Quantitative historical analysis uncovers a single dimension of complexity that structures global variation in human social organization

Peter Turchin[a,b], Thomas E. Currie[c,1], Harvey Whitehouse[d,e], Pieter François[d,f], Kevin Feeney[g], Daniel Mullins[d,h], Daniel Hoyer[i], Christina Collins[c], Stephanie Grohmann[d], Patrick Savage[d], Gavin Mendel-Gleason[g], Edward Turner[i], Agathe Dupeyron[i], Enrico Cioni[i], Jenny Reddish[i], Jill Levine[i], Greine Jordan[i], Eva Brandl[i,j], Alice Williams[c], Rudolf Cesaretti[k], Marta Krueger[l], Alessandro Ceccarelli[m], Joe Figliulo-Rosswurm[n], Po-Ju Tuan[i], Peter Peregrine[o,p], Arkadiusz Marciniak[l], Johannes Preiser-Kapeller[q], Nikolay Kradin[r], Andrey Korotayev[s], Alessio Palmisano[t], David Baker[u], Julye Bidmead[v], Peter Bol[w], David Christian[u], Connie Cook[x,y], Alan Covey[z], Gary Feinman[aa], Árni Daníel Júlíusson[bb], Axel Kristinsson[cc], John Miksic[dd], Ruth Mostern[ee], Cameron Petrie[m,ff], Peter Rudiak-Gould[gg], Barend ter Haar[hh], Vesna Wallace[v], Victor Mair[ii], Liye Xie[jj], John Baines[kk], Elizabeth Bridges[ll], Joseph Manning[mm], Bruce Lockhart[nn], Amy Bogaard[oo], and Charles Spencer[pp,1]

[a]Department of Ecology and Evolutionary Biology, University of Connecticut, CT 06269; [b]Complexity Science Hub Vienna, 1080 Wien, Austria; [c]Human Behaviour & Cultural Evolution Group, Department of Biosciences, University of Exeter, Cornwall TR10 9FE, United Kingdom; [d]Institute of Cognitive and Evolutionary Anthropology, University of Oxford, Oxford OX2 6PE, United Kingdom; [e]Magdalen College, Oxford OX1 4AU, United Kingdom; [f]St. Benet's Hall, Oxford OX1 3LN, United Kingdom; [g]School of Computer Science and Statistics, Trinity College Dublin, Dublin 2, Ireland; [h]Institute of English Studies, University of London, London WC1E 7HU, United Kingdom; [i]Seshat: Global History Databank, Evolution Institute, San Antonio, FL 33576; [j]Department of Anthropology, University College London, London WC1H 0BW, United Kingdom; [k]School of Human Evolution and Social Change, Arizona State University, Tempe, AZ 85287; [l]Institute of Archaeology, Adam Mickiewicz University, 61-614 Poznań, Poland; [m]Department of Archaeology, University of Cambridge, Cambridge CB2 3DZ, United Kingdom; [n]Department of History, University of California Santa Barbara, Santa Barbara, CA 93106; [o]Anthropology and Museum Studies, Lawrence University, Appleton, WI 54911; [p]Santa Fe Institute, Santa Fe, NM 87501; [q]Division for Byzantine Research, Institute for Medieval Research, Austrian Academy of Sciences, 1020 Wien, Austria; [r]Department of Anthropology, Institute of History, Archaeology and Ethnology, Far Eastern Branch of the Russian Academy of Sciences, Vladivostok 690001, Russia; [s]Laboratory of Monitoring of Destabilization Risks, National Research University Higher School of Economics, Moscow 125267, Russia; [t]Institute of Archaeology, University College London, London WC1H 0PY, United Kingdom; [u]Big History Institute, Macquarie University, Sydney NSW 2109, Australia; [v]Department of Religious Studies, Chapman University, Orange, CA 92866; [w]East Asian Languages and Civilizations, Harvard University, Cambridge, MA 02138; [x]Department of Modern Languages & Literatures, Lehigh University, Bethlehem, PA 18015; [y]School of Historical Studies, Institute of Advanced Studies, Princeton, NJ 08540; [z]Department of Anthropology, University of Texas Austin, Austin, TX 78712; [aa]Integrative Research Center, Field Museum of Natural History, Chicago, IL 60605; [bb]History, University of Iceland, IS-108 Reykjavik, Iceland; [cc]History, Reykjavik Academy, IS-108 Reykjavik, Iceland; [dd]Department of Southeast Asian Studies, National University of Singapore, Singapore 119260; [ee]Department of History, University of Pittsburgh, Pittsburgh, PA 15260; [ff]Trinity College, Cambridge CB2 1TQ, United Kingdom; [gg]Independent Scholar, Toronto, ON M6P 1T6, Canada; [hh]Faculty of Oriental Studies, University of Oxford, Oxford OX2 6LU, United Kingdom; [ii]Department of East Asian Languages and Civilizations, University of Pennsylvania, Philadelphia, PA 19104; [jj]Department of Anthropology, University of Toronto, Toronto, ON M5S 2S2, Canada; [kk]Faculty of Oriental Studies, Oriental Institute, University of Oxford, Oxford OX1 2LE, United Kingdom; [ll]Institute of Archaeology and Anthropology, University of South Carolina, Columbia, SC 29208; [mm]Department of History, Yale University, New Haven, CT 06520; [nn]Department of History, National University of Singapore, Singapore 117570; [oo]Institute of Archaeology, University of Oxford, Oxford OX1 2PG, United Kingdom; and [pp]Division of Anthropology, American Museum of Natural History, New York, NY 10024

Do human societies from around the world exhibit similarities in the way that they are structured, and show commonalities in the ways that they have evolved? These are long-standing questions that have proven difficult to answer. To test between competing hypotheses, we constructed a massive repository of historical and archaeological information known as "Seshat: Global History Databank." We systematically coded data on 414 societies from 30 regions around the world spanning the last 10,000 years. We were able to capture information on 51 variables reflecting nine characteristics of human societies, such as social scale, economy, features of governance, and information systems. Our analyses revealed that these different characteristics show strong relationships with each other and that a single principal component captures around three-quarters of the observed variation. Furthermore, we found that different characteristics of social complexity are highly predictable across different world regions. These results suggest that key aspects of social organization are functionally related and do indeed coevolve in predictable ways. Our findings highlight the power of the sciences and humanities working together to rigorously test hypotheses about general rules that may have shaped human history.

cultural evolution | sociopolitical complexity | comparative history | comparative archaeology | quantitative history

T he scale and organization of human societies changed dramatically over the last 10,000 y: from small egalitarian groups integrated by face-to-face interactions to much larger societies with specialized governance, complex economies, and sophisticated information systems. This change is reflected materially in public buildings and monuments, agricultural and transport infrastructure, and written records and texts. Social complexity, however, is a characteristic that has proven difficult to conceptualize and quantify (1, 2). One argument is that these features of societies are functionally interrelated and tend to coevolve together in predictable ways (3, 4). Thus, societies in different places and at different points in time can be meaningfully compared using an overall

Significance

Do human societies from around the world exhibit similarities in the way that they are structured and show commonalities in the ways that they have evolved? To address these long-standing questions, we constructed a database of historical and archaeological information from 30 regions around the world over the last 10,000 years. Our analyses revealed that characteristics, such as social scale, economy, features of governance, and information systems, show strong evolutionary relationships with each other and that complexity of a society across different world regions can be meaningfully measured using a single principal component of variation. Our findings highlight the power of the sciences and humanities working together to rigorously test hypotheses about general rules that may have shaped human history.

measure of social complexity (2). Several researchers have attempted to come up with a single measure to capture social complexity (5–7), but a more common approach has been to use proxy measures, such as the population size of the largest settlement (7, 8), number of decision-making levels (9), number of levels of settlement hierarchy (10), or extent of controlled territory (11). Others have criticized this approach on the grounds that these proposed measures focus too narrowly on size and hierarchy (12, 13) or that there are multiple dimensions or variable manifestations of complexity (14). However, another common view is that different societies have unique histories and cannot be meaningfully compared in this way (15). Indeed, most historians have abandoned the search for general principles governing the evolution of human societies (16, 17). However, although every society is unique in its own ways, this does not preclude the possibility that common features are independently shared by multiple societies. How can we study both the diversity and commonalities in social arrangements found in the human past?

In this paper, we address these issues by building a global historical and archaeological database that takes into account the fragmentary and disputed nature of information about the human past. To test hypotheses about the underlying structure of variation in human social organization, we apply a suite of statistical techniques to these data, including principal component analysis (PCA). We then compare evolutionary trajectories in world regions by plotting the estimated first principal component (PC) of variation against time.

Building a Comparative Database of Human History

Previous attempts to address these questions have been limited by a reliance on verbal arguments (15, 18, 19), comparisons involving a small number of polities (20, 21), noncomprehensive data samples (3, 22), or nonsystematic methods of data coding and purely descriptive analyses (6, 23–25). To advance beyond purely theoretical debates and comparisons based on limited samples, we have built a massive repository of systematically collected, structured historical and archaeological data known as "Seshat: Global History Databank" (26) (*Materials and Methods*). In collecting data, we used a targeted, stratified sampling technique that aims to maximize the variation in forms of social organization captured from as wide a geographic range as possible [thus minimizing pseudoreplication of data points (27)]. Specifically, we divided the world into 10 regions and in each, selected three locations or "Natural Geographic Areas" (NGAs), representing early, intermediate, and late appearance of politically centralized societies (Fig. 1). The construction of this databank has been accomplished in collaboration with a large number of historical and archaeological experts. Our goal is to capture the state of the art knowledge about past societies, including where information

is uncertain or there are disagreements between researchers (*Materials and Methods*). The online version of the databank (seshatdatabank.info/) illustrates how entries in the databank are supported by explanations of coding choices and references (*SI Appendix, SI Methods*).

Our unit of analysis is a polity: an independent political unit that ranges in scale from groups organized as independent local communities to territorially expansive, multiethnic empires. To populate the databank, we coded information on all identifiable polities ($n = 414$) that occupied each of the 30 NGAs at 100-y time slices from the beginnings of agriculture (in some cases, as far back as 9600 BCE) to the modern period (in some cases, as late as 1900 CE) (*SI Appendix, SI Methods*). To capture different aspects of social complexity, we systematically collected data on 51 variables that could be reliably identified and categorized from the historical and archaeological records. These variables were then aggregated into nine "complexity characteristics" (CCs) (Fig. 2A). The first set of variables relates to the size of polities: polity population (CC1), extent of polity territory (CC2), and "capital" population (the size of the largest urban center; CC3). A second set of variables measures hierarchical complexity (CC4), focusing on the number of control/decision levels in the administrative, religious, and military hierarchies and on the hierarchy of settlement types (village, town, provincial capital, etc.). Government (CC5) variables code for the presence or absence of official specialized positions that perform various functions in the polity: professional soldiers, officers, priests, bureaucrats, and judges. This class also includes characteristics of the bureaucracy (e.g., presence of an examination system), the judicial system, and specialized buildings (e.g., courts). Infrastructure (CC6) captures the variety of observable structures and facilities that are involved in the functioning of the polity. Information system (CC7) codes the characteristics of writing, record-keeping, etc. We also record whether the society created literature on specialized topics, including history, philosophy, and fiction (texts; CC8). Finally, economic development is reflected in monetary system (CC9), which represents the "most sophisticated" monetary instrument present in the coded society, and indicates the degree of economic complexity that would be possible. Our data collection process also allows us to incorporate uncertainty in this coding or disagreement among sources (*Materials and Methods*).

Testing Hypotheses About the Evolution of Social Complexity

To test between the different hypotheses laid out above, we analyzed these data using PCA, which assesses the extent to which different variables are tapping into shared dimensions of variation. We expected CC1–CC3 to cluster tightly together, as they all measure size, albeit in somewhat different ways. Beyond this, if the variation in social organization across different societies can be meaningfully captured by a single measure of social complexity, we would predict that the different CCs would correlate strongly with each other and be captured in one PC of variation onto which all CCs load. If social complexity is predictably multidimensional, then other PCs capturing significant amounts of variation might also be present.

We hypothesized that social complexity could be captured by two PCs (7). Size variables (CC1–CC3) should exhibit a strong relationship with hierarchical organization (CC4), as hierarchy is often thought to be a necessary mechanism for enabling effective information flows in large polities (19). We refer to the combination of size and hierarchy as "scale" (Fig. 2A). The other variables might form another dimension of "nonscale" complexity, perhaps reflecting specialization of roles and the products that emerge from such specialization. Another possibility is that these CCs covary in other ways or are free to vary independently (that is, they do not evolve together in a predictable manner). In the latter situation, we would not expect correlational analysis or the PCA to reveal any structure in terms of the relationships of these variables with each other.
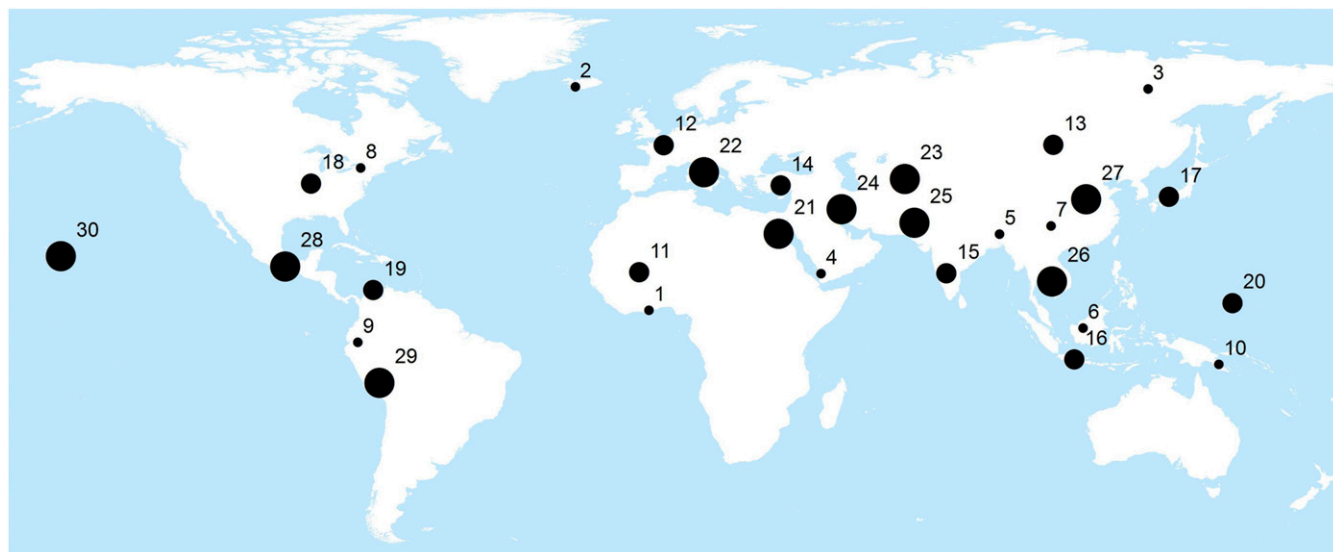
Fig. 1. Locations of the 30 sampling points on the world map (the size of the dot reflects the antiquity of centralized societies within the world region). The key to the numbers is in *SI Appendix*, Table S1.

Contrary to these expectations, all nine CCs showed substantial and statistically significant correlations with each other, with coefficients ranging from 0.49 to 0.88 (*SI Appendix*, Table S4). We found that a single PC, PC1, explains 77.2 ± 0.4% of variance. The proportion of variance explained by other PCs drops rapidly toward zero (Fig. 2*B*). Furthermore, all CCs load equally strongly onto PC1, indicating that PC1 captures contributions from across the multiple measures of social organization used here (Fig. 2*C* and *SI*

Appendix). This result provides strong support for the hypothesis that social complexity can be captured well by a single measure. In running these analyses, we have to take into account a number of factors, including missing data and various sources of autocorrelation. However, our results are robust to a large number of different assumptions and potential sources of error and bias (*SI Appendix, SI Results*).

