LETTER

doi:10.1038/nature16447

Holocene shifts in the assembly of plant and animal communities implicate human impacts

S. Kathleen Lyons[1], Kathryn L. Amatangelo[2], Anna K. Behrensmeyer[1], Antoine Bercovici[1], Jessica L. Blois[3], Matt Davis[1,4], William A. DiMichele[1], Andrew Du[5], Jussi T. Eronen[6], J. Tyler Faith[7], Gary R. Graves[8,9], Nathan Jud[10,11], Conrad Labandeira[1,12,13], Cindy V. Looy[14], Brian McGill[15], Joshua H. Miller[16], David Patterson[5], Silvia Pineda-Munoz[17], Richard Potts[18], Brett Riddle[19], Rebecca Terry[20], Anikó Tóth[1], Werner Ulrich[21], Amelia Villaseñor[5], Scott Wing[1], Heidi Anderson[22], John Anderson[22], Donald Waller[23] & Nicholas J. Gotelli[24]

Understanding how ecological communities are organized and how they change through time is critical to predicting the effects of climate change[1]. Recent work documenting the co-occurrence structure of modern communities found that most significant species pairs co-occur less frequently than would be expected by chance[2,3]. However, little is known about how co-occurrence structure changes through time. Here we evaluate changes in plant and animal community organization over geological time by quantifying the co-occurrence structure of 359,896 unique taxon pairs in 80 assemblages spanning the past 300 million years. Co-occurrences of most taxon pairs were statistically random, but a significant fraction were spatially aggregated or segregated. Aggregated pairs dominated from the Carboniferous period (307 million years ago) to the early Holocene epoch (11,700 years before present), when there was a pronounced shift to more segregated pairs, a trend that continues in modern assemblages. The shift began during the Holocene and coincided with increasing human population size[4,5] and the spread of agriculture in North America[6,7]. Before the shift, an average of 64% of significant pairs were aggregated; after the shift, the average dropped to 37%. The organization of modern and late Holocene plant and animal assemblages differs fundamentally from that of assemblages over the past 300 million years that predate the large-scale impacts of humans. Our results suggest that the rules governing the assembly of communities have recently been changed by human activity.

How are plant and animal communities organized, and does their structure change through time? This question has dominated many decades of research on community assembly rules and is critical to charting the future of biodiversity[1]. Whereas most studies have described overall community structure with simple indices such as species richness[8] and average co-occurrence[3], some analyses categorize individual species pairs in assemblages as random, aggregated, or segregated[2,9]. Segregated species pairs may be generated by processes such as negative species interactions, distinct habitat preferences, and dispersal limitation. Aggregated species pairs may be generated by processes such as positive species interactions, shared habitat preferences, and concordant dispersal[2]. Recent meta-analyses document an excess of segregated species pairs in modern communities: most significant species pairs co-occur less frequently than would be expected by chance[2,10]. The relative dominance of segregated versus aggregated species pairs suggests an important role for biotic interactions such as competition and predation, habitat selectivity, and dispersal limitation in structuring modern communities.

Do the patterns of species segregation that characterize modern assemblages also hold in the fossil record, or is the present different? If there was a change, when did the modern condition arise? There are many examples from the fossil record of times of major reorganization in ecological communities, such as a shift in the complexity of marine invertebrate communities after the end-Permian mass extinction[11]. But even during the lengthy periods between mass extinctions, the nature of species interactions may change. For example, the diversity and intensity of insect herbivory increased during a warming trend from the Late Palaeocene to the Eocene[12]. Moreover, many late Pleistocene plant and animal assemblages that contain some extant species have no modern analogues[13,14]. Such results hint that general patterns of species associations observed in contemporary assemblages could have been quite different in the past.

Here we ask whether non-random species associations of plant and mammal assemblages over the past 300 million years (Myr) are dominated by segregated or aggregated species pairs. This novel analysis is designed to compare statistical patterns of taxon associations for fossil and modern data using a consistent set of methodologies. We analysed 80 well-sampled fossil and recent assemblages: 38 for mammals and 42 for plants (see Supplementary Information, Extended Data Fig. 1 and Extended Data Table 1). Each data set contained information on taxon presence and absence across multiple localities in a particular time period (Extended Data Fig. 1 and Extended Data Table 1). Ages of plant data sets range from 307 million years ago (Ma) to the present and are from North America and Africa. Mammal data sets range in age from 21.4 Ma to the present and are from North America, Eurasia, and Africa. We compared each data set to a 'null' assemblage generated by randomization, scored each taxon pair as random, aggregated, or segregated, and used an empirical Bayes approach to control for the rate of false positive discoveries[15]; see Methods). Finally we

[1]Department of Paleobiology, National Museum of Natural History, Smithsonian Institution, Washington DC 20013, USA. [2]Department of Environmental Science and Biology, The College at Brockport – SUNY, Brockport, New York 14420, USA. [3]School of Natural Sciences, University of California, Merced, 5200 North Lake Road, Merced, California 95343, USA. [4]Department of Geology and Geophysics, Yale University, New Haven, Connecticut 06520, USA. [5]Hominid Paleobiology Doctoral Program, Center for the Advanced Study of Hominid Paleobiology, Department of Anthropology, George Washington University, Washington DC 20052, USA. [6]Department of Geosciences and Geography, University of Helsinki, PO Box 64, 00014 University of Helsinki, Finland. [7]School of Social Science, The University of Queensland, Brisbane, Queensland 4072, Australia. [8]Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington DC 20013, USA. [9]Center for Macroecology, Evolution and Climate, University of Copenhagen, Copenhagen 2100, Denmark. [10]Biological Sciences Graduate Program, University of Maryland, College Park, Maryland 20742, USA. [11]Florida Museum of Natural History, University of Florida, Gainsville, Florida 32611, USA. [12]Department of Entomology, University of Maryland College Park, College Park, Maryland 20742, USA. [13]Key Lab of Insect Evolution and Environmental Changes, Capital Normal University, Beijing 100048, China. [14]Department of Integrative Biology and Museum of Paleontology, University of California Berkeley, Berkeley, California 94720, USA. [15]School Biology and Ecology & Sustainability Solutions Initiative, University of Maine, Orono, Maine 04469, USA. [16]Department of Geology, University of Cincinnati, Cincinnati, Ohio 45221, USA. [17]Department of Biological Sciences, Macquarie University, Sydney, New South Wales 2109, Australia. [18]Department of Anthropology, Human Origins Program, National Museum of Natural History, Smithsonian Institution, Washington DC 20013, USA. [19]School of Life Sciences, University of Nevada-Las Vegas, Las Vegas, Nevada 89154, USA. [20]Department of Integrative Biology, Oregon State University, Corvallis, Oregon 97331, USA. [21]Chair of Ecology and Biogeography, Nicolaus Copernicus University, Lwowska 1, 87-100 Torun, Poland. [22]Evolutionary Studies Institute, University of the Witwatersrand, Jorissen Street, Braamfontein, Johannesburg 2001, South Africa. [23]Department of Botany, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA. [24]Department of Biology, University of Vermont, Burlington, Vermont 05405, USA.

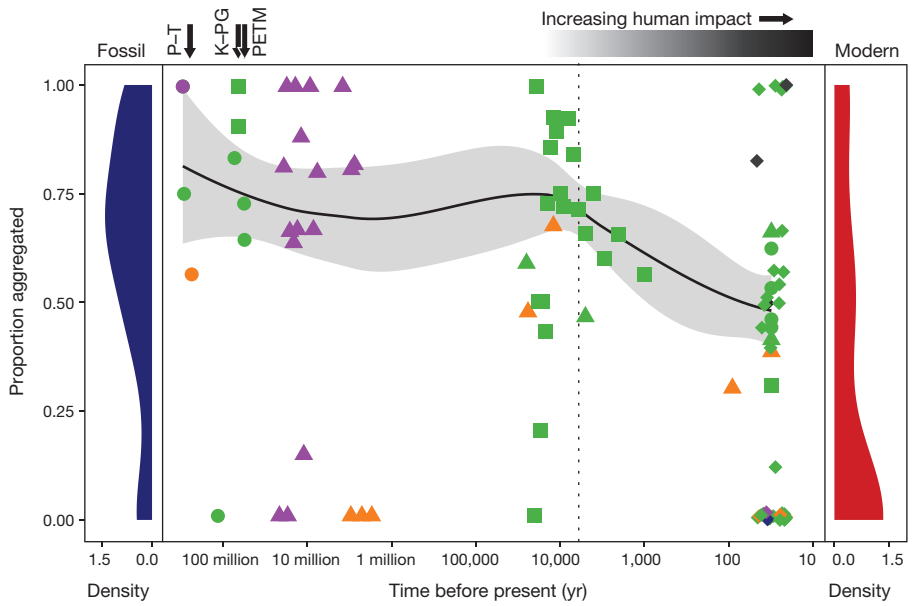


Figure 1 | Proportion of aggregated pairs over the past 300 Myr. Weighted Loess curve with shaded 95% confidence intervals illustrates reduction in the proportion of aggregated species pairs in the Holocene (log scale). Dotted vertical line at 5,998 years delineates the linear model breakpoint in the trend (Methods and Extended Data Fig. 2). Non-random species pairs of 'Fossil' data (blue density profile) are predominantly aggregated, whereas 'Modern' data (red density profile) are predominantly segregated. Colours indicate continent: North America (green), Eurasia (purple), Australia (dark grey), South America (dark blue), Africa (orange). Point shapes indicate type of data: pollen (square), mammals (triangle), macroplants (circle). Data on terrestrial communities from ref. 2 are diamonds. All fossil and modern data are from mainland sites; no island sites were included. Time values of modern data points were assigned a single age (see Supplementary Information data sets), but are jittered for visual representation. P–T, Permo–Triassic transition; K–Pg, Cretaceous–Palaeogene transition; PETM, Palaeocene–Eocene thermal maximum.

synthesized our results with those from a meta-analysis of 39 modern communities that used the same methodology[2,10].

For all fossil data sets, most taxon pairs were random (87–100% of possible pairs; Extended Data Table 1), which is also typical for modern assemblages[2]. This result reflects the statistically conservative nature of the tests used to identify significantly associated pairs, and the fact that most taxon pairs in a diverse, well-sampled assemblage interact weakly, or not at all. In 62 of 80 assemblages analysed here, a subset of taxon pairs showed significant associations that are stronger than can be explained by the null model, even after controlling for the false discovery rate (Fig. 1). Unlike modern mainland assemblages, most significant associations in the fossil record are aggregated, positive associations (Fig. 1). This pattern is consistent across the past 300 Myr for the diverse fossil assemblages in this study, which encompass mammals, plant macrofossils, and pollen from multiple continents and time slices.

