M. Wagner, The spread of agriculture into northern Central Asia: Timing, pathways, and environmental feedbacks. *The Holocene* **26**, 1527 (2016/10/01, 2016).
11. Z. Zhao, New Archaeobotanic Data for the Study of the Origins of Agriculture in China. *Current Anthropology* **52**, S295 (2011).
12. G.-A. Lee, G. W. Crawford, L. Liu, X. Chen, Plants and people from the Early Neolithic to Shang periods in North China. *Proceedings of the National Academy of Sciences* **104**, 1087 (January 16, 2007, 2007).
13. K. Mizoguchi, *The Archaeology of Japan: From the Earliest Rice Farming Villages to the Rise of the State*. (Cambridge University Press, 2013).
14. J. J. E. Kidder, in *The Cambridge History of Japan:,* D. M. Brown, Ed. (Cambridge University Press, Cambridge, 1993), vol. 1, pp. 48-107.
15. R. D. Gray, A. J. Drummond, S. J. Greenhill, Language Phylogenies Reveal Expansion Pulses and Pauses in Pacific Settlement. *Science* **323**, 479 (Jan 23, 2009).
16. P. Bellwood, J. J. Fox, D. Tryon, *The Austronesians: Historical and Comparative Perspectives*. (Australian National University, 2006).
17. P. Bellwood, *Prehistory of the Indo_Malaysian Archipelago*. (University of Hawai'i Press, Honolulu, 1997).
18. P. Bellwood, in *Southeast Asia: from prehistory to history,* I. Glover, P. Bellwood, Eds. (RoutledgeCurzon Press, NY, USA, 2004), pp. 21-40.
19. B. E. Crowley, A refined chronology of prehistoric Madagascar and the demise of the megafauna. *Quaternary Science Reviews* **29**, 2591 (9//, 2010).
20. T. Denham, Early Agriculture and Plant Domestication in New Guinea and Island Southeast Asia. *Current Anthropology* **52**, S379 (2011).
21. T. P. Denham *et al.*, Origins of Agriculture at Kuk Swamp in the Highlands of New Guinea. *Science* **301**, 189 (2003).
22. P. V. Kirch, *On the Road of the Winds: An Archaeological History of the Pacific Islands before European Contact*. (University of California Press, Berkeley and Los Angeles, 2000).