We can also test directly the idea that societies that developed on distant world continents share enough similarities in their
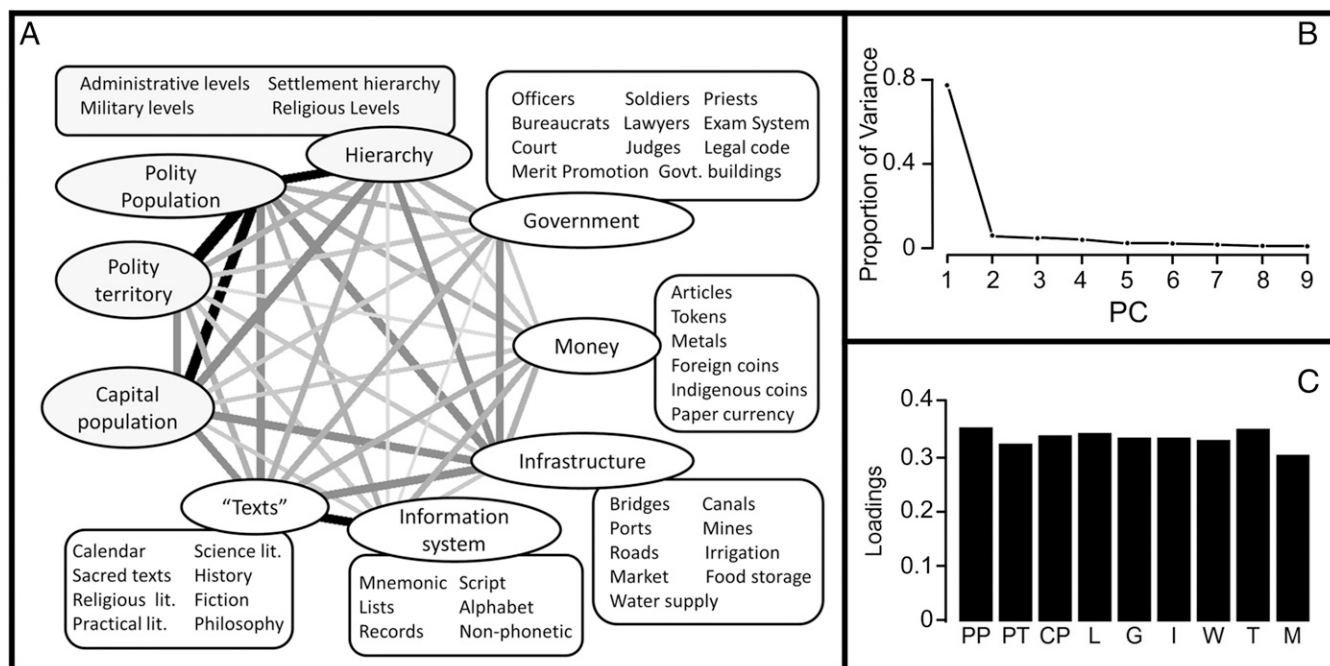
Fig. 2. (*A*) Nine CCs (ovals) aggregating 51 variables (*SI Appendix* has details on all CCs). Line width and color are proportional to the correlation coefficients between CCs (darker and thicker lines indicate stronger correlations). All CCs are significantly correlated with one another (correlation coefficients range between 0.49 and 0.88). Some variables show stronger linkages with each other, such as the scale variables (ovals shaded in gray), whereas money is less strongly correlated with the other variables. (*B*) Proportion of variance explained by PCs. (*C*) Factor loadings for CCs on PC1 indicating strong contributions by all CCs to a single dimension of social complexity. CP, capital population; G, government; I, infrastructure; L, levels; M, money; PP, polity population; PT, polity territory; T, texts; W, information system (writing).

Table 1. Cross-validation results for out of sample prediction of CCs across all world regions

Predicted CC	Overall ρ^2
Polity population	0.84
Polity territory	0.76
Capital population	0.71
Levels of hierarchy	0.60
Government	0.53
Infrastructure	0.62
Information system	0.59
Texts	0.73
Monetary system	0.53

Prediction accuracy is measured with prediction ρ^2 (*SI Appendix*, Table S2). Overall ρ^2 values are calculated as an average of the ρ^2 values weighted by the number of polities from which they are drawn.

complexity dimensions to allow for meaningful comparisons. We used the statistical technique of *k*-fold cross-validation (28), in which models are fitted on one set of data ("training set") and evaluated on another independent set ("testing set"). We reserved all data for polities in a particular world region, such as North America, as the testing set; developed predictive models on the rest of the data (by regressing each CC in turn on other CCs); and then, used the fitted models to predict each CC for North American polities. We then repeated this analysis for all other world regions. The accuracy of prediction is measured by the coefficient of prediction, ρ^2, which approaches one if prediction is very accurate, takes the value of zero when prediction is only as good as simply using the mean, and can take negative values if model prediction is worse than the mean.

Our results show that the values of CCs can be predicted by knowledge of other CCs (Table 1), and as Table 2 shows, median ρ^2 ranges between 0.08 (Southeast Asia) and 0.91 (North America), indicating that this predictive ability holds across all world regions. Low ρ^2 values do occur for some variables and seem to be lowest for those regions with the fewest number of polities to be predicted (*SI Appendix, SI Results*). This is to be expected, as with fewer cases to predict, there is less chance for general relationships to be detected. Some decreases in ρ^2 may also occur if smaller societies adopt some of the features, which make up CCs, from other societies, because they may be useful in dealing with larger societies (perhaps especially aspects of money and writing). Such selective adoption may not necessarily lead to the rapid devel-

opment of other aspects of complexity. Lower ρ^2 may also occur if some traits are retained when others are lost (see below).
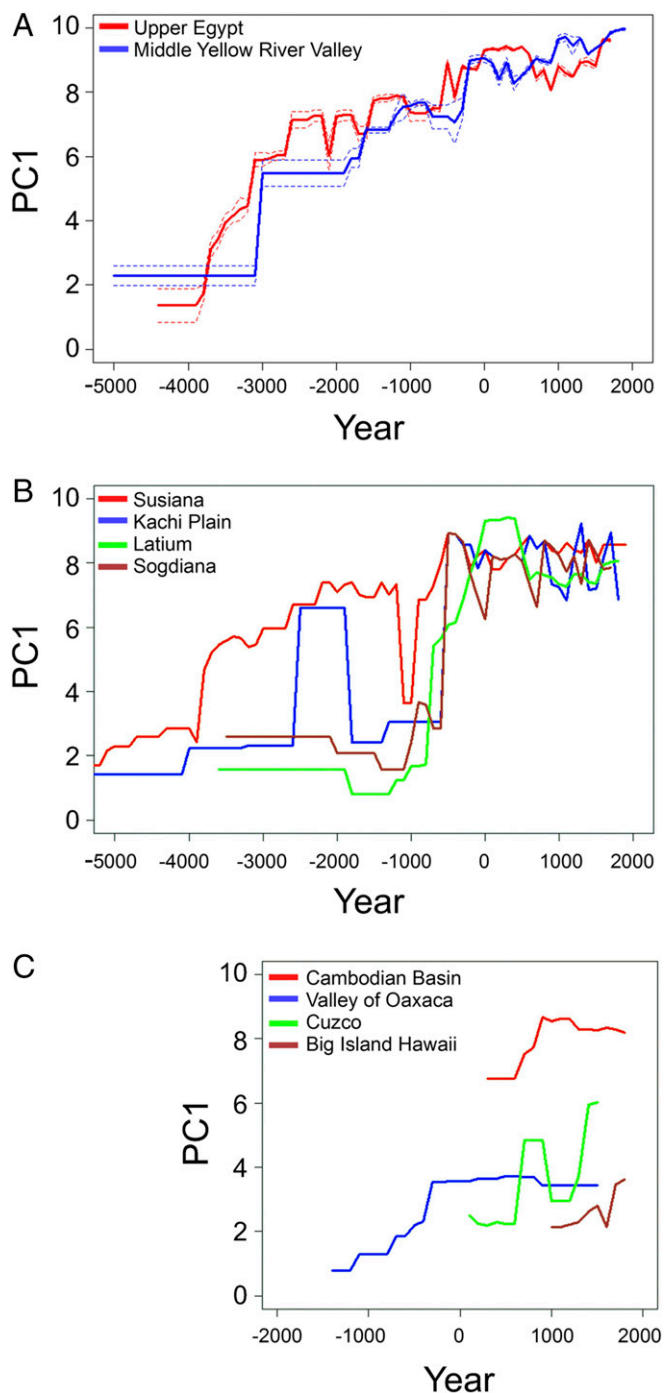
Comparing Evolutionary Trajectories

Our results, thus, indicate that there is striking similarity in the way that the societies in our global historical sample are organized. Examining PC1 enables us to compare how social complexity evolved in different parts of the globe over time. We plotted PC1 values estimated for each polity that occupied each of the 30 NGAs at 100-y time intervals. Fig. 3 compares the trajectories of the NGAs with early appearance of politically centralized societies in each of the 10 world regions (*SI Appendix* has all 30 trajectories). These trajectories indicate a general increase in complexity over time, albeit with occasionally substantial decreases in complexity (29). This comparison shows that there are crucial differences in the timing of takeoff and the rate of change as well as level of social complexity reached in different regions by 1900—differences that become clearly revealed through the analyses performed here. For example, although it is well-known that complex societies of the Americas emerged later than those in Eurasia, using our data, we can quantify their differences in social complexity. The difference in PC1 levels indicates that societies in the Americas were not as complex as those from Eurasia at time of contact, which may be a contributing factor in explaining why European societies were able to invade and colonize the Americas (30).

The tight relationships between different CCs provide support for the idea that there are functional relationships between these characteristics that cause them to coevolve (3). Scale variables are likely to be tightly linked, since increases or decreases in size may require changes in the degree of hierarchy (both too few and too many decision-making levels create organizational problems) (19). A similar argument has been put forward for size and governance (20). The production of public goods, such as infrastructure, may require solutions to collective action problems (31), and these can be provided by governance institutions and professional officials (32). Despite these linkages, because of their nature, different CCs are likely to show different temporal dynamics. Levels of nonscale characteristics, such as information systems, monetary systems, or infrastructure, may be retained and used even if a polity does decrease in size. Indeed, by retaining such features, the scale of the polity may more readily bounce back and return to its former level. This cultural continuity may be one reason why the trends that we see in our data are for social complexity to increase over time in a cumulative, ratchet-like manner (3, 33–35). For example, polities in our Italian NGA had writing, texts, and coins before the dramatic rises in scale of the Roman republic and empire, and they retained these features after the fall of Rome.

Discussion

One major conclusion from these analyses is that key aspects of human social organization tend to coevolve in predictable ways. This result supports the hypothesis that there are substantial commonalities in the ways that human societies evolve. Thus, societies can be meaningfully compared along a single dimension, which can be referred to as social complexity. Our analyses suggest that the estimated first PC of social complexity can be interpreted as a composite measure of the various roles, institutions, and technologies that enable the coordination of large numbers of people to act in a politically unified manner. However, as noted in the Introduction to this paper, the term "social complexity" has previously been defined and discussed in many ways. Indeed, complexity is a term that has many colloquial meanings, and there are many valid ways in which it could be applied to human social organization. For example, the kinship systems of some Australian Aboriginal groups, such as the Aranda, involve many complicated rules that determine who can marry whom (36, 37), and Turkana pastoralists have sophisticated social rules

Table 2. Cross-validation results for out of sample prediction of CCs summarized for different world regions

	ρ^2			
Predicted region	Median	Minimum	Maximum	n
Africa	0.72	0.37	0.90	41
Central Eurasia	0.63	−0.38	0.86	9
East Asia	0.70	0.30	0.93	34
Europe	0.53	−0.31	0.84	43
North America	0.91	0.79	0.97	11
Oceania–Australia	0.14	−3.21	0.97	1
South America	0.74	−24.57	0.97	5
South Asia	0.46	−0.05	0.69	12
Southeast Asia	0.08	−4.27	0.91	8
Southwest Asia	0.71	0.19	0.79	39
All regions	0.62	0.53	0.84	203

Prediction accuracy is measured with prediction ρ^2 (*SI Appendix*, Table S2). Median, minimum, and maximum indicate the median, smallest, and largest ρ^2 values across the nine CCs for the region, respectively.

Fig. 3. Trajectories of social complexity in 10 world regions quantified by PC1 values for locations where centralized, hierarchical polities first appeared in a particular region. (*A*) Africa and east Asia. Broken lines indicate 95% confidence intervals. (*B*) Southwest Asia, south Asia, Europe, and central Asia. (*C*) Southeast Asia, North America, South America, and Oceania. Confidence intervals for *B* and *C* are shown in *SI Appendix*, Figs. S4 and S5. PC1 has been rescaled to fall between 0 (low complexity) and 10 (high complexity) to aid interpretation. Flat horizontal lines indicate periods when there is no evidence of change from our polity data.

and norms that enable them to join together in large groups to conduct cooperative raiding missions (38).

Building historical databases, such as Seshat, allows us to take the vast amount of information about the human past and use it to test and reject competing hypotheses in the same cumulative process that characterizes the sciences (39, 40). It is important to

emphasize that we attach no normative judgment to the measure of social complexity that we have identified here; more complex societies are not necessarily "better" than less complex societies. We need to separate out these issues as well as ethnocentric judgments about non-European societies (2) from the kind of questions about how societies have actually evolved that we address here (3).

Our purpose here is not to propose that one definition of social complexity is superior to another. Instead, by supplying evidence that at least some aspects of human societies evolve in predictable and interconnected ways, this study illustrates that it is possible to move beyond the kind of verbal arguments that too often dominate debates about the evolution of human social organization. Furthermore, quantitative comparative analysis forces us to be more explicit about the evidence needed to support different claims and brings greater clarity to debates and discussions. It is important to recognize that, in any study, including this one, there are many subjective judgments about the coding of variables. Our goal in establishing the databank is to provide a summary of what is currently known about past human societies based on the literature and the expert knowledge of academics. It is not our aim to provide a more objective or definitive representation of such evidence but rather, to make the decisions and assumptions behind our data more explicit than has often been the case in the past. Our databank thus allows others viewing these data to challenge these decisions and provide alternative assessments. Future analyses can then assess whether alternative coding decisions substantially affect the results presented here.

The choice of variables and CCs themselves is also an important consideration in evaluating these results. We have attempted to be inclusive by choosing variables that would not favor particular forms of governance from certain parts of the world as being more complex. The variables are broad enough to allow for such features to come from a variety of specific institutions and are not biased toward Western forms of governance, which ultimately have their origins in early states in Greece and Mesopotamia. Our government variables (CC5), for example, capture the degree of specialization and professionalization of those involved in decision-making in sociopolitical affairs, a characteristic that has long been central to discussion of social complexity in different parts of the world (41). Our information system and texts variables (CC7 and CC8, respectively) capture the extent to which different types of information are being recorded and transmitted and reflect diversity and specialization in learning. Such information is potentially important in organizing societies or enabling societies to solve adaptive problems. Again, the variables within this category are broad enough to not be specific to any particular cultural tradition a priori. In particular, writing has been independently invented in such distant world regions as western Eurasia, east Asia, and Mesoamerica. As with the coding of specific variables, future analyses could assess whether the inclusion of alternative variables substantially affects the results presented here. Importantly, if our choice of variables was biased toward certain cultural–historical traditions, then this would reduce the correlations between different aspects of complexity, and these patterns would be different in different parts of the world. However, the overall high degree of correlation between CCs, as our cross-validation results indicate, suggests that the patterns that we have identified are relatively stable across regions.

The approach that we have taken in this paper can be used to resolve other long-standing controversies in the study of human societies. For example, some researchers have argued that traditional approaches to social complexity have overemphasized hierarchical relationships and did not pay enough attention to more horizontal or heterarchical forms of complexity (13, 42). Power relationships within societies can range from being autocratic or exclusionary (certain individuals or groups aim to control sources of power) to more corporate/collective, in which power is broadly shared across different sectors of societies (12, 43, 44). Other authors have

identified additional patterns that might be seen in human social evolution (21, 45), which can be fruitfully studied with the approach in this article. Indeed, some of the features that we have already coded, such as types and numbers of official positions, could be important in addressing such issues. We are already collecting data to test the idea that the balance between autocratic and collective forms of power has changed systematically over time, with autocratic forms being more prevalent in chiefdoms and early states. The emergence of institutions that held despotic leaders to account is argued to have occurred later (26), perhaps in connection with the emergence of certain religions (46, 47).

Our approach is also well-suited to go beyond identifying patterns and investigate the processes of sociopolitical evolution. The systematic compilation of long-term diachronic data for multiple variables on a large number of societies has been relatively rare in comparative history and archaeology (refs. 20, 35, and 48–50 have comparative studies of evolutionary trajectories for a smaller number of cases or time periods). Previous large-scale comparative approaches have generally focused on comparing evolutionary outcomes (end points) or snapshots at a single period of time rather than entire long-term trajectories (25, 51–54). By analyzing trajectories, we can both examine the processes that lead to variation in human societies across space and time and also take into account the historical changes that are contingent on the particular conditions and past history of the societies involved (3, 4, 55, 56).

In this study, the focus on looking at comparative changes over time enables us to investigate questions about the tempo of evolutionary change in human social systems. One pattern that is already apparent (Fig. 3 and *SI Appendix*, Fig. S6) is that many trajectories exhibit long periods of stasis or gradual, slow change interspersed with sudden large increases in the measure of social complexity over a relatively short time span. This pattern is consistent with a punctuational model of social evolution, in which the evolution of larger polities requires a relatively rapid change in sociopolitical organization, including the development of new governing institutions and social roles, to be stable (3, 4, 57). One example that has been investigated in previous work is the emergence of bureaucratic forms of governance, which tend to develop around the time when polities first extend political control beyond more than a day's round trip from the capital (20). A related idea is that, if there are strong relationships between these variables and if change is relatively rapid, then societies may tend to evolve toward certain types of sociopolitical organization based on associations between certain combinations of traits (3, 24, 57). Cluster analysis of PC1 shows some initial support for this idea, indicating a clear distinction between large societies that exhibit many of the nonscale features of complexity and smaller societies that lack most of these features, with other potential groupings within these clusters (*SI Appendix, SI Discussion* and Figs. S12 and S13).