However, beginning in the Holocene, there was a significant temporal trend towards a greater proportion of segregated species pairs, which is consistent with the results for modern assemblages. A breakpoint analysis indicates that the shift began approximately 6,000 years ago (Extended Data Fig. 2). Confidence intervals of the breakpoint are large owing to a lack of appropriate data sets between 20,000 and 1 million years ago. Therefore, it is difficult to pinpoint the exact time of the shift, but a closer examination of the data suggests that placing it within the Holocene is reasonable. Before the breakpoint, on average 64% of significant pairs were aggregated (median = 73%). After the breakpoint, the average dropped to 37% (median = 42%). This trend is not driven by the modern data and persists when only fossil data are analysed (Extended Data Fig. 3).

Why are species associations so different in fossil versus modern assemblages? We first tested and eliminated five potential 'artefact' hypotheses that are related to sampling issues (see Methods for details). (1) Collection modes were discounted because they were heterogeneous both for the modern and for fossil assemblages, and because the decrease in aggregated pairs was strong in fossil pollen and mammal assemblages that spanned the shift. Moreover, sampling methodology was consistent within an assemblage type across periods that encompass the change (Extended Data Fig. 4). (2) Scale was discounted because there was no relationship between the spatial or temporal extent and grain of each data set and the percentage of aggregated pairs (Fig. 2 and Extended Data Fig. 5). (3) Taphonomic bias was discounted because the null model algorithm preserved the marginal totals of the data matrix in each randomized assemblage, controlling for simple taphonomic biases that could generate heterogeneity in the number of species per site or the number of occupied sites per species. (4) Taxonomic resolution was discounted because parallel analyses at the genus and species levels did not produce systematic changes in the proportions of aggregated pairs (Extended Data Table 2). (5) Increased sampling of rare species in modern data sets was discounted because segregated pairs tend to form in species with intermediate occupancy, whereas aggregated pairs form both in common and in rare species in modern and fossil data sets. All of these mechanisms can potentially affect assemblage structure in fossil and modern data sets. However, our analyses suggest that these mechanisms cannot account for the prominent decrease in aggregated species pairs that began during the Holocene (Fig. 1).

The failure of sampling issues to account for the temporal change in the percentage of non-randomly associated taxon pairs suggests that a mechanistic explanation is required. We consider two hypotheses that invoke a systematic change in either abiotic or biotic factors as drivers of co-occurrence patterns.

One of the most obvious differences between the present interval and the past 300 Myr of geological history represented by these fossil assemblages is the increasing variability of climate towards the present, associated with the glacial–interglacial cycles of the Quaternary period[16]. This is not to say that there were no periods of high climate variability before the ice ages, but that our data do not regularly sample times of high climate variability in deep time. If climate variability is responsible for the shift in the frequency of aggregated species pairs, there should be a negative relationship between climate variability

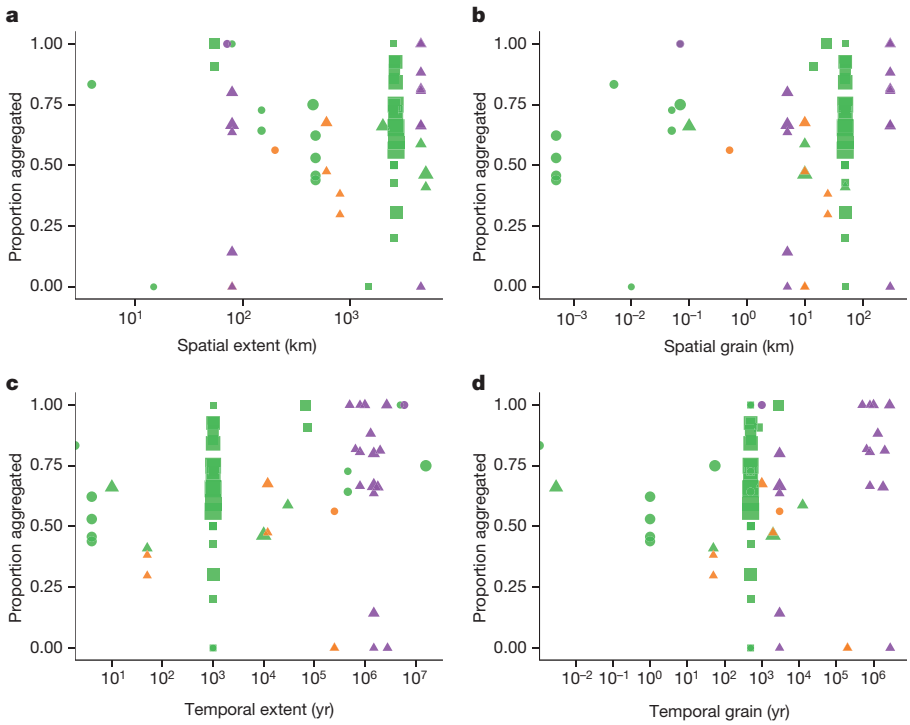


Figure 2 | Relationship between scale and proportion of aggregated pairs. The proportion of significant pairs that are aggregated does not depend on the temporal or spatial scale of data. Each point represents a single data set. **a**, **b**, Aggregated pairs versus spatial extent (longest linear distance between any two sites in a data set; **a**) or spatial grain (estimated radius of collection area that fossil specimens would have been transported to the depositional environment in a typical locality; **b**). **c**, **d**, Proportion of aggregated pairs versus temporal extent (duration from the oldest to youngest locality in a data set; **c**) or temporal grain (typical amount of time-averaging of localities in a data set; **d**). Colours and shapes as in Fig. 1. Note the logarithmic scale of the *x* axes. Modern data from ref. 2 are excluded from this analysis.

and the percentage of aggregations. We quantified climate variability within the temporal extent of each data set for the past 65 Myr, using climate data from ice[17] and deep sea[16] cores that were standardized to a common scale (Methods). We found no relationship between the proportion of aggregated pairs and the standard deviation of climate within the sampled time slice (Extended Data Figs 6 and 7), or the standard deviation of the first differences of climate within the sampled time slice (Fig. 3a). Collectively, these results suggest that the increasing variability in climate in the Quaternary cannot explain the decreased frequency of aggregation.

An alternative explanation is that the mid- to late Holocene is unusual in the evolutionary history of terrestrial ecosystems, and that biotic drivers (as opposed to climate) now are different from what they have typically been over the past 300 Myr. First, we asked whether there was a significant shift in the proportions of aggregated versus segregated pairs across critical geological intervals that spanned periods of mass extinctions or major climatic change during the past 300 Myr (Extended Data Table 3 and Extended Data Fig. 4). We found a significant decrease in the percentage of positive associations only in data sets that spanned the Pleistocene–Holocene transition (11,700 years ago). With the exception of large-bodied mammals in North America and Africa (Extended Data Fig. 4b), aggregated species pairs decreased in all data sets through the Pleistocene–Holocene transition. In contrast, there was no significant change in the percentage of aggregations across the three other critical intervals that were encompassed by these data: the Permo–Triassic transition (252 Ma), the Cretaceous–Palaeogene transition (66 Ma), and the Palaeocene–Eocene thermal maximum (56 Ma). These intervals include the Cretaceous–Palaeogene mass extinction, responsible for the loss of the non-avian dinosaurs[18], and the Permo–Triassic extinction, the largest mass extinction ever recorded[18]. Even the Palaeocene–Eocene thermal maximum, a period of major climatic change in which global temperatures increased ~5–8 °C in a few millennia[19], did not coincide with a change in the relative proportions of aggregated versus segregated pairs.

It is difficult to pinpoint the exact mechanism responsible for the uniqueness of the present time interval. However, our analyses provide some clues about possible cause. Data that encompass the shift towards the modern pattern are almost exclusively North American (Fig. 1 and Extended Data Fig. 4). The statistical confidence interval bracketing the breakpoint at 6,000 years ago encompasses the beginning of agriculture in North America around 8,000 years ago[6] and the increase in human populations during the Holocene[4,5]. The trend towards greater segregations in North American pollen (Fig. 1 and Extended Data Fig. 4), with particularly strong shifts occurring in the past 2,000 years[20], is also consistent with the history of agriculture in North America. Cultivation of multiple species of domesticated plants began approximately 3,800 years ago[6,7], with evidence for more general dependency on agriculture in North America beginning 1,300 years ago[6,7]. Estimates of global land area under cultivation increase rapidly starting 6,000 years ago and are as high as 4×10^8 hectares ($1\ \text{hectare} = 10^4\ \text{m}^2$) by 2,000 years ago[4,5].

Possible drivers by which increasing human impacts led to an increase in segregated pairs include (1) increases in hunting and domestication of particular species[21,22], (2) changes in land use[4,5], (3) increases in the frequency of fire[4], (4) increases in habitat fragmentation and dispersal barriers[23,24], and (5) deliberate and accidental spread of species beyond their native geographical ranges[25–27]. We note that modern island assemblages (which we excluded from our comparisons with fossil assemblages) are more segregated than modern mainland assemblages (Fig. 3b), which is consistent with the hypothesis that habitat fragmentation and dispersal limitation have increased segregated pairs. Possibly all of the processes listed play a role. Although their combined effects on taxon pairs are difficult to predict, the relative importance of factors structuring species co-occurrence appears to have changed through the Holocene. Future work comparing the co-occurrence structure of fossil and modern communities should allow us to better understand how this alteration will play out in the future. Regardless of the precise mechanisms,

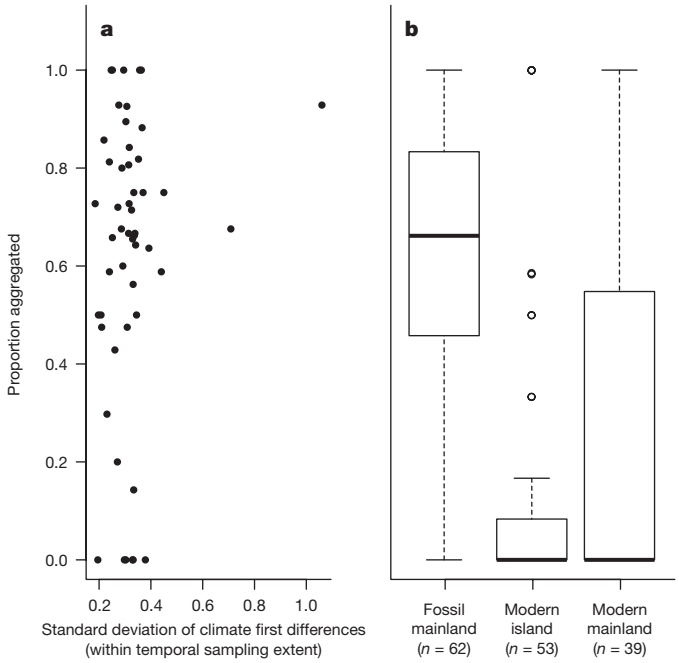


Figure 3 | Tests of possible mechanisms for decreasing spatial aggregation through time. a, Climate variability within the temporal extents of the fossil data sets is uncorrelated with the proportion of aggregated species pairs. Climate variability is measured as the standard deviation of the first differences in climate across the interval (see also Extended Data Fig. 6). **b**, Box plots show the proportion of aggregated species pairs for fossil data, modern islands and modern mainland assemblages. Dashed lines indicate maximum and minimum values, circles are outliers. Island assemblages, with more limited capacity for dispersal, have the smallest and least variable fraction of aggregated pairs. Mainland fossil assemblages are significantly more aggregated than mainland modern assemblages. Note, these island data were excluded from other analyses.

humans appear to be agents of disturbance on a large scale and have been so for longer than is often recognized.

Our results suggest that assemblage co-occurrence patterns remained relatively consistent for 300 Myr but have changed over the Holocene as the impact of humans has dramatically increased. Across shallower time intervals, other studies have documented that local and regional species composition has changed substantially over recent decades[28,29] and millennia[30]. The rules governing community assembly, at least as implicated by co-occurrence patterns, seem to have changed during the Holocene and continue to change with the increasing influence of human activity. The co-occurrence structure of modern and recent plant and animal assemblages thus appears to be unique in the evolutionary history of terrestrial ecosystems, an important perspective for assessing challenges to these ecosystems in the face of present and future human impacts.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 August; accepted 13 November 2015.
Published online 16 December 2015; corrected online 6 January 2016 (see full-text HTML version for details).

1. Botkin, D. B. *et al.* Forecasting the effects of global warming on biodiversity. *Bioscience* **57,** 227–236 (2007).
2. Gotelli, N. J. & Ulrich, W. The empirical Bayes approach as a tool to identify non-random species associations. *Oecologia* **162,** 463–477 (2010).
3. Gotelli, N. J. & McCabe, D. J. Species co-occurrence: a meta-analysis of J.M. Diamond's assembly rules model. *Ecology* **83,** 2091–2096 (2002).
4. Marlon, J. R. *et al.* Global biomass burning: a synthesis and review of Holocene paleofire records and their controls. *Quat. Sci. Rev.* **65,** 5–25 (2013).
5. Ruddiman, W. F. & Ellis, E. C. Effect of per-capita land use changes on Holocene forest clearance and CO_2 emissions. *Quat. Sci. Rev.* **28,** 3011–3015 (2009).
6. Zeder, M. A. & Smith, B. D. A conversation on agricultural origins: talking past each other in a crowded room. *Curr. Anthropol.* **50,** 681–691 (2009).
7. Smith, B. D. & Yarnell, R. A. Initial formation of an indigenous crop complex in eastern North America at 3800 B.P. *Proc. Natl Acad. Sci. USA* **106,** 6561–6566 (2009).
8. Connor, E. F. & McCoy, E. D. The statistics and biology of the species-area relationship. *Am. Nat.* **113,** 791–833 (1979).
9. Connor, E. F., Collins, M. D. & Simberloff, D. The checkered history of checkerboard distributions. *Ecology* **94,** 2403–2414 (2013).
10. Ulrich, W. & Gotelli, N. J. Null model analysis of species associations using abundance data. *Ecology* **91,** 3384–3397 (2010).
11. Wagner, P. J., Kosnik, M. A. & Lidgard, S. Abundance distributions imply elevated complexity of post-Paleozoic marine ecosystems. *Science* **314,** 1289–1292 (2006).
12. Wilf, P. & Labandeira, C. C. Response of plant-insect associations to Paleocene-Eocene warming. *Science* **284,** 2153–2156 (1999).
13. Lyons, S. K. A quantitative model for assessing community dynamics of Pleistocene mammals. *Am. Nat.* **165,** E168–E185 (2005).
14. Jackson, S. T. & Williams, J. W. Modern analogs in Quaternary paleoecology: here today, gone yesterday, gone tomorrow? *Annu. Rev. Earth Planet. Sci.* **32,** 495–537 (2004).
15. Ulrich, W. *Pairs* – a FORTRAN program for studying pair wise species associations in ecological matrices. http://www.keib.umk.pl/wp-content/uploads/2013/04/PairsManual.pdf (2008).
16. Zachos, J. C., Dickens, G. R. & Zeebe, R. E. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* **451,** 279–283 (2008).
17. Augustin, L. *et al.*; EPICA community members. Eight glacial cycles from an Antarctic ice core. *Nature* **429,** 623–628 (2004).
18. Bambach, R. K. Phanerozoic biodiversity mass extinctions. *Annu. Rev. Earth Planet. Sci.* **34,** 127–155 (2006).
19. Bowen, G. J. *et al.* Two massive, rapid releases of carbon during the onset of the Palaeocene–Eocene thermal maximum. *Nature Geosci.* **8,** 44–47 (2015).
20. Blois, J. L. *et al.* A framework for evaluating the influence of climate, dispersal limitation, and biotic interactions using fossil pollen associations across the late Quaternary. *Ecography* **37,** 1095–1108 (2014).
21. Delcourt, P. A. & Delcourt, H. R. *Prehistoric Native Americans and Ecological Change: Human Ecosystems in Eastern North American since the Pleistocene.* (Cambridge Univ. Press, 2004).
22. Barnosky, A. D. Colloquium paper: megafauna biomass tradeoff as a driver of Quaternary and future extinctions. *Proc. Natl Acad. Sci. USA* **105** (suppl. 1), 11543–11548 (2008).
23. Haddad, N. M. *et al.* Habitat fragmentation and its lasting impact on Earth's ecosystems. *Sci. Adv.* **1,** e1500052 (2015).
24. Williams, M. *Deforesting the Earth: From Prehistory To Global Crisis* (Univ. Chicago Press, 2003).
25. Grayson, D. K. The archaeological record of human impacts on animal populations. *J. World Prehist.* **15,** 1–68 (2001).
26. van Kleunen, M. *et al.* Global exchange and accumulation of non-native plants. *Nature* **525,** 100–103 (2015).
27. Capinha, C., Essl, F., Seebens, H., Moser, D. & Pereira, H. M. The dispersal of alien species redefines biogeography in the Anthropocene. *Science* **348,** 1248–1251 (2015).
28. Dornelas, M. *et al.* Assemblage time series reveal biodiversity change but not systematic loss. *Science* **344,** 296–299 (2014).
29. Tóth, A. B., Lyons, S. K. & Behrensmeyer, A. K. A century of change in Kenya's mammal communities: increased richness and decreased uniqueness in six protected areas. *PLoS ONE* **9,** e93092 (2014).
30. Thompson, J. R., Carpenter, D. N., Cogbill, C. V. & Foster, D. R. Four centuries of change in northeastern United States forests. *PLoS ONE* **8,** e72540 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank G. Dietl for comments that improved the manuscript. Support for this research was provided by a National Museum of Natural History Program grant to the Evolution of Terrestrial Ecosystems Program (ETE) and NSF-DEB 1257625. This is ETE publication 338.

Author Contributions All authors provided input into the final manuscript. S.K.L. helped design the study, contributed and analysed data, and wrote the paper. K.L.A., J.L.B., J.T.F., and W.U. were involved in the study design, and contributed and analysed data. A.T. was involved in the study design, contributed and analysed data, and created Supplementary Fig. 1. A.K.B., A.B., W.A.D., J.T.E., N.J., C.L., C.V.L., S.W., and B.R. were involved in the study design and contributed data. H.A., J.A., and D.W. contributed data. M.D. and J.H.M. were involved in the study design, analysed data, created figures and wrote text for the supplementary information. A.D. was involved in the study design and analysed data. G.R.G., B.M., D.P., S.P.-M., R.T., R.P., and A.V. were involved in the study design. N.J.G. designed the study and helped write the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.K.L. (lyonss2@si.edu).

METHODS

No statistical methods were used to predetermine sample size.

Detection of non-random species pairs. The data for each analysis consist of a binary presence–absence matrix in which each row is a taxon and each column is a sample. The entries represent the presence (1) or absence (0) of a particular taxon in a particular sample. Within this matrix, each of the $S(S-1)/2$ unique species pairs is tested and classified as random, aggregated, or segregated. The tests were performed with the PAIRS version 1.0 software application[15,31]. The methodology is described fully in ref. 2 and is briefly described here.

The analysis begins by calculating a scaled C score[32]: $C_{ij} = (R_i - D)(R_j - D)/R_iR_j$, where C_{ij} is the C score for species pair i and j, R_i is the row total (the number of species occurrences) for species i, R_j is the row total for species j, and D is the number of shared sites in which both species are present. For each species pair, this index ranges from 0.0 (aggregation: maximal co-occurrence of both species) to 1.0 (segregation: minimal co-occurrence of both species). PAIRS calculates the C score for each pair of species and assigns it to a histogram bin. There are 20 bins that range from 0 to 1 in 0.05 intervals, plus a bin for 0.0 (perfectly aggregated pairs) and a bin for 1.0 (perfectly segregated pairs).

We next estimate the P value associated with each species pair by a randomization test. The data matrix is first randomized by reshuffling all matrix elements, with the restriction that the row and column sums of the original matrix are preserved. This 'fixed-fixed' algorithm has been subject to extensive benchmark testing with artificial random and structured matrices[2,33,34]. Compared with a variety of other null model algorithms, the fixed-fixed algorithm most effectively screens against type I errors (incorrectly rejecting the null hypothesis for a random matrix), but is somewhat conservative[33].

An alternative algorithm 'fixed-equiprobable' retains row sums (species occurrence frequencies), but allows column totals (species richness per site) to vary freely. The fixed-equiprobable algorithm also has good statistical properties, and is appropriate for modern data sets in which sampling effort has been standardized, such as quadrat samples of fixed area. However, this algorithm is not appropriate for fossil data because the number of species detected per site in fossil assemblages is determined primarily by sampling effort of the collector and by site-specific taphonomic biases in preservation.