Our data also indicate a shift toward more complex societies over time in a manner that lends support to the idea of a driving force behind the evolution of increasing complexity (3, 10, 58, 59) (*SI Appendix, SI Discussion*, Fig. S11, and Table S9). Such a driven trend is consistent with the hypothesis that competition between groups, particularly in the form of warfare, has been an important selective force in the emergence and spread of large, complex societies (10, 11, 60). In future work, the kind of systematic approach that we have used here will allow us to assess the large number of alternative mechanisms that have been proposed to explain the evolution of social complexity (2, 11, 14, 26). We are currently expanding the Seshat databank to collect information on agricultural productivity, warfare, religion, ritual, institutions, equity, and wellbeing in past societies to assess such competing hypotheses (26, 47, 61, 62).

Our focus in this paper has been on the increase in social complexity over time. However, understanding the causes of collapses and decreases in social complexity is an equally important research topic. As is clear in the evolutionary trajectories (Fig. 3 and *SI Appendix*, Fig. S6), declines in social complexity, some quite dramatic, are frequently seen in most NGAs. Furthermore, some of the large decreases are "hidden" when a polity collapses, but the NGA is immediately taken over by another large-scale society nearby. While different analytical approaches than the ones used in this article and additional data will be needed to study the processes explaining social collapse, such an investigation is entirely within the scope of the Seshat project.

In summary, our results indicate that it is indeed possible to meaningfully compare the complexity of organization in very different and unconnected societies along a single dimension (6, 30). Although societies in places as distant as Mississippi and China evolved independently on different continents and followed their own trajectories, the structure of social organization, as captured by the interrelations between different CCs, is broadly shared across all continents and historical eras. Key elements of complex social organization have thus coevolved in highly consistent ways across time and space. Differences in the timing of takeoff, the overall rate of increase, and the depth of periodic declines in social complexity provide us with highly informative data for testing theories of social and cultural evolution. Our databank was built via a collaborative relationship with humanities scholars who provided expert knowledge of past societies and helped guide data collection at all stages. This paper has shown the power of the sciences and the humanities working together to help us better understand the past by testing and rejecting alternative hypotheses about the general rules that have shaped human history.

Materials and Methods

Data. Data were collected as part of "Seshat: Global History Databank" (26) (*SI Appendix, SI Methods*). We collected data in a systematic manner by dividing the world into 10 major regions (Fig. 1 and *SI Appendix*, Fig. S1 and Table S1). Within each region, we selected three NGAs to act as our basic geographical sampling unit. Each NGA is spatially defined by a boundary drawn on the world map that encloses an area delimited by naturally occurring geographical features (for example, river basins, coastal plains, valleys, and islands).

Within each world region, we looked for a set of NGAs that would allow us to cover as wide a range of forms of social organization as possible. Accordingly, we selected three NGAs that varied in the antiquity of centralized, stratified societies (giving us one early-complexity, one late-complexity, and one intermediate-complexity NGA per region).

Our unit of analysis is a polity, an independent political unit that ranges in scale from villages (local communities) through simple and complex chiefdoms to states and empires. To code social complexity data, for each NGA, our team chronologically listed all polities that were located in the NGA or encompassed it (*SI Appendix, SI Methods* has a discussion of how we deal with cases where identifying a single polity is not appropriate). For each NGA, we start at a period just before the Industrial Revolution (typically 1800 or 1900 CE depending on the location) and go back in time to the Neolithic (subject to the limitation of data). We chose a temporal sampling rate of 100 y, meaning that we only included polities that spanned a century mark (100, 200 CE, etc.) and omitted any polities of short duration that only inhabited an NGA between these points. Data collection was accomplished by a team of research assistants guided by archaeologists and historians who are experts in the sampled regions and time periods. These experts also checked all data collected by research assistants. *SI Appendix, SI Methods* contains details about coding procedures, including how we decided on the variables to include in the Seshat codebook and how we explicitly engaged with such issues as missing data, uncertainty, and disagreement between experts. We have created a website (seshatdatabank.info/) that illustrates the databank. This online version currently displays information on the social complexity variables in the NGAs and polities analyzed in this study (see also *SI Appendix, SI Methods*). The website shows how entries in the databank are supplemented by explanations of coding decisions and references. The goal of the databank is to make as explicit as possible the evidentiary basis of inferences about the past and to share that information as widely as possible.

Multiple Imputation: Dealing with Missing Data, Uncertainty, and Expert Disagreement. Because of the fragmentary nature of the information that is available about past societies, it was not possible to reliably code all variables for all polities. There is, therefore, a nontrivial amount of data points for which we have been unable to assign even a broad range of possible values because of a lack of evidence (3,700 of the total of 21,000). The presence of such missing data is an important feature of our dataset, in that it accurately reflects our current understanding (or lack of it) about any particular feature in any particular past society. Missing data, however, present a challenge for the statistical analyses.

One way of dealing with incomplete datasets is to simply omit the rows in the data matrix that contain missing values. There are two problems with this approach. First, it can be very wasteful in that omitted rows may contain much useful information relating to the variables that we were able to code. Had we used this approach with our social complexity data, for example, we would have to throw away nearly one-half of the rows. Second, case deletion may lead to biased estimates, because there are often systematic differences between the complete and incomplete cases. In our case, in many NGAs, small-scale societies were present far back in time, and as a result, they are much harder to code. Additionally, some regions of the world have been subject to greater levels of research effort than others. Omitting many of the lesser known cases because of their larger proportion of missing values would give too much weight to later, better known societies from only some parts of the world. As an example, had we used the casewise deletion approach for our current dataset, we would end up with only a single observation for Australia–Oceania. Such unequal dropping of observations would very likely bias the results, since the analysis would be dominated by such regions as Europe and southwest Asia (each with ∼40 complete rows in the data matrix).

To deal with missing values as well as incorporate uncertainty and expert disagreement into our analyses, we use a technique known as multiple imputation (63), which utilizes modern computing power to extract as much information from the data as possible. Imputation involves replacing missing entries with plausible values, and this allows us to retain all cases for the analysis. A simple form of imputation, "single imputation," might replace any unknown cases for a binary "present/absent" variable with simply "absent" or to replace unknown cases of continuous variables with the mean for that variable. These approaches have similar drawbacks to case deletion, in that they tend to introduce a bias. Therefore, in this paper, we perform multiple imputation: analysis done on many datasets, each created with different imputed values that are sampled in probabilistic manner. This approach results in valid statistical inferences that properly reflect the uncertainty caused by missing values (64). Multiple imputation procedures can vary depending on the type of variable and the type of data coding issue faced.

Expert disagreement. In cases where experts disagree, each alternative coding has the same probability of being selected. Thus, if there are two conflicting codings presented by different experts and if we create 20 imputed sets, each alternative will be used roughly 10 times.

Uncertainty. Values that are coded with a confidence interval are sampled from a Gaussian distribution, with mean and variance that are estimated assuming that the interval covers 90% of the probability. For example, if a value of [1,000–2,000] was entered for the polity population variable, we would draw values from a normal distribution centered on 1,500 with an SD of 304. It is worth noting that this procedure means that, in 10% of cases, the value entered into the imputed set will be outside the data interval coded in Seshat. For categorical or binary variables, we sample coded values in proportion to the number of categories that are presented as plausible. For example, if our degree of knowledge does not allow us to tell whether a certain feature was present or absent at a particular time, then the imputed datasets will contain "present" for roughly one-half of the imputed sets and absent for roughly one-half of the sets.

Missing data. For missing data, we impute values as follows. Suppose that, for some polity, we have a missing value for variable A and coded values for variables B–H. We select a subset of cases from the full dataset, in which all values of A–H variables have values and build a regression model for A. Not all predictors B–H may be relevant to predicting A, and thus, the first step is selecting which of the predictors should enter the model (information on model selection is given below). After the optimal model is identified, we estimate its parameters. Then, we go back to the polity (where variable A is missing) and use the known values of predictor variables for this polity to calculate the expected value of A using the estimated regression coefficients. However, we do not simply substitute the missing value with the expected one (because as explained above, this is known to result in biased estimates). Instead, we sample from the posterior distribution characterizing the prediction of the regression model (in practice, we randomly sample the regression residual and add it to the expected value). We applied the same

approach to each missing value in the dataset, yielding an imputed dataset without gaps.

The overall imputation procedure was repeated 20 times, yielding 20 imputed sets that were used in the analyses below. The 20 imputed datasets are available online as Dataset S1.

Statistical Analysis.

PCA. PCA was used to investigate the internal correlation structure characterizing the nine measures of social complexity. PCA was run on each imputed dataset to estimate the proportion of variance explained by each PC (PC1–PC9), component loadings (correlations between the original variables and the PCs), and the values of PCs for each polity. Because we have 20 sets of all of these results, we also report the confidence intervals associated with these estimates. Values for PC1 derived from the 20 imputed datasets are available online as Dataset S2.

Cross-validation. For the multiple imputation to be a worthwhile procedure, we need to ascertain that the stochastic regression approach for predicting missing values actually yields better estimates than, for example, simply using the mean of the variable. To do this, we used a statistical technique known as *k*-fold cross-validation (28). In addition to this methodological issue, this cross-validation procedure allows us to address another substantive question, namely the extent to which the relationships between variables are consistent across different parts of the world. This is done by quantifying how well we can predict the value of a particular feature of a particular society based on known information about the values of other features in that society and the observed relationships between the known and the unknown variables in other societies.

Cross-validation estimates the true predictability characterizing a statistical model by splitting data into two sets. The parameters of the statistical model are estimated on the fitting set. Next, this fitted model is used to predict the data in the testing set. Because the prediction is evaluated on the "out of sample" data (data that were not used for fitting the model), the results of the prediction exercise give us a much better idea of how generalizable the model is compared with, for example, such regression statistics as the coefficient of determination, R^2.

The accuracy of prediction is often quantified with the coefficient of prediction (65):

$$\rho^2 = 1 - \frac{\sum_{i=1}^{n} \left(Y_i^* - Y_i \right)^2}{\sum_{i=1}^{n} \left(\bar{Y} - Y_i \right)^2},$$

where Y_i indicates the observations from the testing set (the omitted values), Y_i^* is the predicted value, $\bar{Y}$ is the mean of Y_i, and n is the number of values to be predicted. The coefficient of prediction ρ^2 equals one if all data are perfectly predicted and zero if the regression model predicts as well as the data average (in other words, if the model is simply $Y_i^* = \bar{Y}$). Unlike the regression R^2, which can vary between zero and one, prediction ρ^2 can be negative—when the regression model predicts data worse than the data mean. Prediction ρ^2 becomes negative when the sum of squares of deviations between predicted and observed is greater than the sum of squares of deviations from the mean.

In *k*-fold cross-validation, rather than having simply a single fitting set and one testing set, we divide the data into *k* sets. We selected those cases that had complete coding for all variables and divided our dataset into 10 sets for each of our 10 world regions. Next, we set aside one region (for example, Africa) and used the other nine regions to fit a regression model for the variable of interest. Let us say that *Y* is polity population, and we are interested in how well it can be predicted from knowing the population of the capital, hierarchy levels, writing, etc. We fit a regression model to the data from the other nine regions. We then predict the values of *Y* (polity population in this case) for Africa using the known values for other variables in African polities and the regression coefficients. Next, we omit another region (for example, Europe) and repeat the exercise. At the end, we have predicted all data points by the out of sample method, while fitting the model on 9/10th of data at any given step.

One important aspect of this procedure is to guard against overfitting (i.e., including too many predictor variables in the model), which is known to yield much worse predictability than a model that uses the "right" number of predictors (66). We have experimented with several methods of model selection that prevent overfitting. We found that a frequentist approach in which predictor variables are selected based on their *P* values (using the 0.05 threshold) does as well as the more commonly used model selection approach using the Akaike Information Criterion (AIC) (66). In fact, AIC tended

to slightly overfit compared with the frequentist approach. As the frequentist approach has an additional advantage of consuming less computer time, we used this approach for all cross-validation analyses reported below.

Multiple imputation, cross-validation, and PCA were all conducted using scripts written in the R statistical programming language (67).