For these reasons, we have used only the fixed-fixed model, both for the analysis of fossil assemblages and for comparison with modern assemblages. Details of the randomization are discussed further in refs 2, 35. Using 1,000 randomizations, we create a simple P value (two-tailed test) for each species pair, which leads to a classification of each species pair as aggregated, random, or segregated.

However, with a total of $S(S-1)/2$ unique pairs in a matrix of S species, retaining all of the significant pairs ($P < 0.05$) would generate a potentially large number of false positive results. This problem has frequently arisen in the analysis of micro-arrays, genomic surveys, and other examples of 'big data'[36]. The PAIRS analysis relies on an empirical Bayes approach by creating a histogram of C score values based on the pairs generated in each null assemblage. To screen out false positives, we calculated the average number of species pairs in each bin of the histogram. Next, we determined the observed number of pairs from the empirical assemblage in each bin, ordered by their P values from the simulation. We retained only those pairs that were above the mean number for each bin and were statistically significant on the basis of the simple P value criterion. This double screen effectively eliminates many of the false-positives that can arise in random data sets[2].

Weighted Loess regression. A Loess smoothing line was created with the stat_smooth function in the R package ggplot2 version 1.0.0 (ref. 37) using default parameters. For Loess fitting, the fit at point x is made using points in the neighbourhood of x (closest 75% of total points), with tricubic weighting (proportional to $(1 - (\text{distance/maximum distance})^3)^3$). Points were additionally weighted by the number of sites in each matrix and 95% confidence intervals were generated using a t-based approximation.

Analysis of climate variability. To examine the how climate variability impacts the percentage of aggregated species pairs, we used climate proxy data from ice[17] and deep sea cores[16], which collectively encompass the past 65 Myr of the assembled data sets. The European Project for Ice Coring in Antarctica (EPICA) data were used preferentially when there was temporal overlap between proxy data sets. Climate data were mean centred and standardized before pooling into a single data time series. We then sampled the climate data across the 'temporal extents' (Extended Data Table 1) of the individual Evolution of Terrestrial Ecosystems Program (ETE) data sets to test if there were relationships between the percentage of aggregated species pairs and climate variability. Climate variability was calculated in two ways: (1) as the standard deviation of climate within the temporal extent of each data set and (2) as the standard deviation of the first differences (changes in climate from available time-step to time-step within the temporal extent of a data set) of climate. We used standard deviation because it helps address issues with changes in data density over time. Estimated rates of change are sensitive to the time span over which they are measured and more closely spaced data would shift apparent rates of change. Approaches using standard deviation are less sensitive to this issue. We also compared climate variability with age (years before present) of ETE data sets to test for Phanerozoic-scale trends in climate variability sampled by ETE data sets.

Breakpoint analysis. We used a maximum likelihood approach, available in the R package 'segmented' version 1.1, to estimate the breakpoint time at which the sharper decline in aggregated species pairs began. This analysis used an initial linear model of the proportion of aggregated pairs as a function of community age ($\log_{10}$ of years before present) to generate a best-fitting number of breakpoints, with separate regression lines fit to each segment. A bootstrap of 1,000 replicates was used to estimate uncertainty in the model parameters (including uncertainty in the time of the breakpoint).

Tests of artefacts. *Collection modes*. We thought that differences in the way fossil and modern data are collected might be responsible for the observed difference in the relative proportions of aggregated versus segregated species pairs in modern communities[2,10] and fossil communities. There are two reasons why collection modes are not likely to be responsible for this difference. First, fossil collections are heterogenous by nature. Different collecting methods are used for different taxonomic groups (for example, bulk sampling, surface sampling, cores). Moreover, even within a taxonomic group, the type of depositional environment imposes different kinds of bias (for example, cave sites versus open pits for Pleistocene mammals). Second, we see a switch from species pairs that are dominated by aggregations to those dominated by segregations in our data sets that span the Pleistocene–Holocene transition (Extended Data Fig. 4 and Extended Data Table 1). In particular, mammal assemblages show a switch from >50% aggregations in the Pleistocene to <50% aggregations in the Holocene. The data encompassing this switch are all fossil localities and there are similar biases in both time slices. Although there is variation in the pollen assemblages, a weighted regression that takes into account the sampling in each time slice shows a significant decrease through time ($P = 0.04$, $R^2 = 0.15$). This trend of increasing percentage of segregated pairs begins approximately 14,000 years ago and continues across the Holocene with the switch occurring in the final 1,000 year time slice[20]. The fact that these data were all collected using the same sampling techniques suggests that sampling cannot account for this pattern.

Issues of scale. It is generally assumed that fossil data are biased. Although the type of bias is not the same for all taxonomic groups, most fossil assemblages contain some degree of temporal or spatial averaging[38]. That is, they represent accumulations of species that can occur over hundreds or thousands of years and may mix species that did not exist at the locality at the same time[39]. The fossil data sets in this analysis include assemblages that range from no time-averaging (for example, fossil leaves preserved in volcanic event beds) to those that are time-averaged over thousands or hundreds of thousands of years (for example, some mammal assemblages). In addition, some data sets could not be resolved to time bins of less than a million years. Spatial averaging is less of an issue in these data sets, but individual samples are drawn from areas with diameters ranging from a few metres to more than 300 km (Supplementary Table 1).

If issues of scale are contributing to the pattern found here, there should be a relationship between the proportion of significant pairs that are aggregated and the spatial or temporal scale of the data. We evaluated this by estimating the spatial or temporal grain and extent of each data set included in the analyses (Extended Data Table 1) and determining if there was a significant relationship with the percentage of aggregations. The spatial grain is the estimated radius of collection area over which fossil specimens would have been transported to the depositional environment in a typical locality. The temporal grain is the typical amount of time-averaging of localities in a data set. Spatial extent is the longest linear distance between any two sites in a data set and temporal extent is the duration from the oldest to youngest locality in a data set.

We found no relationship between the scale of the data sets and the proportion of significant pairs that were aggregated versus segregated (Fig. 2 and Extended Data Fig. 5). Regression analyses were not significant and explained very little of the variation in the data (Extended Data Fig. 6). The pattern of segregated versus aggregated pairs was not different in fossil versus modern assemblages because of biases related to the scale of fossil data.

Taphonomic bias. How can taphonomy and palaeoenvironment affect species frequencies (richness) and spatial representation? The fossil record contains buried assemblages of species that were derived from living communities at different times in the past. Species representation (presence or absence) in individual fossil assemblages is a critical attribute of our data sets, therefore we need to consider how this variable might be biased relative to original associations of species in communities. Taphonomic processes operate during the transition of dead remains into preserved samples and thus control the biological information that passes

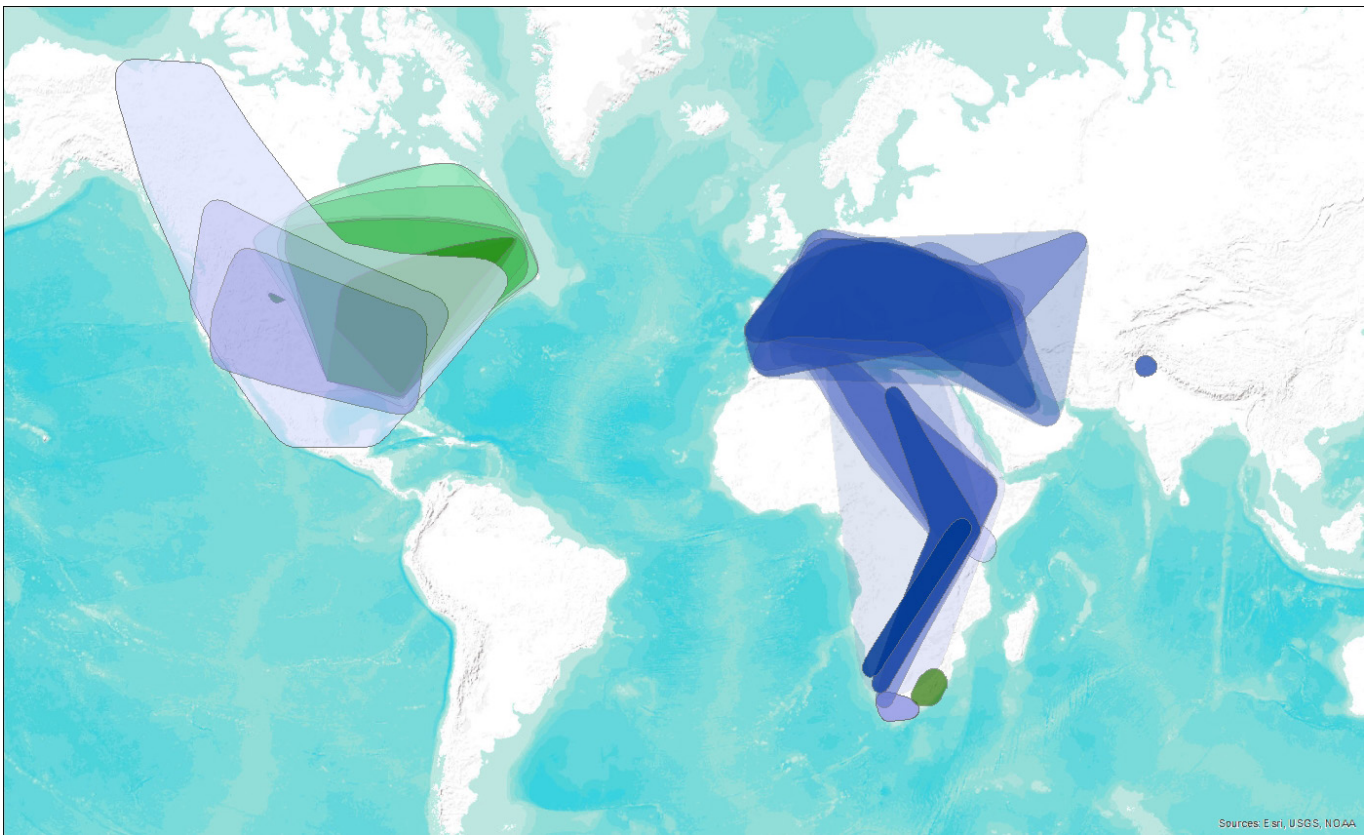
from the living community into the fossil record[39]. These processes act as filters that can alter species representation in fossil samples in a variety of ways: (1) selective preservation of organisms with particular body compositions and sizes, for example organisms with and without mineralized skeletons, larger versus smaller individuals; (2) variable preservation of organisms depending on their population abundance, spatial distribution and life habits, for example aquatic versus terrestrial; (3) post-mortem or depositional mixing of species that did not live together (time-averaging), or separation of species that did (selective transport or destruction). Additionally, some types of environment are better represented in the depositional record than others, such as wetlands versus dry land surfaces. All of these add up to potential biases that could affect biological signals and the proportions of random versus significant species pairs, or the proportions of segregated versus aggregated pairs, in our analyses.