1. Kohler TA, Crabtree SA, Bocinsky RK, Hooper PL (2015) Sociopolitical evolution in midrange societies: The pre-Hispanic Pueblo case. Available at https://www.santafe.edu/research/results/working-papers/sociopolitical-evolution-in-midrange-societies-the. Accessed May 24, 2017.
2. Carneiro RL (2003) *Evolutionism in Cultural Anthropology* (Westview, Boulder, CO).
3. Currie TE, Mace R (2011) Mode and tempo in the evolution of socio-political organization: Reconciling "Darwinian" and "Spencerian" evolutionary approaches in anthropology. *Philos Trans R Soc Lond B Biol Sci* 366:1108–1117.
4. Spencer CS (1990) On the tempo and mode of state formation–Neoevolutionism reconsidered. *J Anthropol Archaeol* 9:1–30.
5. White LA (1949) *The Science of Culture: A Study of Man and Civilization* (Farrar, Straus and Company, New York).
6. Morris I (2013) *The Measure of Civilization: How Social Development Decides the Fate of Nations* (Princeton Univ Press, Princeton).
7. Chick G (1997) Cultural complexity: The concept and its measurement. *Cross Cult Res* 31:275–307.
8. Ortman SG, Cabaniss AHF, Sturm JO, Bettencourt LMA (2015) Settlement scaling and increasing returns in an ancient society. *Sci Adv* 1:e1400066.
9. Currie TE, Mace R (2009) Political complexity predicts the spread of ethnolinguistic groups. *Proc Natl Acad Sci USA* 106:7339–7344.
10. Spencer CS, Redmond EM (2001) Multilevel selection and political evolution in the Valley of Oaxaca, 500-100 BC. *J Anthropol Archaeol* 20:195–229.
11. Turchin P, Currie TE, Turner EAL, Gavrilets S (2013) War, space, and the evolution of Old World complex societies. *Proc Natl Acad Sci USA* 110:16384–16389.
12. Blanton RE, Feinman GM, Kowalewski SA, Peregrine PN (1996) A dual-processual theory for the evolution of Mesoamerican civilization. *Curr Anthropol* 37:1–14.
13. Keech McIntosh S (1999) *Beyond Chiefdoms: Pathways to Complexity in Africa*, ed Keech McIntosh S (Cambridge Univ Press, Cambridge, UK), pp 1–30.
14. Feinman G (2013) *Cooperation and Collective Action: Archaeological Perspectives*, ed Carballo DM (Univ Press of Colorado, Boulder, CO), pp 35–56.
15. Yoffee N (2005) *Myths of the Archaic State: Evolution of the Earliest Cities, States, and Civilizations* (Cambridge Univ Press, Cambridge, UK).
16. Hunt L (1989) *The New Cultural History* (Univ of California, Berkeley, CA).
17. Sheehan JJ (2005) President's Column: How do we learn from history? *Perspect Hist* 43:1–3.
18. Feinman GM (2008) Variability in states: Comparative frameworks. *Soc Evol Hist* 7:54–66.
19. Turchin P, Gavrilets S (2009) Evolution of complex hierarchical societies. *Soc Evol Hist* 8:167–198.
20. Spencer CS (2010) Territorial expansion and primary state formation. *Proc Natl Acad Sci USA* 107:7119–7126.
21. Drennan RD, Peterson CE (2006) Patterned variation in prehistoric chiefdoms. *Proc Natl Acad Sci USA* 103:3960–3967.
22. Feinman GM, Neitzel J (1984) *Advances in Archaeological Method and Theory*, ed Schiffer MB (Academic, Orlando, FL), Vol 7, pp 39–102.
23. Currie TE (2014) Developing scales of development. *Cliodynamics* 5:65–74.
24. Peregrine PN, Ember CR, Ember M (2004) Universal patterns in cultural evolution: An empirical analysis using Guttman scaling. *Am Anthropol* 106:145–149.
25. Peregrine P, Ember CR, Ember M (2007) Modeling state origins using cross-cultural data. *Cross-Cultural Res* 41:75–86.
26. Turchin P, et al. (2015) Seshat: The Global History Databank. *Cliodynamics* 6:77–107.
27. Eff EA, Does Mr (2004) Galton still have a problem? Autocorrelation in the standard cross-cultural sample. *World Cult* 15:153–170.
28. Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence* (Morgan Kaufmann Publishers Inc., San Francisco), Vol 2, pp 1137–1145.
29. Currie TE, Greenhill SJ, Gray RD, Hasegawa T, Mace R (2010) Rise and fall of political complexity in island South-East Asia and the Pacific. *Nature* 467:801–804.
30. Diamond J (1997) *Guns, Germs and Steel* (Vintage, London).
31. Gavrilets S (2015) Collective action problem in heterogeneous groups. *Philos Trans R Soc Lond B Biol Sci* 370:20150016.
32. Mattison SM, Smith EA, Shenk MK, Cochrane EE (2016) The evolution of inequality. *Evol Anthropol* 25:184–199.
33. Richerson PJ, Boyd R (2001) *The Origin of Human Social Institutions*, ed Runciman WG (Oxford Univ Press, Oxford), pp 197–234.
34. Henrich J (2015) *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter* (Princeton Univ Press, Princeton).
35. Marcus J (1998) The peaks and valleys of ancient states: An extension of the dynamic model. *Archaic States* (School of American Research Press, Santa Fe, NM), pp 59–94.
36. Denham WW, McDaniel CK, Atkins JR (1979) Aranda and Alyawara kinship: A quantitative argument for a double helix model. *Am Ethnologist* 6:1–24.
37. Cook M (2003) *A Brief History of the Human Race* (Granta Books, London).
38. Mathew S, Boyd R (2011) Punishment sustains large-scale cooperation in prestate warfare. *Proc Natl Acad Sci USA* 108:11375–11380.
39. Dunbar RIM (1995) *The Trouble with Science* (Fabe & Faber, London).
40. Collins R (1994) *Sociological Forum* (Springer, Berlin), Vol 9, pp 155–177.
41. Wright HT (1977) Recent research on the origin of the state. *Annu Rev Anthropol* 6:379–397.
42. Crumley CL (1995) *Heterarchy and the Analysis of Complex Societies*, eds Ehrenreich RM, Crumley CL, Levy JE (Archaeological Papers of the American Anthropological Association, Arlington, VA), Vol 6, pp 1–5.
43. Blanton R, Fargher L (2008) *Collective Action in the Formation of Pre-Modern States* (Springer, Berlin).
44. Carballo DM, Roscoe P, Feinman GM (2014) Cooperation and collective action in the cultural evolution of complex societies. *J Archaeol Method Theory* 21:98–133.
45. Bondarenko DM, Grinin LE, Korotayev AV (2002) Alternative pathways of social evolution. *Soc Evol Hist* 1:54.
46. Bellah RN (2011) *Religion in Human Evolution: From the Paleolithic to the Axial Age* (Harvard Univ Press, Cambridge, MA).
47. Turchin P (2016) *Ultrasociety: How 10,000 Years of War Made Humans the Greatest Cooperators on Earth* (Beresta Books, Chaplin, CT).
48. Wright HT (1986) *American Archeology Past and Future*, ed Meltzer DJ (Smithsonian Institute, Washington, DC), pp 323–365.
49. Drennan RD (1991) Pre-hispanic chiefdom trajectories in Mesoamerica, Central America, and northern South America. *Chiefdoms: Power, Economy, and Ideology*, ed Earle TK (Cambridge Univ Press, Cambridge, UK), p 263.
50. Kirch PV (1984) *The Evolution of the Polynesian Chiefdoms*, New Studies in Archaeology (Cambridge Univ Press, Cambridge, UK).
51. Claessen HJM, Skalník P (1978) *The Early State* (Moulton Publishers, The Hague, The Netherlands).
52. Trigger BG (2003) *Understanding Early Civilizations: A Comparative Study* (Cambridge Univ Press, Cambridge, UK).
53. Johnson AW, Earle T (2000) *The Evolution of Human Societies* (Stanford Univ Press, Stanford, CA).
54. Flannery K, Marcus J (2012) *The Creation of Inequality: How Our Prehistoric Ancestors Set the Stage for Monarchy, Slavery, and Empire* (Harvard Univ Press, Cambridge, MA).
55. Spencer CS (1997) Evolutionary approaches in archaeology. *J Archaeol Res* 5:209–264.
56. Kirch PV, Green RC (1997) History, phylogeny, and evolution in Polynesia. *Curr Anthropol* 33:161–186.
57. Spencer CS (2009) *Macroevolution in Human Prehistory*, eds Prentiss AM, Kuijt I, Chatters JC (Springer, New York), pp 133–156.
58. McShea DW (1994) Mechanisms of large-scale evolutionary trends. *Evolution* 48:1747–1763.
59. McShea DW (2001) The minor transitions in hierarchical evolution and the question of a directional bias. *J Evol Biol* 14:502–518.
60. Carneiro RL (1970) A theory of the origin of the state: Traditional theories of state origins are considered and rejected in favor of a new ecological hypothesis. *Science* 169:733–738.
61. Currie TE, et al. (2015) Agricultural productivity in past societies: Toward an empirically informed model for testing cultural evolutionary hypotheses. *Cliodynamics* 6:24–56.
62. Whitehouse H, François P, Turchin P (2015) The role of ritual in the evolution of social complexity: Five predictions and a drum roll. *Cliodynamics* 6:199–216.
63. Rubin DB (1987) *Multiple Imputation for Nonresponse in Surveys* (Wiley, New York).
64. Yuan Y (2000) *Multiple Imputation for Missing Data: Concepts and New Developments* (SAS Institute, Rockville, MD), p 267.
65. Turchin P (2003) *Complex Population Dynamics: A Theoretical/Empirical Synthesis* (Princeton Univ Press, Princeton).
66. Burnham KP, Anderson DR (2002) *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach* (Springer, New York).
67. R Core Team (2015) R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna).

Supporting Information

Turchin et al. 10.1073/pnas.1708800115
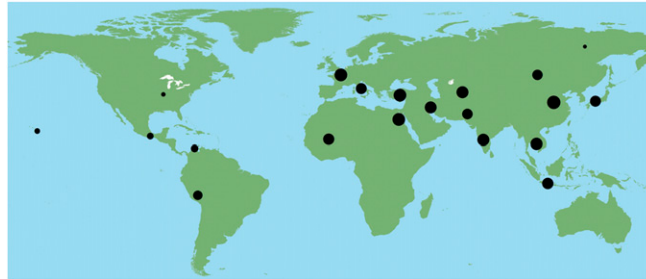
Movie S1.   The dynamics of social complexity through space and time. PC1 values mapped to our 30 NGAs for the time periods for which we have data. Larger dots indicate larger PC1 values (i.e., higher social complexity). Data presented in this video reflect the state of the database in March 2017.

[Movie S1](#)

Dataset S1.   Information on complexity characteristics (CCs) for each NGA over time for twenty replicates of the multiple imputation procedure

[Dataset S1](#)

   CapPop, capital population; infrastr, infrastructure; levels, hierarchical levels; irep, number of multiple imputation replicate; NGA, natural geographic area; PolID, shortened polity name; PolPop, polity population; PolTerr, polity territory; Time, starting date from which CC data applies; writing, information systems.

Dataset S2.   Values for PC1 for each NGA over time for twenty replicates of the multiple imputation procedure (V1-20)

[Dataset S2](#)

   NGA, natural geographic area; PolID, shortened polity name; Time, starting date from which CC data applies.

Other Supporting Information Files

[SI Appendix (PDF)](#)

Quantitative Historical Analyses Uncover a Single Dimension of Complexity that Structures Global Variation in Human Social Organization

Peter Turchin, Thomas E. Currie, Harvey Whitehouse, Pieter Francois, Kevin Feeney, Daniel Mullins, Daniel Hoyer, Christina Collins, Stephanie Grohmann, Patrick Savage, Gavin Mendel-Gleason, Edward Turner, Agathe Dupeyron, Enrico Cioni, Jenny Reddish, Jill Levine, Greine Jordan, Eva Brandl, Alice Williams, Rudolph Cesaretti, Marta Krueger, Alessandro Ceccarelli, Joe Figliulo-Rosswurm, Po-Ju Tuan, Peter Peregrine, Arkadiusz Marciniak, Johannes Preiser-Kapeller, Nikolay Kradin, Andrey Korotayev, Alessio Palmisano, David Baker, Julye Bidmead, Peter Bol, David Christian, Connie Cook, Alan Covey, Gary Feinman, Árni Daníel Júlíusson, Axel Kristinsson, John Miksic, Ruth Mostern, Cameron Petrie, Peter Rudiak-Gould, Barend ter Haar, Vesna Wallace, Victor Mair, Liye Xie, John Baines, Elizabeth Bridges, Joseph Manning, Bruce Lockhart, Amy Bogaard, Charles Spencer

Supplementary Information

Supplementary Methods

Structure of Seshat: Global History Databank

***Background*:** Our collective knowledge about past societies is almost entirely in a form inaccessible to scientific analysis; stored in historians' brains or scattered over heterogeneous notes and publications. The huge potential of this knowledge for testing theories about political and economic development has been largely untapped. Founded in 2011, *Seshat: Global History Databank* brings together the most current and comprehensive body of knowledge about human history in one place (*1*). The Databank systematically collects what is currently known about the social and political organization of human societies and how they have evolved over time (*2*).
***Goal*:** The goal of *Seshat* is to enable researchers to conduct comparative analyses of human societies and rigorously test different hypotheses about the social and cultural evolution of societies across the globe and over long periods of human history.
***Time frame*:** Currently *Seshat* focuses on the time period between the Agricultural and Industrial Revolutions. The spatial reach is global, and eventually we plan to include in the Databank information on any past societies, up to the present, for which historical or archaeological data are available. However, reaching this goal will require many years and, as a first step, we analyze a sample of 30 locations across the globe, stratified by the world region and the antiquity of complex societies (see below). For each of the 30 global points we start at a period just before the Industrial Revolution (typically, 1800 or 1900 CE depending on the location) and go back in time to the Neolithic or equivalent period (subject to the limitation of data).
***Unit of analysis*:** Our unit of analysis is a *polity*, an independent political unit that ranges in scale from villages (local communities) through simple and complex chiefdoms to states and empires.
***Variables*:** In addition to the social complexity variables analyzed in this paper, we also code variables on warfare, religion and rituals, agriculture and resources, institutions, well-being, and the production of public goods. Overall, the current codebook includes 1500 variables. These variables are coded for any past polity that occupied one of our 30 world locations between the Neolithic and Industrial Revolutions. Currently there are 414 such polities in *Seshat*. As of September 2017, the Databank contains >200,000 coded values ("Seshat records", see below).

Systematic sampling of past societies
In order to assess whether different societies show commonalities in the way they have evolved we developed a geo-temporal, stratified sampling scheme to select the societies on which to collect data. We designed our sampling scheme with two goals in mind: 1) to include as much variation among the sampled societies as possible in terms of social organization, and 2) to ensure representation of different parts of the world. This issue is challenging as societies can expand or contract in geographical space, appear or disappear in the historical & archaeological records, and show varying degrees of continuity with earlier or later societies.
***Geographic sampling & Natural Geographic Areas (NGAs)*:** To overcome these issues and ensure that we collected data in a systematic manner we divided the world into ten major regions (see Figure SI1 and Table SI1). Within each region we selected three natural geographic areas (NGAs) to act as our basic geographical sampling unit. Each NGA is defined spatially by a boundary drawn on the world map that encloses an area delimited by naturally occurring geographical features (for example, river basins, coastal plains, valleys, and islands). The extent

of the NGAs does not change over time, and NGAs thus act as our fixed points which determine which societies we collected data for.

Stratification for maximizing variation in socio-political organization: Within each world region we looked for NGAs that would allow us to cover as wide a range of forms of social organization as possible. In effect we wanted to ensure that we captured information about the kinds of societies that researchers have previously discussed in relation to social complexity ("states", "chiefdoms", "stratified societies", "empires" etc.) without using typological definitions of such societies or employing a strong, limiting definition about what features such societies should have. We also wanted to make sure that we captured information about societies that are not traditionally thought of as complex ("small scale societies", "egalitarian tribes", "acephalous societies"). This approach enables us to assess whether the different features of these societies tend to co-occur and evolve in somewhat regular ways across time and space. Accordingly, within each world region one NGA was selected that saw some of the earliest developments of some kind of large-scale or centralized, stratified society that existing scholarship would refer to as a "complex society". We also chose another sampling point that was the opposite; ideally, it was free of such societies until the modern or colonial period. Finally, the third NGA was intermediate in terms of the time that political centralization emerged. Because different world regions acquired centralized societies at different times there can be substantial variation across 'early complexity' NGAs both in the time at which our measures of social complexity start increasing and the degree of social complexity that is eventually reached at the end of our sampling period. For example, Susiana, the early complexity NGA in Southwest Asia has much longer history of large societies than Hawaii, the early complexity NGA in the Pacific region.

Temporal sampling of polities: To populate the Databank, for each NGA we consulted the literature and chronologically listed all polities that were located in the NGA, or encompassed it. We chose a temporal sampling rate of one hundred years meaning that we only included polities that spanned a century mark (100AD, 200AD etc.) and omitted any polities of short duration that only inhabited an NGA between these points. This is short enough to capture meaningful changes in the social complexity of historical societies, but not too short to lead to *oversampled* data ("oversampling" results when the succeeding point in time contains the same information as the preceding one, thus not adding to the overall information content of the data set in terms of variability).

For those periods when the NGA is divided up among a multitude of small-scale polities (e.g., independent villages, or small chiefdoms) it is not feasible to code each individual polity. In such instances we use the concept of 'quasi-polity,' which is defined as a cultural area with some degree of cultural homogeneity that is distinct from surrounding areas and approximately corresponds to an ethnological "culture" (*3-5*) or an archaeological sub-tradition (*6*). We then collect data for each quasi-polity as a whole. This way we can integrate over (often patchy) data from different sites and different polities within the NGA to estimate what a 'generic' polity was like. This approach is especially useful for societies known only archaeologically, for which we usually don't know polity boundaries.

It is important to point out that our use of polities and quasi-polities is best understood as a means of sampling the vast literature on past human societies rather than trying to impose a rigid framework on the human past. Our data coding procedures enable us to capture changes in a particular variable within the lifetime of a polity and also allow us to capture variation within a

polity or quasi-polity where there is such evidence. We are also able to flexibly incorporate multiple lines of evidence and uncertainty as we outline below.
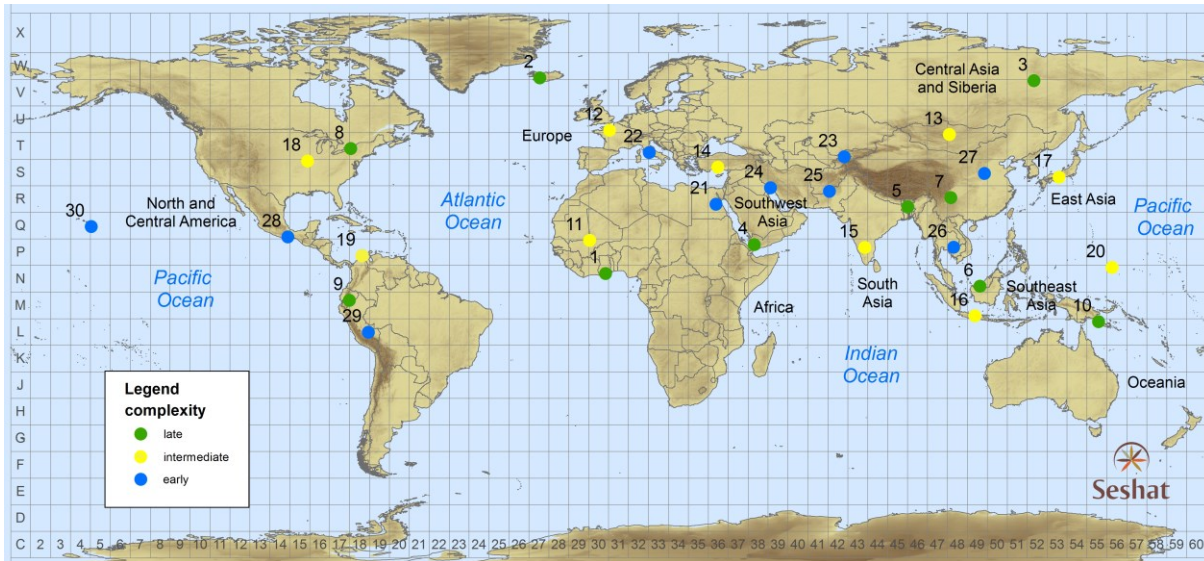
Figure SI1. *Locations of the 30 sampling points (Natural Geographic Areas) on the world map. For the key to the NGA numbers see Table SI1. Map adapted from (1).*

Table SI1. *The World Sample-30. The numbers of NGAs correspond to the numbers in Figures 1 and SI1.*

World Region	Late Complexity	Intermediate Complexity	Early Complexity
Africa	Ghanaian Coast (1)	Niger Inland Delta (11)	Upper Egypt (21)
Europe	Iceland (2)	Paris Basin (12)	Latium (22)
Central Eurasia	Lena River Valley (3)	Orkhon Valley (13)	Sogdiana (23)
Southwest Asia	Yemeni Coastal Plain (4)	Konya Plain (14)	Susiana (24)
South Asia	Garo Hills (5)	Deccan (15)	Kachi Plain (25)
Southeast Asia	Kapuasi Basin (6)	Central Java (16)	Cambodian Basin (26)
East Asia	Southern China Hills (7)	Kansai (17)	Middle Yellow River Valley (27)
North America	Finger Lakes (8)	Cahokia (18)	Valley of Oaxaca (28)
South America	Lowland Andes (9)	North Colombia (19)	Cuzco (29)
Oceania-Australia	Oro, PNG (10)	Chuuk Islands (20)	Big Island Hawaii (30)

Data Collection

Identifying social complexity variables and creating complexity characteristic measures:
Researchers from different disciplines have defined social complexity in different ways, each definition emphasizing different aspects, and with different measures being put forward to capture social complexity (*7-14*). As we stated in the Introduction, our approach is to be inclusive in that we make an attempt to code a variety of aspects of what different disciplines understand by social complexity, and attempted to be as "theory neutral" as possible in deciding on the list of variables to collect information on. In coming up with this list of variables we consulted a number of researchers who are historical and archaeological experts on societies from a variety of regions and time periods, and who represent a variety of theoretical persuasions. In total we identified c.70 variables relating to social complexity that could potentially be coded across different societies (see Codebook: http://seshatdatabank.info/methods/codebook/). Through our data collection process we found that some of these variables were easier to capture than others, or had information that was more widely recorded. For our final analyses we used information on the 51 variables that could reliably be identified and coded. The nature of the historical and archaeological records means that information can be patchy so we deliberately built some redundancy into our coding procedures meaning that different variables act as proxies for nine complexity characteristics. The first set of variables relates to the scale of societies: the total population of the polity, the extent of territory it controls, and the size of the largest urban center (Figure 1 of the main article). These variables were log-transformed prior to analysis.

Next come measures of hierarchical or vertical complexity ("levels of hierarchy" in Figure 1). These focus on the number of control/decision levels in the administrative, religious, and military hierarchies. Another measure of vertical complexity is the number of levels in the settlement hierarchy. The four hierarchical variables were averaged to yield the "levels of hierarchy" variable.

"Government" variables code for presence or absence of professional soldiers and officers, priests, bureaucrats, and judges. This class also includes characteristics of the bureaucracy and of the judicial system, and presence of specialized buildings (e.g., courts). Government variables were aggregated by adding the number of binary codes indicating "present" and dividing them by the total number of variables. The aggregated variable Government, thus is scaled between 0 and 1.

The variety of public goods and public works provided by the community is captured in "Infrastructure." Informational complexity is coded by the characteristics of the writing and record-keeping (more generally, informational) systems. We also record whether the society has developed specialized literature, including history, philosophy, and fiction. These binary codes were treated the same way as Government, yielding aggregated variables Infrastructure, Writing, and Texts (see Figure 1 of the main article).

Finally, the sophistication of the cash economy is reflected in Monetary System, which can take values between 0 and 6, reflecting the "most sophisticated" monetary instrument present in the coded society (Figure 1 in the main article). For example, if precious metals were used as money, while foreign and indigenous coins and paper currency were absent, Money would take the value of 3. If on the other hand, paper currency was present, the value of the aggregated variable is 6. Presence of "less sophisticated" instruments does not affect the value of Money.

It was not possible to code data for all variables for all polities (see below). For our final dataset we set a threshold that for a polity to be included 30% of the variables had to be coded (i.e., at least 16 of the 51 social complexity variables). This was to strike a balance between unnecessarily throwing away information by setting the threshold too high on the one hand, and including too many poorly covered polities that might create problems in the analysis stage on the other. We explored the effects of adjusting this threshold in confirmatory analyses below.

Data Coding Approach
Having identified the polities and quasi-polities, and defined our social complexity codebook data collection occurred in two phases. In Phase I research assistants searched published articles and books on a particular polity (often with advice from a regional or polity expert on what sources were likely to be most useful) in order to find information about each variable and enter it into the databank. In Phase II, where possible, experts on the polity, academic historians or archaeologists, went over the data to check coding decisions made by RAs and help us fill the gaps. Experts also indicate when the value should be coded as "unknown." When two or more experts disagree about the value or there is ongoing debate in the literature, all choices are entered as alternatives. For quantitative variables whose values are known only approximately, coders are instructed to enter a likely range [min, max] that roughly corresponds to a 90 percent confidence interval (i.e., omitting possible, but unlikely or unrepresentative values).

We refer to a coded value for a particular variable for a particular polity as a "Seshat record." Seshat records have complex internal structure. First, there is the value of the coded variable. For a numerical variable the value can be either a point estimate, or a range approximating the 90-percent confidence interval. Binary variables can take the following values: present, absent, inferred present, inferred absent, and unknown (a numerical variable can also be coded as unknown). "Inferred" presence or absence indicates some degree of uncertainty: when direct evidence of presence (for example) is lacking, but the expert can confidently infer it. For example, if iron smelting has been attested both for the period preceding the one that is coded, and for the subsequent period, we code it as "inferred present" even though there is no direct evidence for it (assuming there are no indications that this technology was lost and then regained). To incorporate this uncertainty into our analyses an inferred present codings is given a value of 0.9 (rather than 1), and and inferred absent is given a value of 0.1 (rather than 0).

Binary variables can also have temporal uncertainty associated with them. For example, if we know that iron smelting appeared in the NGA at some point between 300 and 600 CE, we code period previous to 300 CE as absent, the period following 600 CE as present, and the period between 300 and 600 CE as effectively "either absent, or present".

As mentioned above, *Seshat* also reflects disagreements among the experts. When two or more experts propose different values for the same variable, all are entered. These values can also contain uncertainty. For example, a Seshat record may state that the population of a particular polity at 300 BCE was either between 30,000 and 40,000 people (according to Expert I) or between 60,000 and 120,000 (according to Expert II).

The second important part of a Seshat record is a narrative explaining why this particular variable was coded in this particular way. Typically, this narrative is first written by an RA, who may quote the relevant text from a reference (a book or an article). The narrative is then checked and edited by experts as needed. Subsequent experts can add to it and disagree with previously recorded estimates.

The third part of a Seshat record is the references to publications or other databases. As not all the knowledge that can be brought to bear on these issues is necessarily in the literature a

reference can also be attributed to an expert with knowledge of the polity. In such cases the expert makes a judgment on the coding themselves and provides a justification.

We expect that Seshat records will evolve as more experts are involved in checking them, and as new insights or evidence are produced by academic historians and archaeologists. As such changes occur, they do not simply overwrite the previous information; instead, the Databank stores these changes so that the evolution of any record can be examined at any later time. This feature of Seshat Databank ensures continuity and accumulation of knowledge. It also identifies gaps in our knowledge, where a lack of evidence prevents us from being certain about features of societies in the past.

Data Availability

We have created a website (http://seshatdatabank.info/) that illustrates the Seshat and shows how entries in the databank are supported by references, and explanations & justifications of the codes. The full set of NGAs with information on the social complexity variables is open access as an accompaniment to this publication. The databank is continually expanding and new variables are being added in order to address other research questions. All data in Seshat will eventually be made open access a certain period after data collection and analysis, creating a unique resource for building and sharing knowledge about the human past.

Data on complexity characteristics and principal components are available as online supplementary files (SI Datasets S1, S2). In future, researchers will also be able to download updated or expanded versions of the databank from the website above as text files suitable for analysis and reuse.

Supplementary Results

Cross-Validation

Predictability of variables: *K*-fold cross-validation was applied to the subset of data in which all rows lacked missing values ($n = 203$). For these cross-validation analyses where there was a range of estimates, we used the midpoint; similarly, we took an average of values where experts disagreed. Cross-validation results indicate that regression models can predict all variables much better than the mean (Table 1, Table SI2), with overall predictability (ρ^2) varying between 0.53 and 0.84. Overall ρ^2 in Table 1 and the values for all regions in bold in Table SI2 are calculated as an average of the ρ^2 values weighted by the number of polities from which they are drawn.

Predictability between regions: Overall different world regions are well predicted by the relationships between variables observed in other world regions (Table 1, Table SI2). However, there is some degree of variability between world regions in this respect. In several regions (Africa, Central Eurasia, East Asia, North America, and Southwest Asia) regression models predict all variables better than the mean (no negative prediction ρ^2). For Europe, only writing has a negative ρ^2. Other regions (Oceania-Australia, South America, and Southeast Asia) have between two and four negative ρ^2s (Table SI2). However, these are the same regions that have

7

very few complete observations ($n < 10$). Thus, it appears that the probability that variables are not well-predicted is a function of the sample size being predicted (Fig SI2), rather than some sort of difference between the predicted region and the "global norm". The fact that Europe, Central Eurasia, East Asia, Africa and Southwest Asia are generally well-predicted supports the hypothesis that these different aspects of complexity are functionally linked and co-evolve together. However, an alternative explanation is that the co-occurrence of traits may simply be due to the fact that these regions have been historically connected and traits have tended to spread between them. Going against this alternative explanation it should be noted that North America is also well predicted by models built on the data from those other regions, even though it developed largely in isolation from other world regions prior to 1500 CE.

Table SI2. *Prediction ρ^2 as estimated by Cross-Validation. Italics mark world regions with $n < 10$ observations. "New World" refers to the combined results of predicting North American, South American, and Oceania-Australia polities by fitting regression models on "Old World" polities (see "confirmatory analyses").*

Region	PolPop	PolTerr	CapPop	Levels	gov't	infra	writing	Texts	money	n
All Regions	**0.84**	**0.76**	**0.71**	**0.60**	**0.53**	**0.62**	**0.59**	**0.73**	**0.53**	**203**
Africa	0.90	0.89	0.72	0.68	0.63	0.57	0.73	0.83	0.37	41
Central Eurasia	0.64	0.34	0.63	0.29	0.42	0.76	-0.38	0.86	0.76	9
East Asia	0.84	0.70	0.77	0.30	0.65	0.70	0.73	0.93	0.37	34
Europe	0.84	0.68	0.69	0.57	0.40	0.53	-0.31	0.36	0.20	43
North America	0.92	0.92	0.80	0.97	0.91	0.84	0.79	0.96	0.89	11
Oceania–Australia	0.92	-0.46	0.97	0.74	-3.21	0.60	-2.60	0.14	-1.69	1
South America	0.97	0.95	0.78	0.59	-4.15	-24.57	0.89	0.48	0.74	5
South Asia	0.56	0.46	0.69	-0.05	0.62	0.69	0.40	0.46	0.46	12
Southeast Asia	-0.35	-4.27	0.30	0.60	0.08	-0.25	0.47	0.91	-1.15	8
Southwest Asia	0.79	0.75	0.72	0.57	0.35	0.78	0.19	0.71	0.58	39
"New World"	0.92	0.92	0.81	0.89	0.50	0.71	0.72	0.81	0.82	17
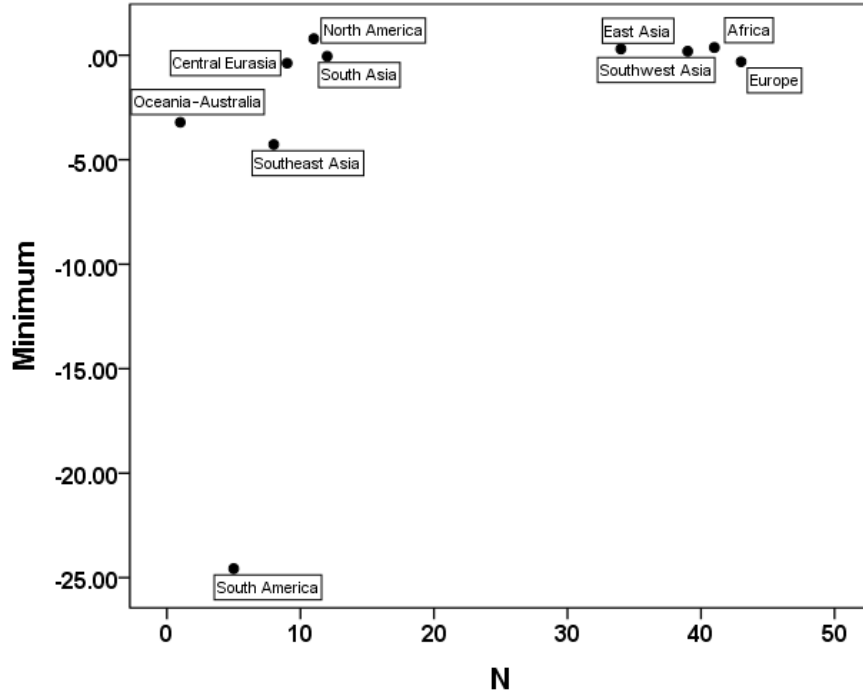
Fig SI2. *Minimum (Min) Prediction ρ² shows a positive relationship with number of polities (N) being predicted. This indicates that with fewer polities to predict the chances of obtaining low levels is greater, as general relationships have less opportunity to be observed.*

Table SI3. *Optimal number of predictor variables varies across response variables. Significant predictor variables in these minimum adequate models are indicated by asterisks.*

	Predictor Variables								
Response	*PolPop*	*PolTerr*	*CapPop*	*levels*	*gov't*	*infra*	*Writing*	*texts*	*money*
PolPop		*	*	*			*		
PolTerr	*								
CapPop	*			*					
levels	*		*			*	*		
gov't						*	*	*	
infra				*	*			*	*
writing	*			*	*			*	*
texts					*	*	*		
money						*	*		

9

Assessing optimal number of predictor variables: In earlier analyses we assessed the number of predictor variables required in the minimum adequate models. The optimal number of predictor variables needed to predict the response variable varied from one (PolTerr) to as many as five (writing, see Table SI3).

A general result of the cross-validation analysis is that it confirms that there is enough information within the dataset to allow internal prediction, which is the basis for the method of multiple imputation. We now turn to the results of multiple imputation for principal component analysis.

Principal components analysis based on multiple imputation

Principal Components Analyses were conducted on 20 imputed datasets. Below we report mean values from across these datasets and 95% confidence intervals.

All nine CCs were highly and significantly correlated with each other. Correlation coefficients varied between 0.49 (government and writing) and 0.88 (polity population and polity territory). Only a single principal component, PC1, has an eigenvalue greater than 1 (Table SI4, Fig SI3 – analyses conducted in SPSS). It explains 77.2±0.4 percent of variance. The proportion of variance explained by other principal components drops rapidly towards zero (e.g. PC2 explains only 6.0±0.4 percent). Furthermore, when we examine the "loadings" of the nine variables on PC1 (correlations between raw variables and PCs), we observe that all variables contribute about equally to PC1 (Figure SI5, Table SI5). Loadings on PC2 (Table SI5) seem to capture a slight residual but negative relationship between "social scale" variables (capital and polity population, hierarchical levels, and polity territory) and informational/economic complexity (writing, texts, and money). This could reflect cases where these features have diffused from large-scale societies where they were originally developed to smaller societies, or cases where large-societies have reduced in size but still retained these features. However, it should be emphasized that PC2 is not well-supported so we should be careful not to over-interpret this result. Overall, these results support the idea that different aspects of social organization have co-evolved in predictable ways, and that social complexity is a concept that can be well-represented by a measure such as PC1.