However, the particular null model algorithm used effectively controls for major sources of taphonomic bias in the data set. This 'fixed-fixed' algorithm[33] creates null assemblages that have the same species richness per sample, and the same number of occurrences per species, as the original data. Thus, if there are preservation biases that generate heterogeneity in the total number of fossil species per sample, or biases in the number of specimens per species, these are effectively controlled for in the analysis. Significant patterns of species aggregation are those measured beyond the effects of sampling heterogeneity in the occurrences of species or the number of species per sample. Similar sampling effects are controlled for in the modern data, which can also exhibit variation in the commonness or rarity of species and in the number of species per sample.

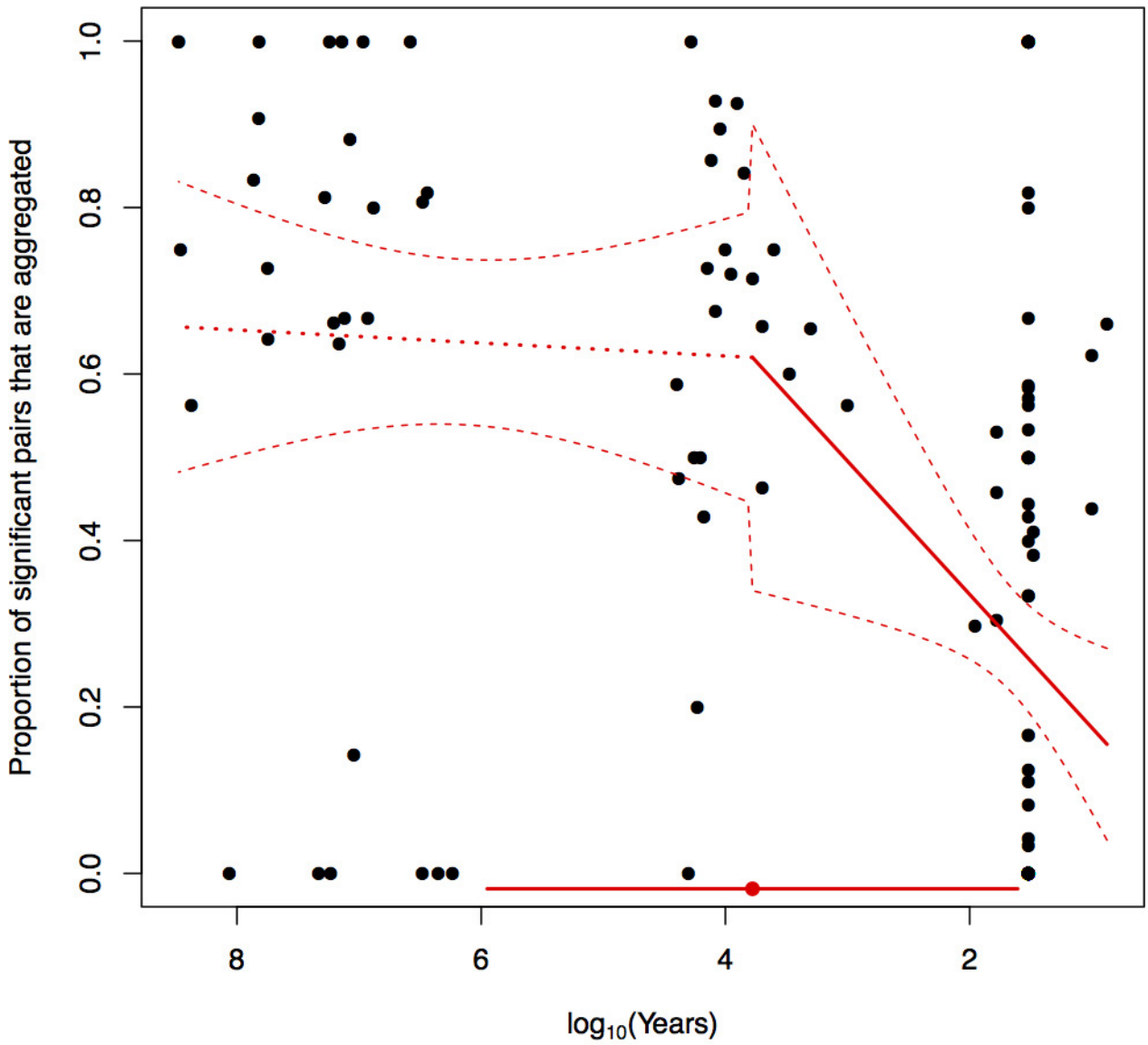
Taxonomic resolution of the data. Fossils are not always resolvable to the species level and are frequently analysed at the genus level. This may have the effect of increasing geographical ranges and overlap between taxa, and may contribute to the dominance of aggregated pairs found in this study. To test whether this was the case, we analysed 18 of the data sets at the species and genus level (16 mammal and 2 plant data sets). If taxonomic resolution is driving the pattern, we expect to see an increase in the proportion of aggregated pairs when species are lumped into genera. We found that six of the data sets showed the expected increase. However, nine showed a decrease and three showed no change (Extended Data Table 2). Interestingly, one of the modern data sets on small mammals from the Great Basin had genetic information that indicated that some were cryptic species. When the analysis was re-run with the cryptic species identified, there was an increase in the proportion of significantly aggregated pairs (from 50% to 61%). This is in the opposite direction that we would expect if lumping species into genera artificially increased aggregated pairs. Taken together, these results suggest that the differences between species associations over the past 300 Myr and the present are not driven by the taxonomic resolution of fossil data.

Sampling of abundant and rare species in fossil and modern data. The results of null model analyses of abundance versus presence–absence data are compared in ref. 10. The two kinds of analysis give qualitatively comparable results, although the abundance analyses are somewhat more powerful in detecting non-randomness. It is generally assumed that fossil deposits miss the rarest species in a community because preservation potential increases with abundance; more individuals means more opportunities for fossilization events. If rare species are more likely to form segregated pairs, we would expect to see more segregations in modern data sets because they should sample more of the rare species than comparable fossil data sets. Within fossil data sets, we would expect to see more segregated pairs in data sets with better sampling and more rare species. We evaluated this using a data visualization technique. We present the results of our analyses as a series of pairwise species by species matrices and order species by occupancy (see Supplementary Information: data sets). Occupancy decreases as one moves to the right on the x axis and up on the y axis. Species with the highest occupancy are close to the origin. The pairwise associations are colour-coded: grey for random pairs, blue for aggregated pairs, and red for segregated pairs. If increasing sampling of rare species is responsible for the pattern we document, then we would expect to see a preponderance of red, segregated pairs in the upper, right-hand corner of the species by species matrices. In particular, this should show up in data sets with better sampling and those that encompass the shift from more aggregated to more segregated pairs (for example, Pleistocene–Holocene mammals and pollen, modern mammals in Kenya, and modern plants in Wisconsin). This is not the pattern that we see. In fact, we find that segregated pairs tend to form with species of intermediate occupancy and that aggregated pairs form both with common species and with rare species. Differences in the sampling of rare species between fossil and modern data sets cannot account for the shift in species associations over time.

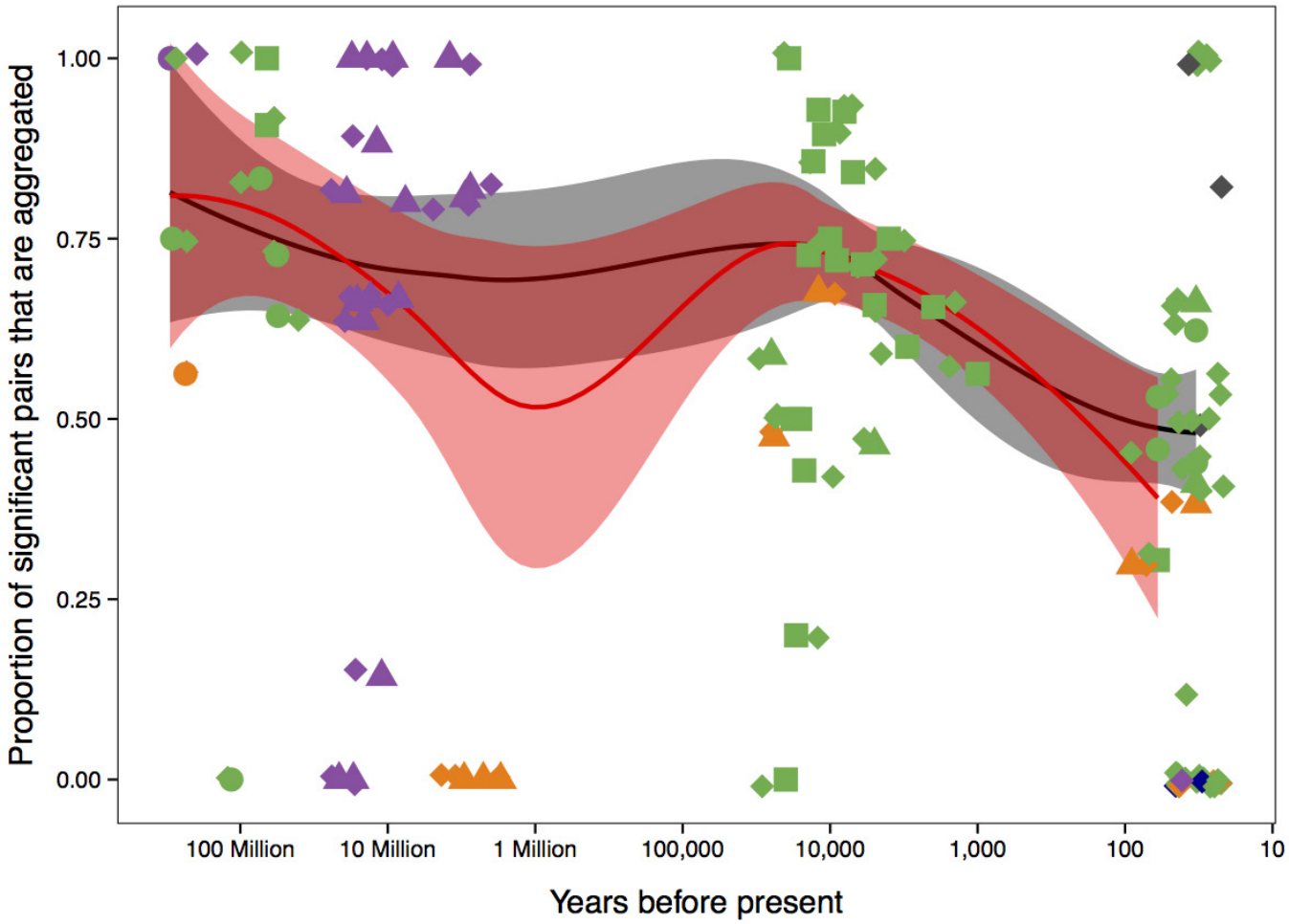
31. Ulrich, W. Pairs – a FORTRAN program for studying pair-wise species associations in ecological matrices. http://www.keib.umk.pl/pairs/?lang=en, (2010).
32. Stone, L. & Roberts, A. The checkerboard score and species distributions. *Oecologia* **85,** 74–79 (1990).
33. Gotelli, N. J. Null model analysis of species co-occurrence patterns. *Ecology* **81,** 2606–2621 (2000).
34. Ulrich, W. & Gotelli, N. J. Null model analysis of species nestedness patterns. *Ecology* **88,** 1824–1831 (2007).
35. Gotelli, N. J. & Ulrich, W. Statistical challenges in null model analysis. *Oikos* **121,** 171–180 (2012).
36. Efron, B. Bayesians, frequentists, and scientists. *J. Am. Stat. Assoc.* **100,** 1–5 (2005).
37. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2009).
38. Kidwell, S. M. & Holland, S. M. The quality of the fossil record: implications for evolutionary analyses. *Annu. Rev. Ecol. Syst.* **33,** 561–588 (2002).
39. Behrensmeyer, A. K., Kidwell, S. M. & Gastaldo, R. A. Taphonomy and paleobiology. *Paleobiology* **26,** 103–147 (2000).