Table SI4. *Eigenvalues (means and standard deviations) for PCAs based on 20 imputed datasets. Only PC1 has an eigenvalue above the standard threshold of 1.*

| PC	Eigenvalue mean	Eigenvalue STD
1	6.95	0.02
2	0.54	0.02
3	0.42	0.01
4	0.35	0.01
5	0.22	0.01
6	0.20	0.00
7	0.15	0.01
8	0.09	0.00
9	0.07	0.00
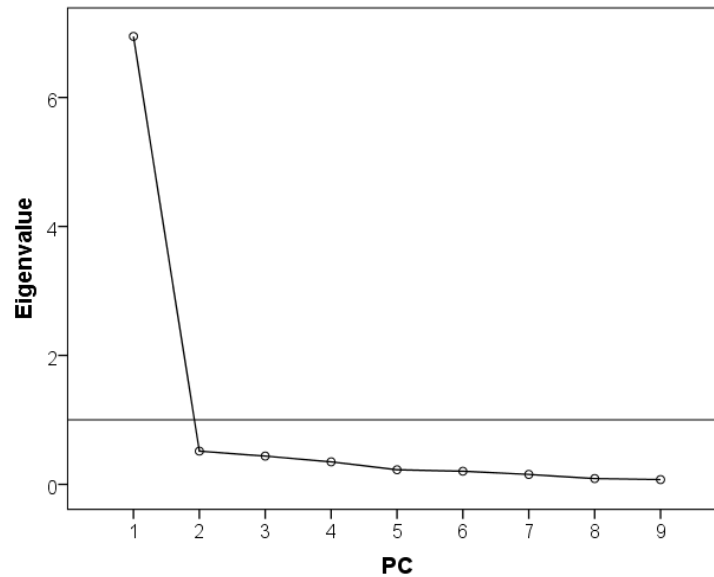
Fig SI3. *Scree plot of mean eigenvalues for PCAs based on 20 imputed datasets. Only PC1 has an eigenvalue above the standard threshold of 1. The PCs also show a characteristic elbow after the first PC.*

Table SI5. *Pairwise correlations between the nine Complexity Components and the first two Principal Components (PC1 and PC2). Correlations between variables were calculated only for cases in which there was no missing data (n=203). Correlations between variables and PCs calculated using full dataset with imputed values (n=414).*

	PolPop	PolTerr	CapPop	levels	gvrnmt	infrastr	writing	texts	money
PolPop		0.88	0.85	0.78	0.66	0.75	0.66	0.77	0.66
PolTerr	0.88		0.74	0.70	0.59	0.63	0.61	0.69	0.54
CapPop	0.85	0.74		0.76	0.61	0.75	0.56	0.71	0.59
levels	0.78	0.70	0.76		0.68	0.72	0.51	0.69	0.63
gvrnmt	0.66	0.59	0.61	0.68		0.74	0.49	0.67	0.57
infrastr	0.75	0.63	0.75	0.72	0.74		0.57	0.76	0.71
writing	0.66	0.61	0.56	0.51	0.49	0.57		0.80	0.65
Texts	0.77	0.69	0.71	0.69	0.67	0.76	0.80		0.70
money	0.66	0.54	0.59	0.63	0.57	0.71	0.65	0.70	
PC1	0.95	0.87	0.91	0.89	0.88	0.88	0.86	0.92	0.79
PC2*	-0.23	-0.30	-0.27	-0.21	0.08	0.09	0.33	0.23	0.34

* PC2 did not produce an eigenvalue greater than 1 in the PCA analysis and its importance should not be over-interpreted
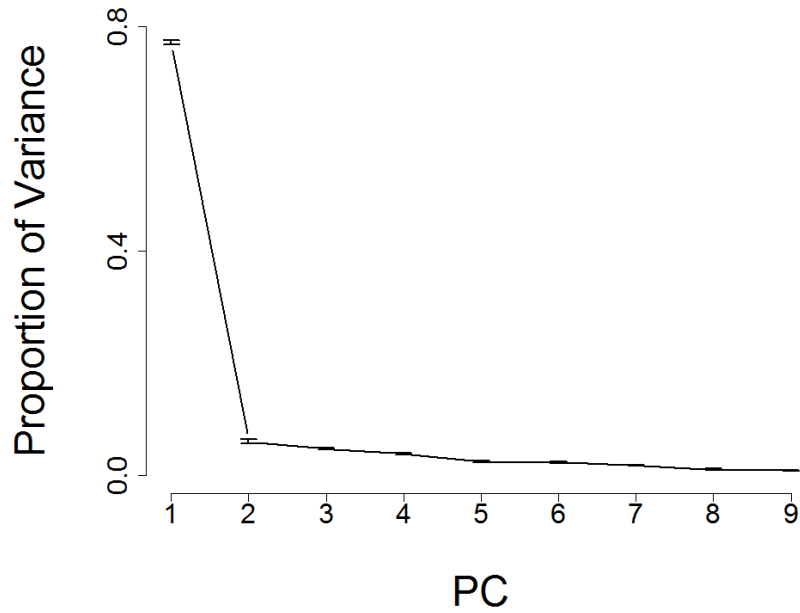
Figure SI4. *Proportion of variance explained by each principal component from the PCA. Error bars indicate 95% confidence intervals.*
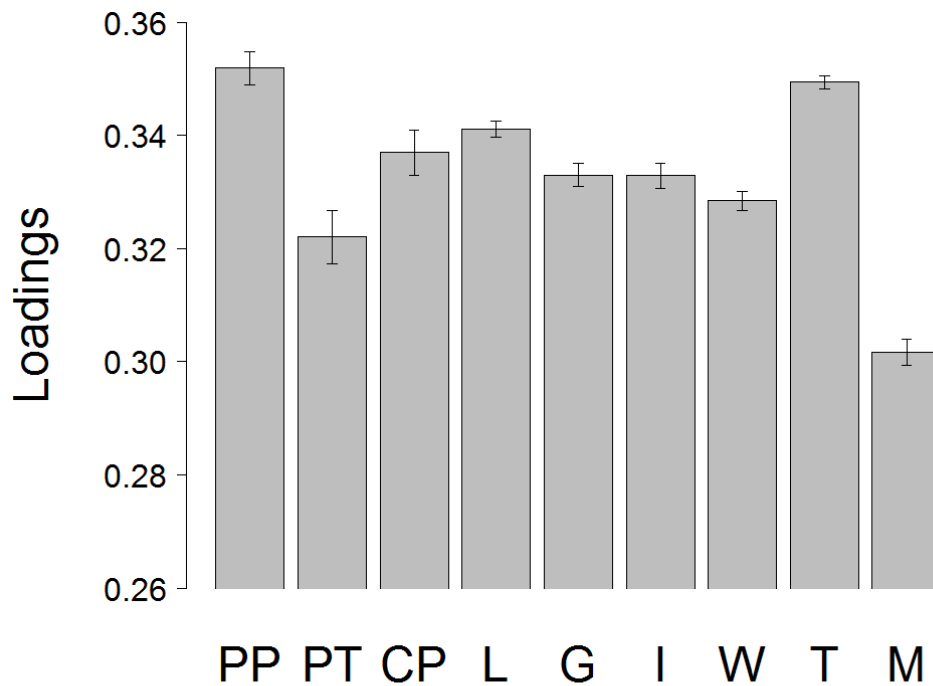

Figure SI5. *CC loadings on PC1. Error bars indicate 95% confidence intervals (note the restricted range on the Y axis, indicating that any differences in loadings between CCs are relatively small).*

Social Complexity Trajectories

Below we present the values for PC1 plotted through time for each NGA (Figure SI6). We have grouped NGAs into their world regions to aid comparisons. The values for PC1 have been rescaled here so that they go from 0 ("low complexity") to 10 ("high complexity"), to assist in interpretation. The broken thin lines indicate 95% confidence intervals. The more missing values, which have to be imputed, and the more uncertainty and disagreement, the wider the confidence interval is. These trajectories are mapped geographically and shown simultaneously in a video (Movie S1).

In line with previous analyses of social evolution, trajectories of the polities in each of our world regions show an overall increase in social complexity (*6, 13*), but also show episodic declines (*15-17*). Additionally, many other interesting features are revealed by these trajectories. For example, the trajectory for Latium (modern day Rome in the "Europe" world region) shows a fairly straightforward pattern and reflects a significant increase of complexity in the early Iron Age (tenth-ninth century BC), an apex during the early-middle Imperial Period (first-third century AD) and a dramatic decrease after the fall of the Roman Empire (476 AD). In the Konya Plain in Southwest Asia there are several of these increases and decreases, but still with an overall upward trend. The trajectory of social complexity dramatically increases at the beginning of the Early Bronze Age (3000 BC), and reaches the peak during the Hittite (1600-1200 BC), the Achaemenid (500-330 BC), the Roman (1-330 AD), the Byzantine (330-1000 AD), and The Ottoman Empire (1453-1922 AD). This also illustrates how the polities that inhabit and control our NGAs (and are thus included in our dataset) may actually originate from outside the NGA. In Susiana, also in Southwest Asia, social complexity significantly increases at the beginning of Susa II period (3800 BCE), and reaches the highest point during the Achaemenid (559-330 BC), the Seleucid (312-63 BC), and the Sasanian Empire (224-651 AD), and with several other fluctuations. The sometimes dramatic increase in social complexity seen at various points in these trajectories could be evidence for the idea that social organization evolves in punctuated bursts, as societies restructure and new forms of organization emerge over relatively short periods (*18-20*). However, in our data, some of these observed changes may actually be due to more complex societies from other regions conquering the NGA, as the Konya case illustrates. This idea will be formally tested in future work, using statistical techniques to test between competing hypotheses about the mode and tempo of social evolution (*19, 20*).
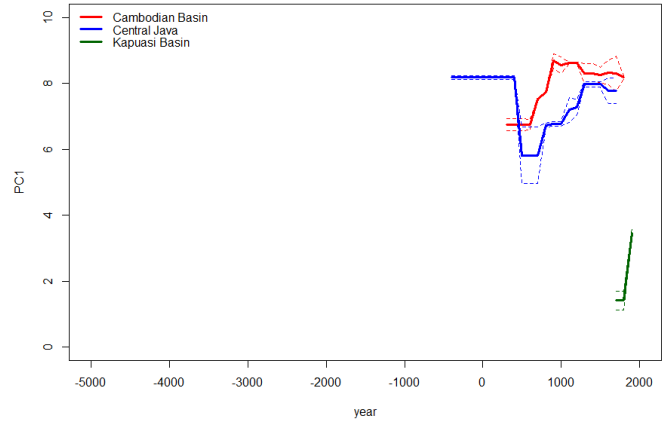
14

Figure SI6 *Evolutionary trajectories of our PC1 variable for each NGA within each of our 10 world regions*
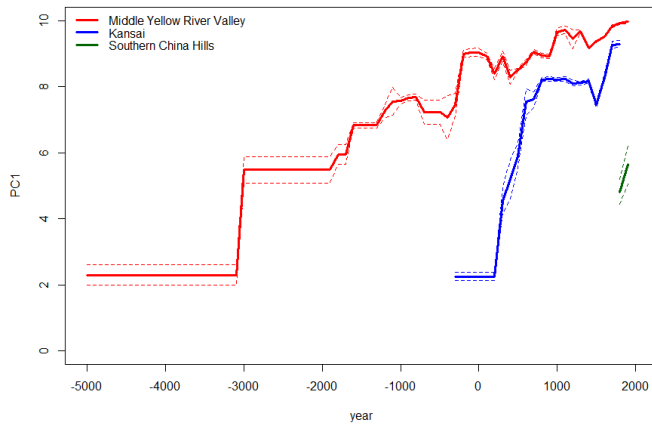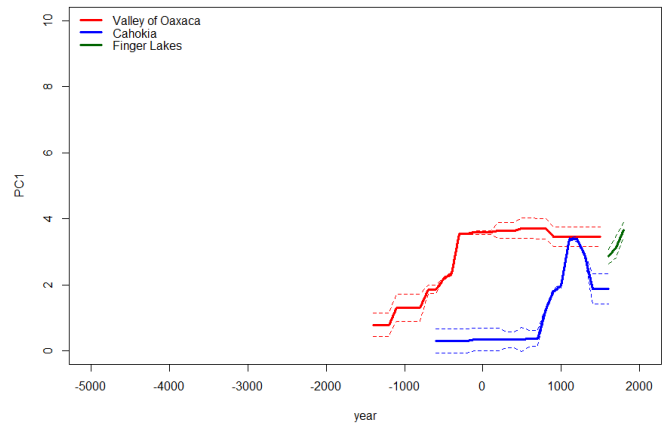

AFRICA

EUROPE

CENTRAL ASIA

SOUTHWEST ASIA
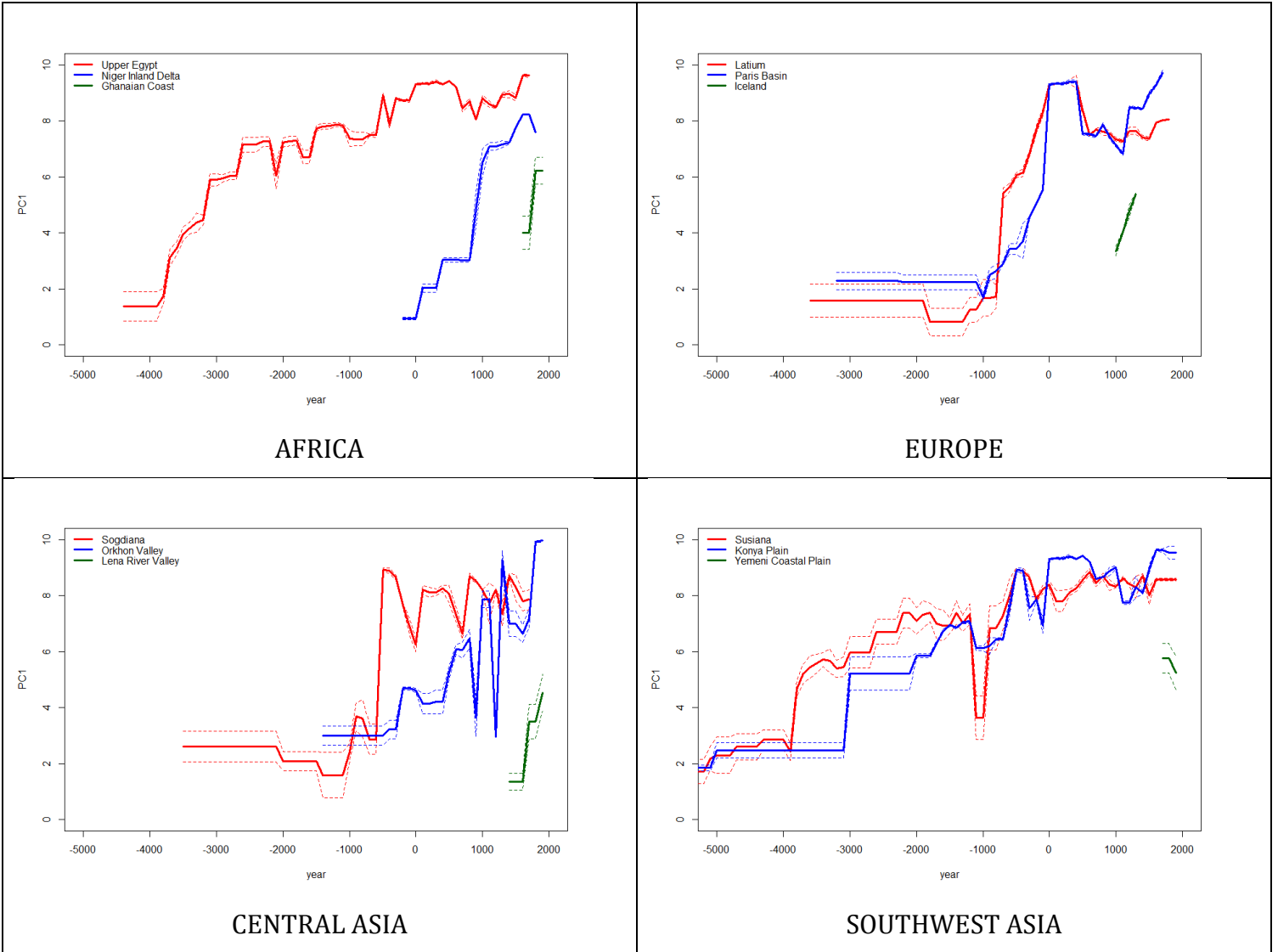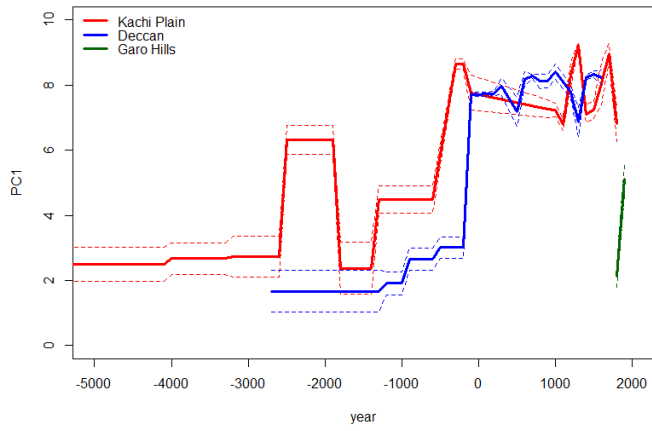
SOUTH ASIA
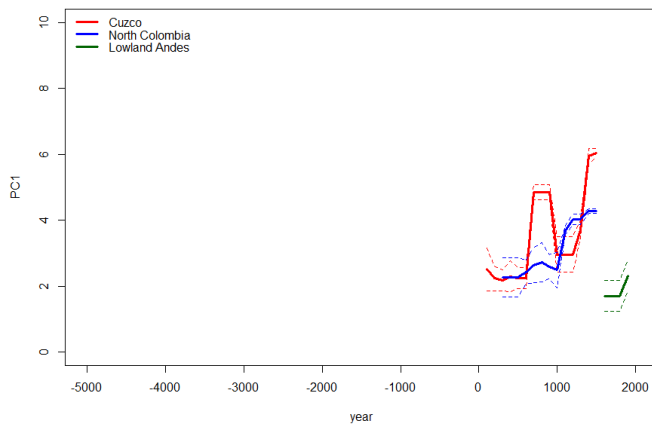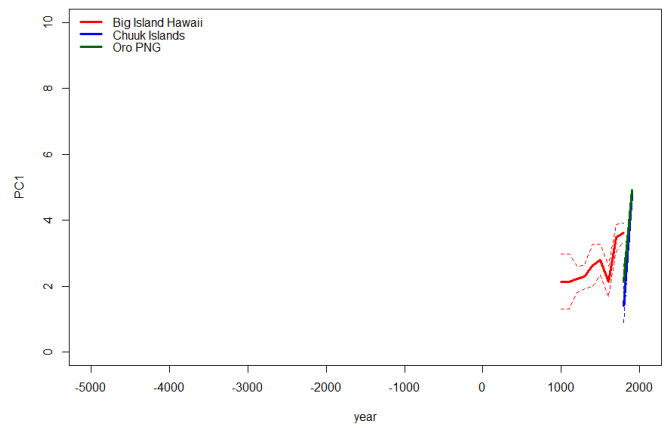
SOUTHEAST ASIA

EAST ASIA

NORTH AMERICA

SOUTH AMERICA

OCEANIA

Confirmatory analyses

In order to produce our main results we had to make a number of decisions and assumptions about how to conduct our statistical analyses. We have performed a number of confirmatory analyses to check the robustness of these results and show the effects of making different assumptions or decisions.

Adjusting the inclusion threshold

In earlier analyses we tested the effects of using different inclusion thresholds (our chosen default value being 30%). We tested the effects of performing PCA on datasets using 10%, 50% , and 100% (i.e. only cases with complete codings) coverage thresholds (in the latter case multiple imputation was not required to impute missing values). Adjusting the inclusion threshold had little effect on the proportion of variance explained by PC1: 10% cutoff – n=409, r^2=0.76; 50% cutoff - n=409, r^2=0.77; 100% - n=205, r^2=70.6). Our results are therefore not an artefact of either our inclusion threshold, or the multiple imputation procedure.

Accounting for sampling biases

In our main dataset some NGAs have a greater coverage than others due to differences in the timing of the beginnings of agriculture in different regions, and the level of research effort that has previously gone in to studying different regions of the world. Although we have attempted to offset some of these biases through our stratified sampling approach, there remains the possibility that parameter estimates from our results may be biased due to uneven coverage. We therefore used bootstrap resampling to create random sub-samples that lead to more balanced datasets. We did this in two ways: 1) Our analysis treats individual polities that span multiple centuries into separate polities for each century. Therefore, for any given polity that produced identical entries across centuries we resampled to produce only one entry per polity. 2) To ensure even geographic coverage, we resampled 10 polities per world NGA. If our main results are due to an overrepresentation of certain NGAs we would expect to see a large drop in the percentage of variance explained by PC1 in these confirmatory analyses. Sampling of one entry per polity had almost no effect on the proportion of variance explained by PC1 (n=285, r^2=0.79), and resampling of 10 polities per NGA only resulted in a relatively small drop in the proportion of variance explained by PC1 (n=300, r^2=0.69).

As a broader check on whether our findings have been driven by a bias to data availability in Africa and Eurasia (the "Old World") we fit models on data from the Old World and predict the remaining three regions (North America, South America, and Oceania-Australia, or the "New World"). Even though the New World polities developed without contacts with the Old World polities, they are highly predictable, with coefficients of prediction ranging between 0.5 and 0.92 (Table SI2). Prediction in the opposite direction is more problematic due to the smaller number of societies and the smaller range of variation. Scale variables are predictable at 0.29-0.77, but all other variables (except money) produced negative values. Rather than being indicative of

different co-evolutionary processes occurring in the Old and New World the lower predictability based on the New World is likely to be due to the smaller amount of variance in our complexity characteristics found in this part of the world. This difference in degree of variance in complexity characteristics between the Old and New world itself could result from differences in the rate at which innovations in the non-scale complexity characteristics that could support large scale societies were developed. It is well known that centres of the emergence of large-scale societies were more isolated from one another in the New World, and large-scale societies there did not spread as much as in the Old World until after European contact (*21*). In the New World the rate of adoption of depended on societies developing them independently. However, in the Old World the rate of adoption was elevated through being able to borrow such innovations from neighbouring societies (*22*). The fact that we can predict the values for complexity characteristics for the New World based on the values and relationships from the Old World suggests that similar co-evolutionary processes were indeed at play in both regions even if societies within these regions had their own unique evolutionary trajectories and specific histories.

Effects of variable choice

To assess whether our results are dependent on the particular variables combination of variables included in the analyses we ran two further sets of analyses: 1) we included only one population variable (polity population) and all the other non-population complexity characteristics, 2) we included only one population variable and removed one of the non-scale complexity characteristics (CC8:"texts"). The first analysis helps us assess whether the results are biased by the inclusion of several population variables, potentially leading to an overestimate of the importance of PC1 and an underestimate of any other dimensions of complexity. The second analysis helped us assess whether the inclusion of a particular non-scale variable, which could be argued to be more relevant to certain cultural traditions, was biasing our results. These additional analyses again had remarkably little impact on our findings: including only one population variable returns a single principal component that explains 78.7% (±0.4%) of the variance, while also removing "texts" returns a single principal component that explains 77.7% (±0.4%) of the variance (see tables SI6&SI7). These results are not that surprising as examining the loadings of the original PCA indicates that all variables load approximately equally onto PC1. Interestingly, the cross-validation analysis shows a slight reduction in the predictability of polity population when the other population variables are not included (table SI8). Again this is perfectly understandable from the previous results - the population variables have some of the strongest correlations with each other (which is, after all, why we are examining the effect of removing them) so removing two of them means the ability to predict polity population relies on the slightly weaker correlations with the other variables.

Overall, these confirmatory analyses suggest that our main findings are robust to the specific choices we have adopted for our analysis.

Table SI6. *Principal Component Analysis including only Polity Population as a direct population variable (analyses conducted in SPSS, variables included: PolPop, Levels, Government, Infrastructure, Information System, Texts, Money)*

Principal component	Eigenvalue mean	Eigenvalue STD	% of Variance mean	% of Variance STD
1	5.51	0.01	78.74	0.19
2	0.42	0.01	6.04	0.08
3	0.36	0.01	5.20	0.09
4	0.27	0.00	3.91	0.06
5	0.19	0.01	2.71	0.08
6	0.16	0.01	2.27	0.10
7	0.08	0.00	1.13	0.04

TableSI 7. *Principal Component Analysis removing the non-scale variable "texts" (analyses conducted in SPSS, variables included: PolPop, Levels, Government, Infrastructure, Information System, Money)*

Principal component	Eigenvalue mean	Eigenvalue STD	% of Variance mean	% of Variance STD
1	4.66	0.01	77.66	0.22
2	0.42	0.00	7.00	0.07
3	0.31	0.01	5.13	0.13
4	0.27	0.00	4.48	0.07
5	0.19	0.01	3.14	0.09
6	0.16	0.01	2.58	0.11

Table SI8. *Comparison of cross-validation (ρ^2) prediction scores from analyses with all variables and analyses that only include Polity Population as population variable.*

| CV Analysis	PolPop	PolTerr	CapPop	Levels	gov't	infra	writing	Texts	money
All population variables	0.84	0.76	0.71	0.60	0.53	0.62	0.59	0.73	0.53
Only Polity Population	0.66	-	-	0.63	0.53	0.58	0.59	0.75	0.53

Testing the Multiple Imputation Method

We have also assessed whether the multiple imputation method used in this study could have introduced bias into our results. The results above indicate that running analyses on cases that are fully coded does not substantially change parameter estimates or overall findings. At an earlier phase in our investigations to further examine this issue we created 100 artificial data sets that randomly introduced missing values into our "complete data set", reproducing the pattern of missing values in the "overall data set" at that time. We then applied the MI procedure to each of them. Each artificial data set was constructed as follows. We started with the first row of the complete data set. The program then chose a random row in the overall data set and determined if there were any missing values in the row. If yes, then missing values were added to the first row of the complete dataset for any variables that had missing data in the row from the overall data set. This procedure was repeated for the second row of the complete data set, and so on. The result was that the artificially constructed data set had the same pattern of missing values as the overall data set. The artificial data set was then subjected to the multiple imputation procedure in exactly the same way as the overall data were analyzed, except the results were based on 10 imputations, to speed up the calculations.

By comparing the PCA results based on the artificial dataset with results from the complete dataset, we see that the Multiple Imputation procedure accurately captures the overall patterns in the data both in terms of the number and pattern of PCs produced (Figure SI7), and the loadings of the different variables on to PC1 (Figure SI8). We repeated this analysis for 100 artificial data sets and compared the distribution of the proportion of variance explained by PC for both the estimated PC1s and the true PC1. The true value of PC1 is 0.706, while the mean and the mode of the estimated PC1, based on 100 artificial data sets with missing values, is 0.685 and 0.695, respectively (Figure SI9). The distribution is asymmetric, suggesting that the estimates based on MI procedure are biased—they tend to under-predict the true PC1. However, the degree of this bias is tiny (0.01 between the true value and the most likely estimate, the mode). Furthermore, the bias is conservative in that replacement of missing values by MI results in slightly under-predicting the true PC1). Taking these considerations together, we conclude that our overall MI procedure works very well for the goals of our study and has not created a bias that is driving our results and conclusions.
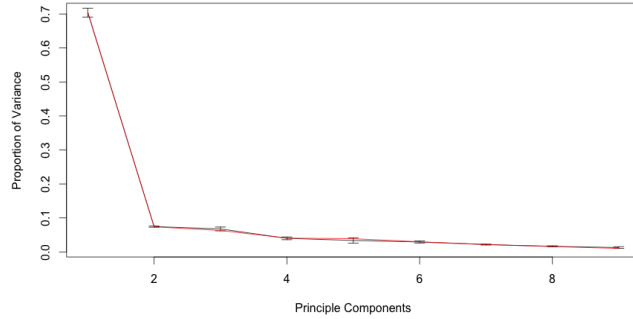
Figure SI7. *Comparison of factors extracted from the real dataset (red line), and the 10 artificial datasets, which had missing values added and then replaced via the multiple imputation procedure. Multiple imputation does not introduce a bias in the artificial datasets in comparison to the real dataset.*
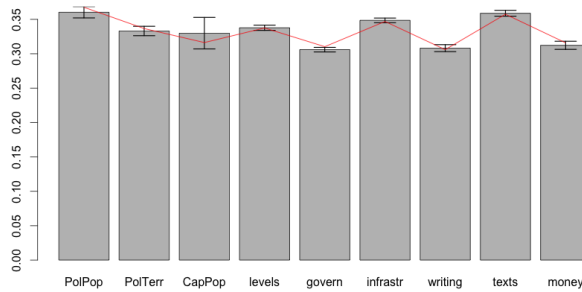
Figure SI8. *Comparison of variable loadings onto PC1 from the real dataset (red line), and the 10 artificial datasets, which had missing values added and then replaced via the multiple imputation procedure. Multiple imputation does not introduce a bias in the artificial datasets in comparison to the real dataset.*
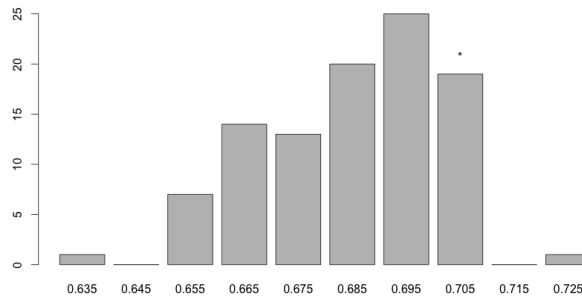
Figure SI9. *Distribution of proportion of variation explained by PC1 in 100 artificial datasets, which had missing values added and then replaced via the multiple imputation procedure. The true value for PC1 (indicated by an asterisk) is well within the range of estimates in the artificial datasets, and is only slightly higher than the modal value for the artificial datasets. Artificial datasets tend to lead to a slight underestimate of the true value for PC1.*

21

Supplementary Discussion

Sampling of NGAs and generality of findings

Currently our database contains 30 NGAs from which we have sampled the polities that controlled these areas over long periods of human history. Our decision to stratify our sampling efforts by broader world regions and onset of large-scale societies means we have extensive coverage of the diversity of human social and political organization. Limiting our sample to 30 NGAs was a necessary practical step in making headway in building a comparative historical database on this scale. Examining the maps (Fig 1, and SI1) it appears that at first glance there are large areas that do not have NGAs from which we have sampled polities. However, large parts of these areas (particularly in North America, South America, and Australia) were not inhabited by agriculturalists at the time of contact with Europeans and therefore are not our focus for constructing our database (which focuses primarily on post-paleolithic, agricultural polities). Other areas, particularly in more northern latitudes or desert areas, were sparsely populated and would therefore not add many extra polities. The choices made about which particular NGAs to choose was also partly based on practical concerns about having sufficient information and having the interest and availability of regional experts who were able to guide the process of data collection. In future it would be good to add further NGAs to regions such as sub-Saharan Africa that are currently relatively sparsely sampled. Collecting data for these areas faces important challenges due to the fact that traditionally such areas have received less academic attention than some other parts of the world, and in some cases there are substantial issues around preservation of archaeological remains (e.g. tropical rainforest soils are often not conducive to preservation).

We currently have no strong reason to suspect that the addition of agricultural polities from other areas would substantially alter our findings (however this remains an open possibility that can be addressed in future). For example, if we examine polities in a region such as North America, where social complexity emerged relatively late, we do not see a substantial difference in the relationships and trajectories we have been able to identify in other places (Fig SI6, Fig SI10,Table SI2). As we increase our coverage of polities in future we will be able to further examine the similarities and differences both within and between regions. Currently there are two areas where we have begun to intensify the coverage of NGAs: Meso-America (where we have begun coding polities in the Basin of Mexico and the Petén Basin [Mexico/Guatemala]), and Europe (where we have identified NGAs relating to the spread of the Neolithic). An interesting point of comparison in future studies may be to include more pastoral or hunter-gatherer societies to assess whether different modes of subsistence affects the patterns we have identified here. Differences in resource type can affect the way individuals are distributed in space, which may have consequences for the types of institutions that are effective for joining individuals and groups together. More generally, the fact that we have been able to detect consistent patterns in the evolution of social complexity indicates that, even though our coverage is not comprehensive, we are still able to uncover important principles that are applicable to wide variety of societies from differing cultural, historical, and ecological contexts.
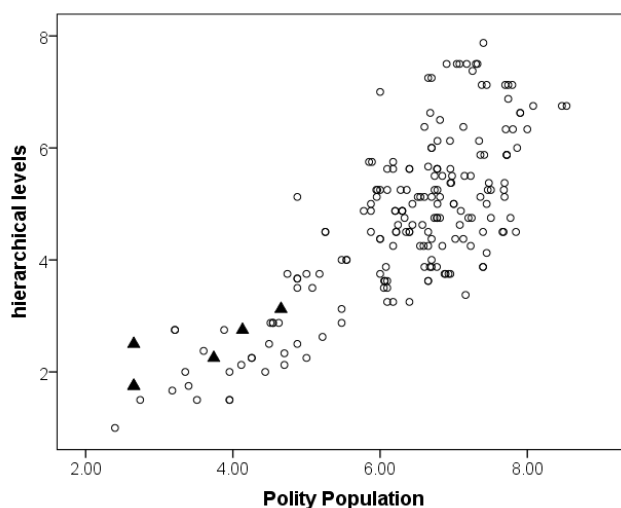
Figure SI10. Relationship between Hierarchy and Polity Population for medium and late complexity polities in the North America region (Cahokia and Finger Lakes NGAs, but not Oaxaca)(black triangles), and all other NGAs. The distribution of North American polities sits within the distribution of the other polities, and there are similar correlation coefficients between these variables for both sets of polities (North America: r=0.83, All other NGAs: r=0.76). This indicates that the North American polities did not evolve in a substantially different way from polities in other regions.

Testing Evolutionary Trend Mechanisms

Our approach is also well-suited to go beyond identifying patterns in socio-political evolution and investigate why social complexity has shown a tendency to increase over time. One idea that we can address by examining the temporal changes in our data is what kind of mechanism lies behind the trend towards increasing complexity. Evolutionary biologists distinguish between two types of general trend mechanisms: *passive* and *driven (23-25)*. This concept has also been applied in previous work to examine related issues around the evolution of socio-political complexity in human societies (*20, 26*). A passive trend relates to the fact that our starting point might be close to the lowest possible value (a "wall"), which means that there is more scope for change in one direction rather than another. Once away from the "wall" increases and decreases are equally likely. Over time the maximum level of complexity is expected to increase as this area of "trait space" can be expanded into, but there is nothing that particularly favours more complex organisms or societies. In a driven trend there is a force that actively favours larger values of the feature in question (i.e. more complex organisms/societies are at a selective advantage and a more likely to produce ancestors). The idea that larger, more complex societies have an advantage in competition between groups has a long history in anthropology, archaeology and related disciplines (*13, 27-32*). Diagnostic features of driven trends are that increases are more common than decreases, which is the case for almost all regions in our data

23

set (the average extent of increases is also generally greater - both in absolute terms and as a proportion of prior complexity)(Table SI9). Another way of distinguishing between trends is to examine how the distribution of the trait in question changes over time. In a passive trend the mean of the distribution increases due to the tail of the distribution increasing. In a driven trend, however, the mode of the distribution also increases, which seems to match what we see in our data (Figure SI11). Overall, the present results are consistent with the idea that competition between groups, particularly in the form of warfare, has been an important driving force in the emergence of large, complex societies. Future work will test competing ideas about the cause of this driven trend towards increased complexity.