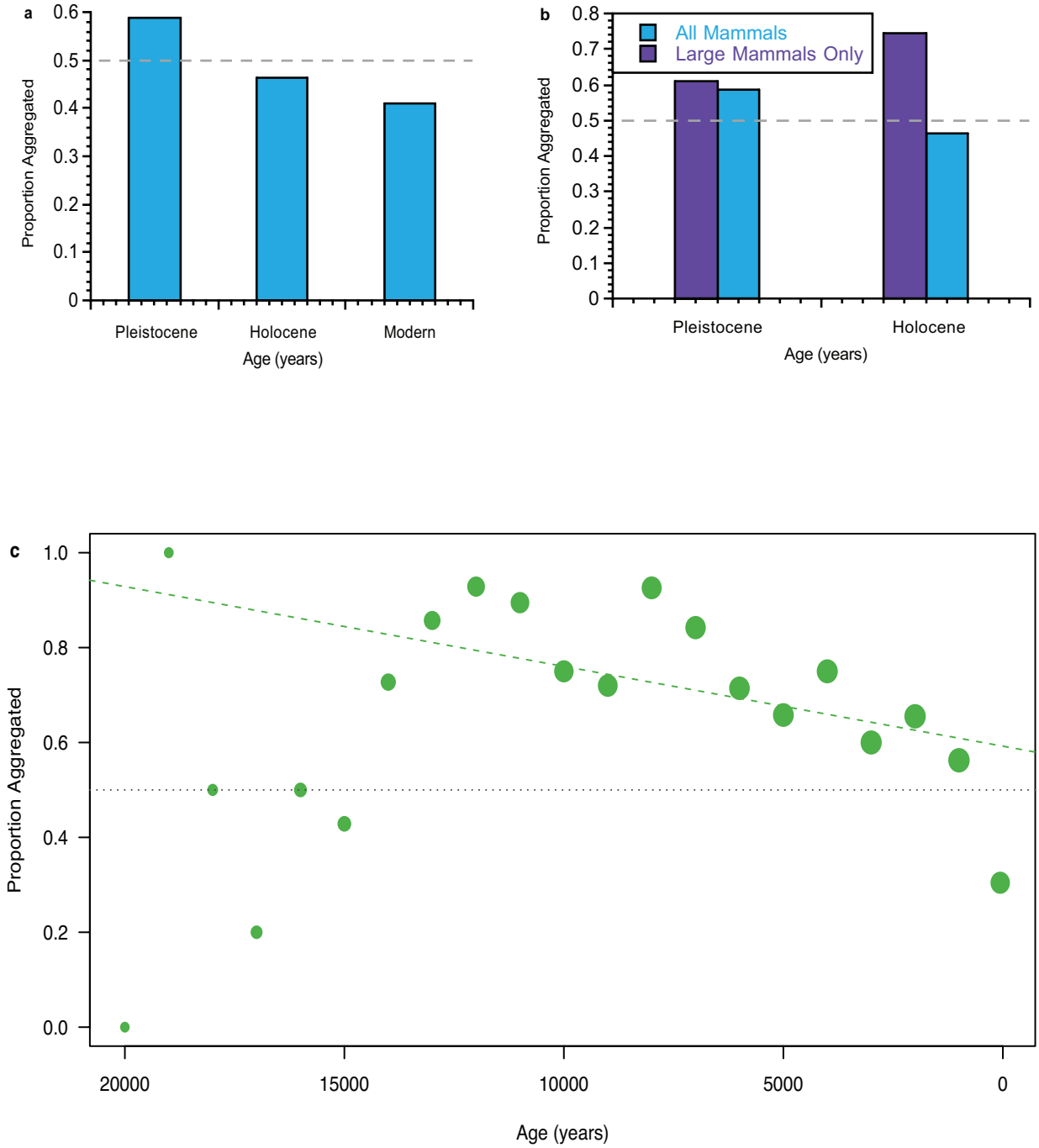
Extended Data Figure 1 | Map showing distribution of fossil data sets. Polygons enclose the localities for each fossil data set included in our analyses. Mammals are in blue, plants are in green. Dark colours represent data sets that are older. This map was created using ArcGIS software by Esri and can be found at http://www.arcgis.com/home/item.html?id=c61ad8ab017d49e1a82f580ee1298931. ArcGIS and ArcMap are the intellectual property of Esri and are used herein under license. Copyright © Esri. All rights reserved. For more information about Esri software, please visit http://www.esri.com.



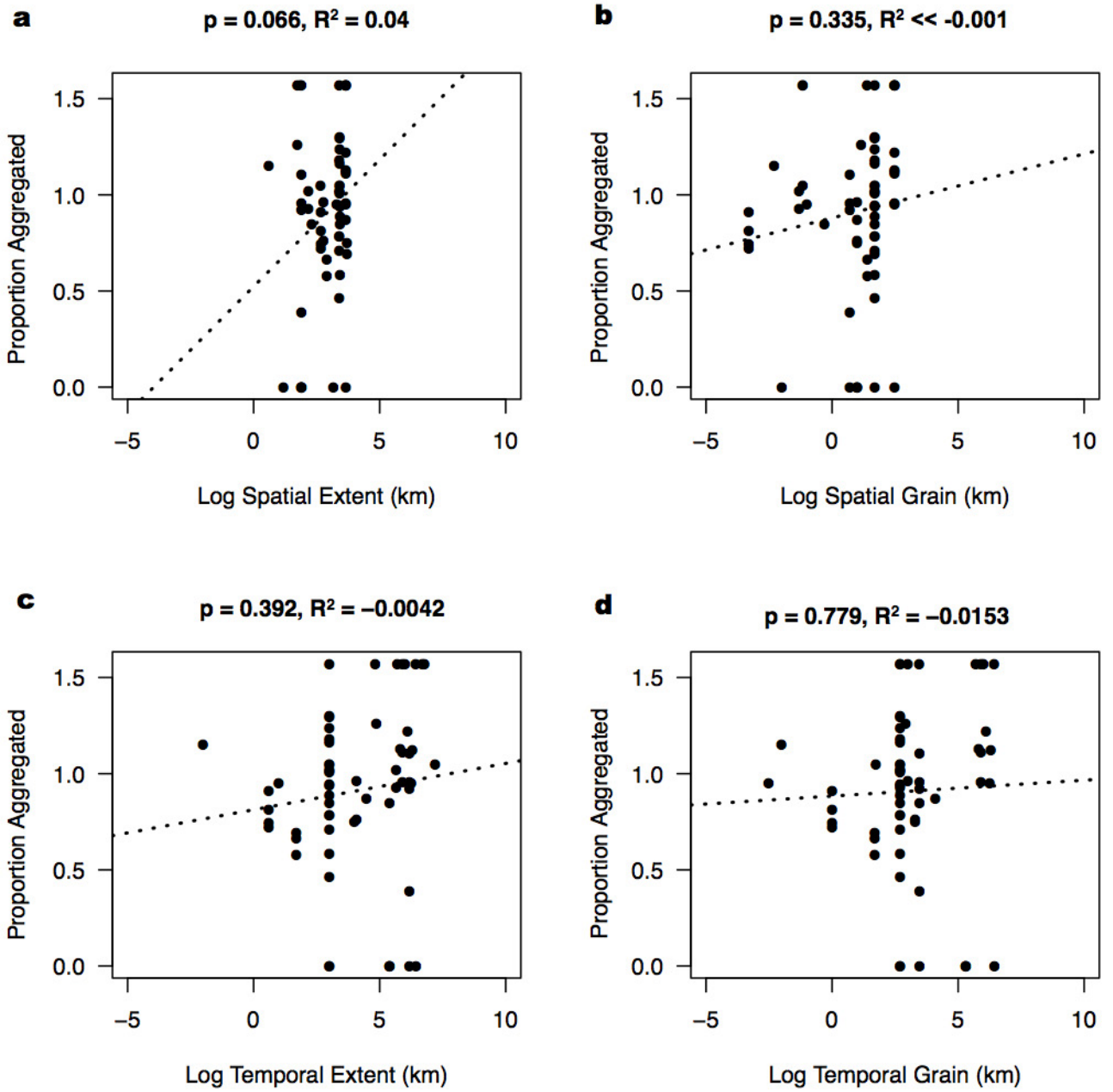
Extended Data Figure 2 | Breakpoint analysis of the composite data. The analysis was performed on all data including the islands (see Fig. 3 main text), showing the mean estimate (red point; $10^{3.778}$ years) and 95% confidence interval (error bar at base of plot; $10^{1.606}$, $10^{5.951}$ years) of the initiation of reduced percentage of aggregated species pairs, as well as the mean and confidence intervals around the change in slope of the two resulting linear models. The breakpoint analysis was run using all the data resolved to the best possible dates to allow for the greatest amount of power in detecting where the switch occurred. However, the results were similar when island data were excluded.



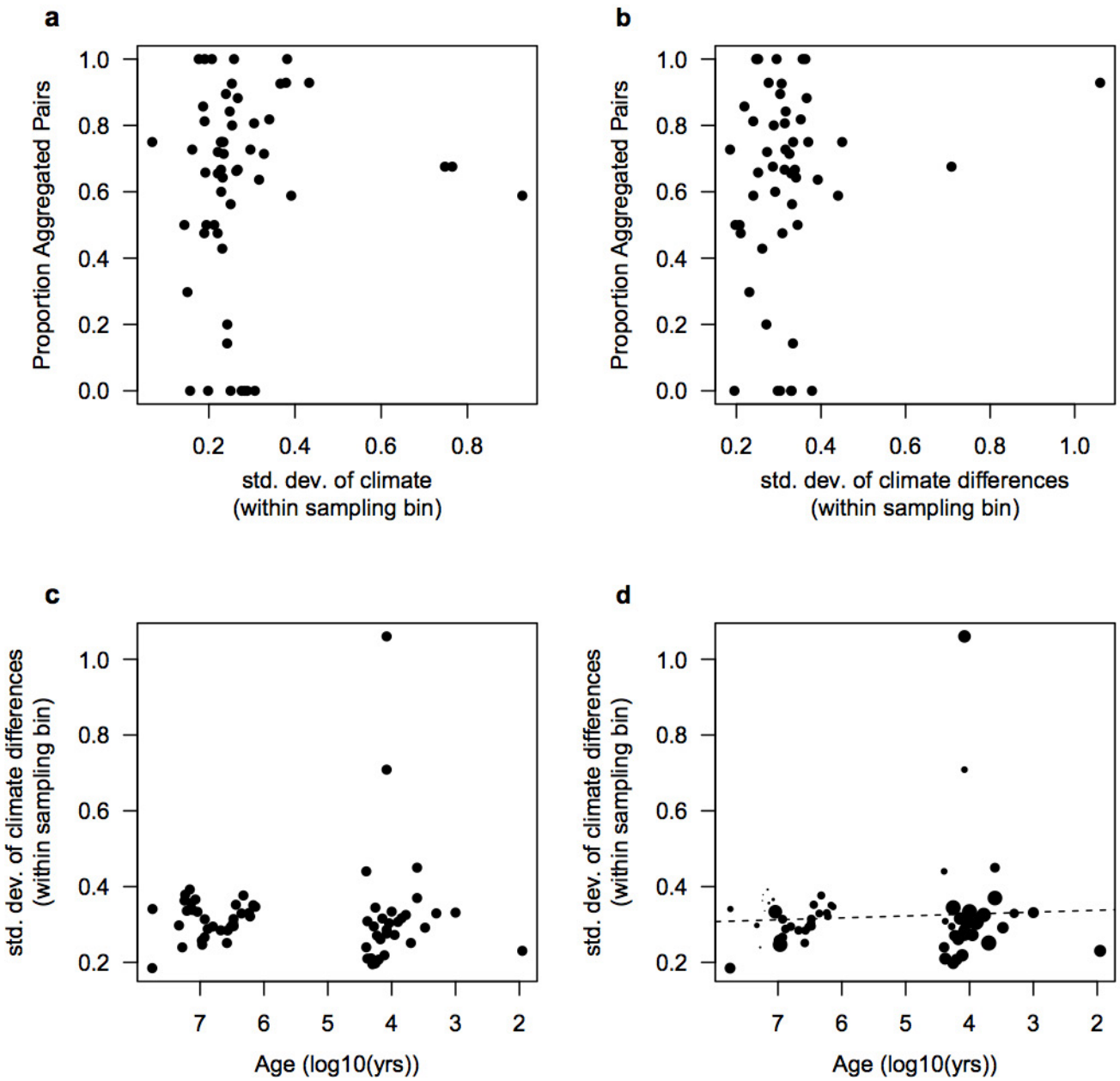
Extended Data Figure 3 | Weighted Loess curve with and without modern data. Loess curve weighted by number of sites with shaded 95% confidence intervals illustrates the reduction in the proportion of aggregated species pairs towards the present. Data are analysed with (black line and shading) and without (red line and shading) the modern data. Colours indicate continent: North America (green), Eurasia (purple), Australia (dark grey), South America (dark blue), Africa (orange). Point shapes indicate type of data: pollen (square), mammals (triangle), macroplants (circle). Data on terrestrial communities from ref. 2 are diamonds. Only mainland assemblages were included in the calculation for the weighted Loess curve and the density plots here and in Fig. 1.



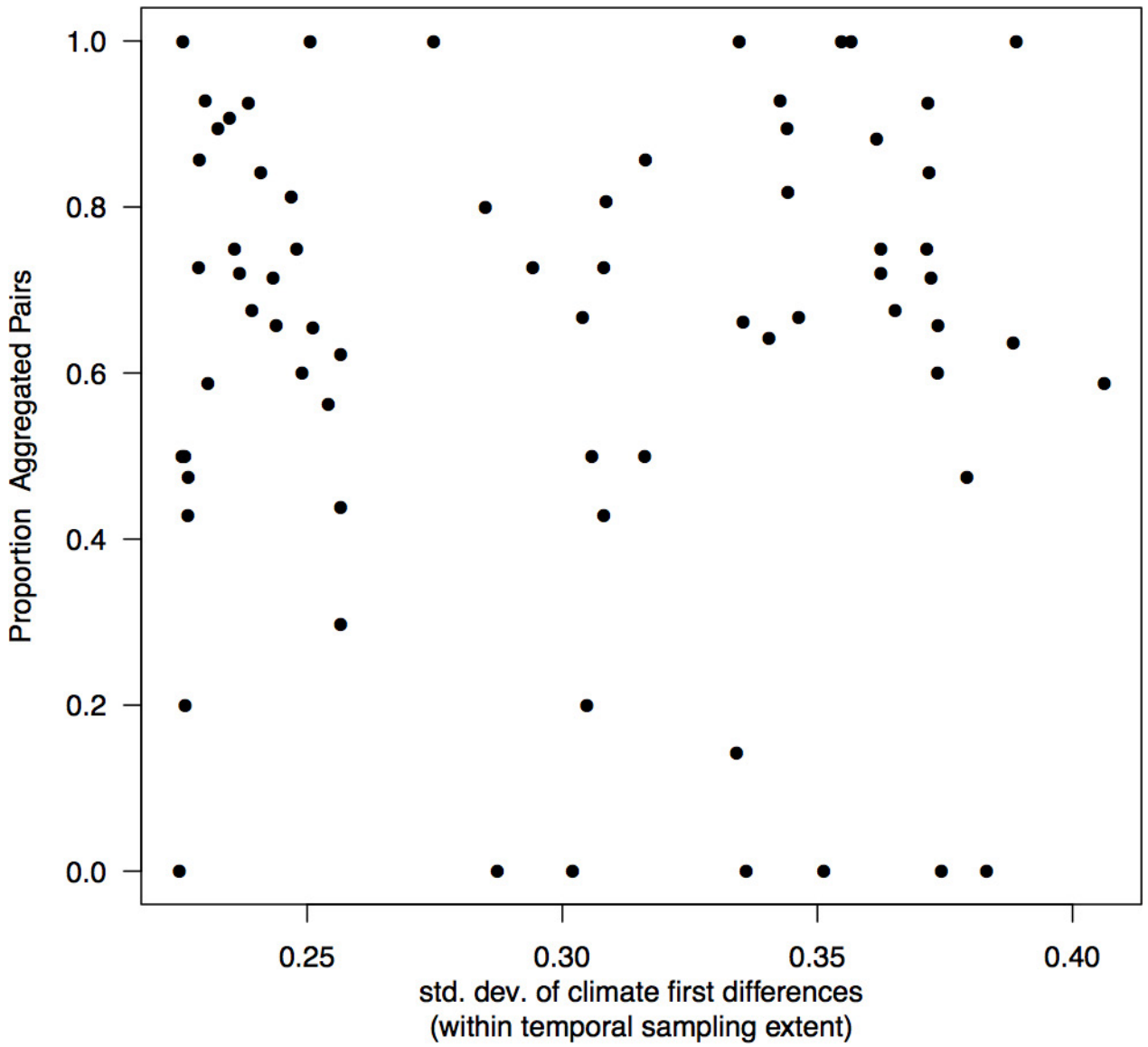
Extended Data Figure 4 | Results of PAIRS analyses of two Pleistocene–Holocene fossil data sets. a, Mammal data for three periods: late Pleistocene (40,000–10,000 years ago), Holocene (10,000–500 years ago), and modern (present, literature survey data). Note the switch from >50% aggregated pairs to <50% aggregated pairs occurs in the Pleistocene versus Holocene data sets. **b**, Results for large and small mammals separately. There is a significant difference ($P < 0.001$) between the Holocene and the Pleistocene for all mammals (blue bars) and for large mammals (purple bars) only ($P = 0.015$). However, the direction of the shift was different. For all mammals, there were fewer positive associations in the Holocene, whereas, for large mammals only, there were more positive associations in the Holocene. **c**, North American pollen data from the past 21,000 years (modified from ref. 20). Data are from cores resolved into 1,000-year time slices. The size of the circle is related to the number of sites in the data set. The point at 0 represents a period from the present to 1,000 years ago, but is sampled from the top of the pollen cores using the same methodology as the older time slices. Note the trend of decreasing percentage of aggregations towards the present, especially in times with the largest numbers of sites (after 14,000 years). A weighted regression that takes into account the number of sites in each time slice is significant (dashed green line; $P = 0.04$, adjusted $R^2 = 0.15$). The final time slice at 0 records a shift from a dominance of aggregated pairs to a dominance of segregated pairs. The sampling methods and data structure are the same for all time slices. Grey dashed line is at 50% in each panel.