Table SI9. *Number and extent of increases and decreases in complexity across regions. Values were calculated from differences between values in PC1 from one polity to the next.*

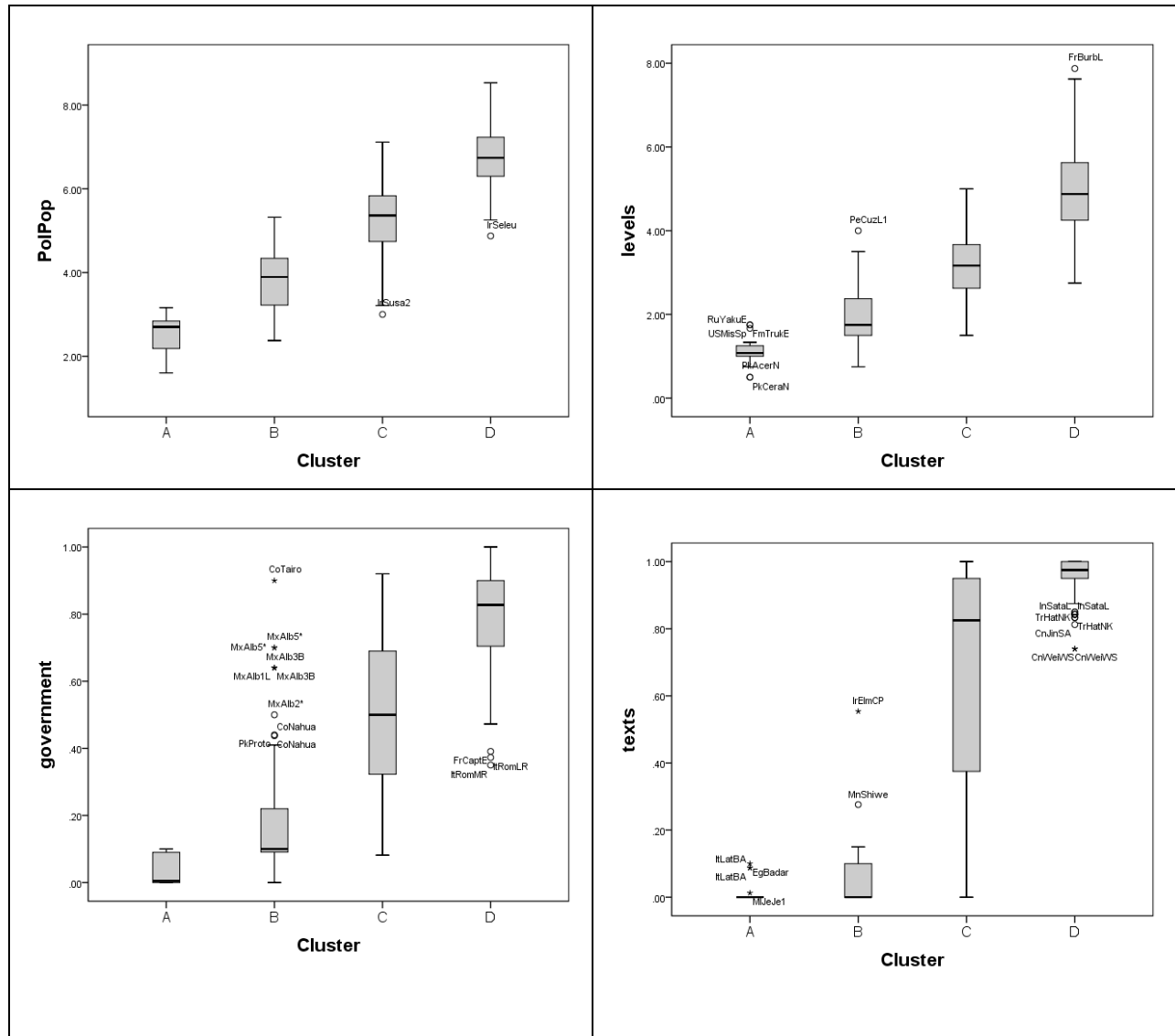| REGION	Inc	Dec	Net Change in PC1			Proportional Change in PC1		
			Min	Max	Mean	Min	Max	Mean
AFRICA	31	19	-1.67	2.59	0.28	-0.24	3.81	0.18
CENTRAL ASIA	19	21	-3.51	4.49	0.17	-0.46	1.17	0.10
EAST ASIA	38	12	-1.14	3.01	0.25	-0.15	1.88	0.08
EUROPE	36	18	-1.67	2.27	0.22	-0.68	1.23	0.08
NORTH AMERICA	17	7	-2.13	2.33	0.17	-1.00	6.67	0.37
OCEANIA	8	1	-0.84	2.60	0.43	-0.30	1.32	0.29
SOUTH AMERICA	11	7	-2.79	3.24	0.27	-0.46	1.13	0.12
SOUTH ASIA	27	11	-3.19	4.78	0.28	-0.56	1.45	0.14
SOUTHEAST ASIA	11	6	-0.64	2.35	0.33	-0.09	1.51	0.14
SOUTHWEST ASIA	50	34	-3.81	5.78	0.16	-0.58	2.07	0.08
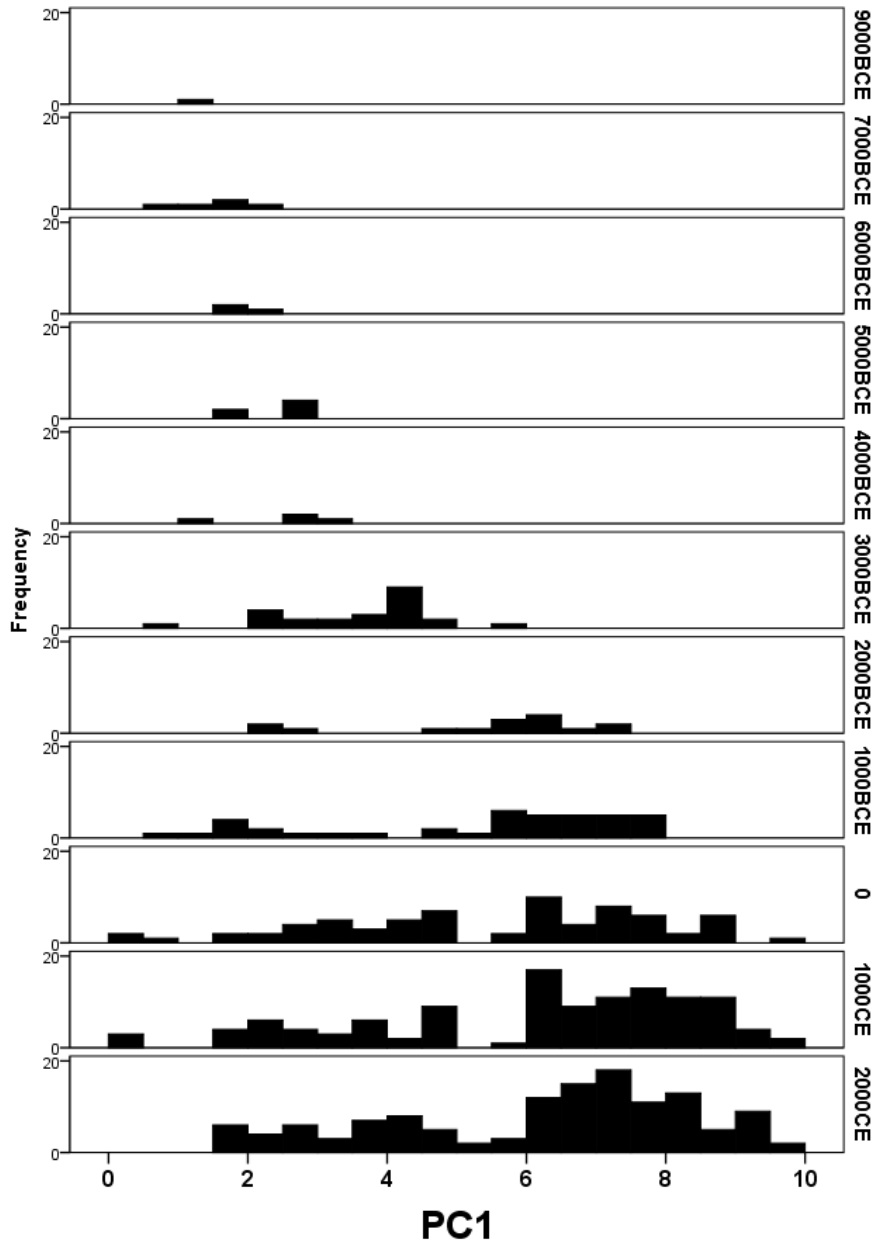
Figure SI11. *Changing distribution of Social Complexity over time. The mode of the distribution takes increasingly greater values over time, which is consistent with a driven evolutionary trend. Rows are 1000-year time slices, dates reflect upper date boundary (e.g. 7000BCE refers to 7999BCE to 7000BCE)*

Co-evolution, punctuated change, and "types" of socio-political organization

Our analyses indicate that are strong co-evolutionary relationships between different features of human societies. If correlated change in these features is relatively rapid, then certain "types" of socio-political organization may become apparent based on recurring associations between certain combinations of traits (*18-20*). Examining the evolutionary trajectories of the different NGAs (Fig 3, SI6) the data appear to show long periods of stasis or gradual, slow change, interspersed with sudden large increases in the measure of social complexity over a relatively short time span. This pattern is consistent with a punctuational model of social evolution, in which the evolution of larger polities requires a relatively rapid change in socio-political organization including the development of new governing institutions and social roles in order to be to stable (*18-20*). The assessment of these evolutionary rates will require more formal testing in future investigations and will need to take into account the fact that some periods of stasis indicate periods when the data from our polities show no change. This may reflect an absence of evidence and could potentially lead to errors in the assessment of rates of change. There do appear to be many instances of limited amounts of change prior, and many horizontal lines are of relatively short duration which would not substantially affect rate estimations. It therefore seems unlikely that these patterns are completely an artefact of the way our data are organized as a time series. Another consideration is the possibility that large changes in PC1 could indicate a new polity taking control of an NGA rather than the kind of change within a society that is envisioned under the punctuational change hypothesis.

To provide an initial assessment of the idea that societies may fall into certain types we conducted a Hierarchical Cluster analysis of PC1. The dendrogram in Figure SI12 shows some initial support for the idea of distinct "types", with a relatively large distance between two main clusters. This indicates a clear distinction between societies with large populations that exhibit many of the non-scale features of complexity, and smaller societies that often lack most of these features (see Fig SI12 (left)). Other potential groupings within these clusters may also indicate important stable combinations of traits that will be investigated in future research (Figure SI13). In line with previous empirical investigations of this question (*11, 19, 20, 33*), the clusters identified in these analyses may indicate that certain combinations of traits are indeed evolutionary stable. Typological schemes of human societies (e.g. band, tribe, chiefdom, state) have been common in studies of socio-political evolution (*13, 34-36*). These schemes have often been criticized, partly because the categories are seen as rigid and do not focus enough on how and why changes occur (see (*20*)). The clusters identified in the present analyses should not be thought of as strict categories as there remains substantial variation within each of the clusters. However, the present study further indicates how we can test hypotheses about the degree to which human groups exhibit the kind of patterned variation that traditional typological schemes are attempting to capture, or indeed test hypotheses about other kinds of patterned variation (*37-39*). Our historical, comparative approach illustrates how this can be done in a manner that enables us to empirically assess the degree of variation within and between categories, and can help in understanding how and why changes between categories can occur.

Fig SI12. *Hierarchical cluster analysis of PC1 from each polity, based on average linkage between groups (x-axis)(analysis conducted using SPSS)(right). Two main clusters are discernible due to the large average distances between them. Within each of these clusters two sub-clusters are identified (A&B and C&D). Values of Polity population and government (below) showed peaked distributions within each of the main clusters in line with the idea that these clusters represent distinct "types" of socio-political organization.*

Fig SI13. *Distributions of different CCs by the four main clusters identified above. Variation in characteristics such as polity population, hierarchical levels, and government seem to be well-summarized by the four clusters, where as a characteristic such as "texts" seems to be better summarized by just two clusters (main clusters 1(A&B) and 2(C&D)).*

Supplementary References

1. P. Turchin *et al.*, Seshat: The Global History Databank. *Cliodynamics: The Journal of Quantitative History and Cultural Evolution* **6**, (2015).
2. P. Francois *et al.*, A Macroscope for Global History: Seshat Global History Databank, a methodological overview. *Digital Humanities Quarterly*, (2016).
3. G. P. Murdock, Ethnographic Atlas: A Summary. *Ethnology* **6**, 109 (Apr., 1967).
4. G. P. Murdock, D. R. White, Standard Cross-Cultural Sample. *Ethnology* **9**, 329 (1969).
5. R. Naroll *et al.*, On Ethnic Unit Classification [and Comments and Reply]. *Current Anthropology* **5**, 283 (1964).
6. P. Peregrine, Atlas of Cultural Evolution. *World Cultures* **14**, 1 (2003).
7. G. P. Murdock, C. Provost, Measurement of Cultural Complexity. *Ethnology* **12**, 379 (1973).
8. G. Chick, Cultural Complexity: The Concept and Its Measurement. *Cross-Cultural Research* **31**, 275 (1997/11/01, 1997).
9. R. Naroll, A Preliminary Index of Social Development. *American Anthropologist*, (Jan 1, 1956).
10. R. L. Carneiro, in *A Handbook of Method in Cultural Anthropology,* R. Naroll, R. Cohen, Eds. (The Natural History Press, Garden City, N.Y., 1970), pp. 834-871.
11. P. N. Peregrine, C. R. Ember, M. Ember, Universal patterns in cultural evolution: An empirical analysis using Guttman scaling. *American Anthropologist* **106**, 145 (2004).
12. L. White, *The Evolution of Culture*. (McGraw-Hill, New York, 1959).
13. R. L. Carneiro, *Evolutionism in Cultural Anthropology*. (Westview Press, Boulder, Colorado, 2003).
14. I. Morris, *The Measure of Civilization: How Social Development Decides the Fate of Nations*. (Princeton University Press, 2013).
15. 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).
16. S. Gavrilets, D. G. Anderson, P. Turchin, Cycling in the Complexity of Early Societies. *Cliodynamics* **1**, (2010).
17. P. Turchin, *Historical dynamics: why states rise and fall*. (Princeton University Press, Princeton, NJ, 2003).
18. C. S. Spencer, On the Tempo and Mode of State Formation - Neoevolutionism Reconsidered. *Journal of Anthropological Archaeology* **9**, 1 (Mar, 1990).
19. C. S. Spencer, in *Macroevolution in Human Prehistory,* A. M. Prentiss, I. Kuijt, J. C. Chatters, Eds. (Springer, New York, 2009), pp. 133-156.
20. 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).
21. J. Diamond, *Guns, germs and Steel*. (Vintage, London, 1997).
22. P. J. Richerson, R. Boyd, in *The Origin of Human Social Institutions,* W. G. Runciman, Ed. (Oxford University Press, Oxford, 2001), pp. 197-234.
23. D. W. McShea, Mechanisms of Large-Scale Evolutionary Trends. *Evolution* **48**, 1747 (1994).
24. D. W. McShea, Trends, tools, and terminology. *Paleobiology* **26**, 330 (SUM, 2000).
25. S. J. Adamowicz, A. Purvis, M. A. Wills, Increasing morphological complexity in multiple parallel lineages of the Crustacea. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 4786 (2008).
26. C. S. Spencer, E. M. Redmond, Multilevel selection and political evolution in the Valley of Oaxaca, 500-100 BC. *Journal of Anthropological Archaeology* **20**, 195 (2001).
27. 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).
28. R. L. Carneiro, in *Origins of the State,* R. Cohen, E. R. Service, Eds. (Institute for the Study of Human Issues, Philadelphia, 1978), pp. 205-223.
29. E. M. Redmond, C. S. Spencer, Chiefdoms at the threshold: The competitive origins of the primary state. *Journal of Anthropological Archaeology* **31**, 22 (2012/03/01/, 2012).
30. P. Turchin, *War and Peace and War: The Life Cycles of Imperial Nations*. (Pi Press, NY, 2006).
31. C. Tilly, *The Formation of national States in Western Europe*. G. Ardant, Ed., Social Science Research Council . Committee on Comparative Politics (Princeton University Press, 1975), vol. 8.

32.   R. L. Carneiro, A Theory of Origin of State. *Science* **169**, 733 (1970).
33.   P. Peregrine, Power and legitimation: political strategies, typology, and cultural evolution. *The comparative archaeology of complex societies*, 165 (2012).
34.   K. V. Flannery, The cultural evolution of civilizations. *Annual Review of Ecology Evolution and Systematics* **3**, 399 (1972).
35.   E. R. Service, *Primitive Social Organization*. (Random House, New York, 1962).
36.   T. Earle, in *Companion Encyclopedia of Anthropology* T. Ingold, Ed. (Routledge, London, 1994), pp. 940-961.
37.   R. D. Drennan, C. E. Peterson, Patterned variation in prehistoric chiefdoms. *Proceedings Of The National Academy Of Sciences Of The United States Of America* **103**, 3960 (Mar 14, 2006).
38.   D. M. Bondarenko, L. E. Grinin, A. V. Korotayev, Alternative Pathways of Social Evolution *Social Evolution & History* **1**, 54 (2002).
39.   R. E. Blanton, G. M. Feinman, S. A. Kowalewski, P. N. Peregrine, A dual-processual theory for the evolution of Mesoamerican civilization. *Current Anthropology* **37**, 1 (1996).