Extended Data Figure 5 | Relationship between the proportion of aggregated pairs and scale. The proportion of significant pairs that are aggregated is not the result of temporal or spatial scale of data. Arcsine transformation of the proportion of significant pairs that are aggregated plotted as a function of spatial (**a**, **b**) or temporal (**c**, **d**) grain (**b**, **d**) or extent (**a**, **c**). Linear regressions are non-significant and adjusted R^2 values are extremely low.



Extended Data Figure 6 | Climate variability measured during the temporal extents of the fossil data sets. Proportion of significant pairs that are aggregated shows no relationship with climate variability within a time interval. **a**, **b**, Climate variability was quantified as the standard deviation of all climate proxy data for that time interval (**a**), or the standard deviation of the first differences in climate across the interval (**b**). **c**, **d**, Climate variability (standard deviation of first differences) and age of data sets show no relationship (**c**), suggesting no trend in climate variability sampled by the fossil data sets across the Phanerozoic. There is a significant, but weak, positive relationship (**d**, dashed line) between climate variability and decreasing age of the data sets when the linear model is weighted by sample size of climate proxy data enveloped by the temporal window of the fossil data sets ($P = 0.007$, adjusted $R^2 = 0.0998$).



Extended Data Figure 7 | Relationship between proportion of aggregated pairs and fixed-width time intervals. High-amplitude Pleistocene climate variability oscillating between glacial and interglacial cycles may have imposed its own novel pressures on floral and faunal communities. Furthermore, ecological impacts may lag behind climate episodes themselves, complicating efforts to quantify climatic links to changes in the proportion of aggregated species pairs over time. Thus, limiting our measure of climate variability to the temporal span of the data sets themselves may potentially not account for important (and possibly ecologically significant) climatic variability from the previous millennia. To incorporate this possibility, we re-analysed the relationship between the proportion of aggregated species pairs and climate variability of each data set, but included climate across the preceding 100,000 years, 10,000 years (not shown), and 1,000 years (not shown). As in the more restrictive analysis (Fig. 3a), there is consistently no relationship between climate variability and the proportion of aggregated species pairs.

Extended Data Table 1 | Raw data for Fig. 1

Dataset	Data Owner	Type	Cont	#Rand	#Agg	#Seg	#Spp	#Sites	Age (yr)	Temp Grain (yr)	Temp Extent (yr)	Spat Grain (km)	Spat Extent (km)
Turkana-Natoo	Behrensmeyer	M	AF	253	0	0	23	10	1365000	200000	250000	10	80
Turkana-Okote	Behrensmeyer	M	AF	1081	0	0	47	16	1470000	200000	250000	10	80
Turkana-Kaitio	Behrensmeyer	M	AF	351	0	0	27	8	1650000	200000	250000	10	80
Turkana-KBS	Behrensmeyer	M	AF	2015	0	1	64	37	1715000	200000	250000	10	80
Turkana-Kalochoro	Behrensmeyer	M	AF	276	0	0	24	7	2100000	200000	250000	10	80
Turkana-Burgi	Behrensmeyer	M	AF	1533	0	7	56	23	2250000	200000	250000	10	80
Turkana-Lomekwi	Behrensmeyer	M	AF	990	0	0	45	10	2985000	200000	250000	10	80
Turkana-TuluBor	Behrensmeyer	M	AF	778	0	2	40	12	3035000	200000	250000	10	80
Turkana-LokochotMoiti	Behrensmeyer	M	AF	496	0	0	32	7	3705000	200000	250000	10	80
Siwalik-3.2-3.6Ma	Behrensmeyer	M	EA	210	0	0	21	16	3300000	3000	1500000	5	80
Siwalik-6-5Ma	Behrensmeyer	M	EA	820	0	0	41	30	6300000	3000	1500000	5	80
Siwalik-7Ma	Behrensmeyer	M	EA	4460	4	1	95	111	7600000	3000	1500000	5	80
Siwalik-8Ma	Behrensmeyer	M	EA	5050	0	0	101	170	8500000	3000	1500000	5	80
Siwalik-9Ma	Behrensmeyer	M	EA	4186	0	0	92	177	9400000	3000	1500000	5	80
Siwalik-11-10Ma	Behrensmeyer	M	EA	10703	4	24	147	92	11000000	3000	1500000	5	80
Siwalik-13-12Ma	Behrensmeyer	M	EA	8376	6	3	130	168	13100000	3000	1500000	5	80
Siwalik-15-14Ma	Behrensmeyer	M	EA	4454	7	4	95	38	14500000	3000	1500000	5	80
Siwalik-17-16Ma	Behrensmeyer	M	EA	2210	0	1	67	20	17100000	3000	1500000	5	80
Eurasia-spp-MN17	Eronen	M	EA	3644	9	2	86	32	2750000	650000	650000	300	4500
Eurasia-spp-MN16	Eronen	M	EA	15545	25	6	177	53	3000000	800000	800000	300	4500
Eurasia-spp-MN15	Eronen	M	EA	5993	2	0	110	27	3800000	800000	800000	300	4500
Eurasia-spp-MN14	Eronen	M	EA	2145	0	0	66	28	4750000	1100000	1100000	300	4500
Eurasia-spp-MN11	Eronen	M	EA	14532	2	1	171	44	8450000	800000	800000	300	4500
Eurasia-spp-MN10	Eronen	M	EA	8514	1	0	131	42	9250000	500000	500000	300	4500
Eurasia-spp-MN07	Eronen	M	EA	9977	30	4	142	65	11850000	1300000	1300000	300	4500
Eurasia-spp-MN06	Eronen	M	EA	9581	10	0	139	87	13850000	2700000	2700000	300	4500
Eurasia-spp-MN05	Eronen	M	EA	14119	51	26	169	96	16100000	1800000	1800000	300	4500
Eurasia-spp-MN04	Eronen	M	EA	7994	7	0	127	65	17500000	1000000	1000000	300	4500
Eurasia-spp-MN03	Eronen	M	EA	4803	39	9	99	57	19000000	2000000	2000000	300	4500
Eurasia-spp-MN02	Eronen	M	EA	3402	0	1	83	35	21400000	2800000	2800000	300	4500
SAfr-LgMammals-MIS1	Faith	M	AF	1091	25	12	48	103	12000	1000	12000	10	600
SAfr-LgMammals-MIS2	Faith	M	AF	906	19	21	44	28	24000	2000	12000	10	600
NAm-mammals-Modern	Lyons	M	NA	9267	541	777	146	67	30	50	50	50	5000
NAm-mammals-Holo	Lyons	M	NA	9860	404	467	147	214	5000	2000	10000	10	5000
NAm-mammals-Pleist	Lyons	M	NA	12727	667	467	167	88	25000	12500	30000	10	4500
crypticTaxaDesertRodents	Riddle	M	NA	1832	39	20	62	171	7.5	0.003	10	0.1	2000
African-SPP-MOD	Tóth	M	AF	15795	52	84	179	14	30	50	50	25	800
African-SPP-HIST	Tóth	M	AF	15632	36	85	178	14	90	50	50	25	800
WI-overstory-SPP-1950	Waller	PI	NA	512	26	23	34	168	60	1	4	0.0005	475
WI-understory-SPP-1950	Waller	PI	NA	12726	596	706	168	152	60	1	4	0.0005	475
WI-overstory-SPP-2000	Waller	PI	NA	688	33	20	39	168	10	1	4	0.0005	475
WI-understory-SPP-2000	Waller	PI	NA	12455	400	511	164	152	10	1	4	0.0005	475
Abo	DiMichele	PI	NA	934	9	3	44	207	290000000	55	16000000	0.07	450
PennWolf	DiMichele	PI	NA	1429	2	0	54	28	300000000	500	5000000	0.07	80
Rotliegend	DiMichele	PI	EA	988	2	0	45	62	300000000	1000	6000000	0.07	72
Calhoun Coal	DiMichele	PI	NA	378	0	0	28	55	304000000	500	10000	0.04	50
Kootenai	Jud	PI	NA	494	0	2	32	17	115000000	500	1000	0.01	15
Molteno5	Labandeira	PI	AF	3655	0	0	94	8	234000000	3000	250000	0.5	200
Molteno4	Labandeira	PI	AF	2278	0	0	68	12	234500000	3000	250000	0.5	200
Molteno3	Labandeira	PI	AF	8778	0	0	133	20	235000000	3000	250000	0.5	200
Molteno2	Labandeira	PI	AF	6312	9	7	113	43	235500000	3000	250000	0.5	200
Molteno1	Labandeira	PI	AF	8256	0	0	129	9	236000000	3000	250000	0.5	200
Burgersdorp	Labandeira	PI	AF	300	0	0	25	8	244000000	3000	250000	0.5	200
EarlyEocenePlants	Wing	PI	NA	3086	99	55	81	78	55600000	500	460000	0.05	150
LatePaleocenePlants	Wing	PI	NA	2101	32	12	66	42	56200000	500	460000	0.05	150
BigCedarRidgePlants	Wing	PI	NA	8238	15	3	129	100	73000000	0	0	0.005	4
pollenPG	Bercovici	Po	NA	3808	20	0	88	150	66027000	2865	68000	25	55
pollenK	Bercovici	Po	NA	4041	49	5	91	71	66065000	825	73000	15	55
NorthAm-pollen-0ka	Blois	Po	NA	3501	21	48	85	252	60	500	1000	50	2645.993
NorthAm-pollen-1ka	Blois	Po	NA	3522	27	21	85	445	1000	500	1000	50	2645.993
NorthAm-pollen-2ka	Blois	Po	NA	3541	19	10	85	438	2000	500	1000	50	2645.993
NorthAm-pollen-3ka	Blois	Po	NA	3286	21	14	82	410	3000	500	1000	50	2645.993
NorthAm-pollen-4ka	Blois	Po	NA	3285	27	9	82	397	4000	500	1000	50	2625.521
NorthAm-pollen-5ka	Blois	Po	NA	3202	25	13	81	397	5000	500	1000	50	2625.521
NorthAm-pollen-6ka	Blois	Po	NA	3472	10	4	84	354	6000	500	1000	50	2625.521
NorthAm-pollen-7ka	Blois	Po	NA	3141	16	3	80	335	7000	500	1000	50	2625.521
NorthAm-pollen-8ka	Blois	Po	NA	2899	25	2	77	300	8000	500	1000	50	2623.018
NorthAm-pollen-9ka	Blois	Po	NA	2901	18	7	77	280	9000	500	1000	50	2623.018
NorthAm-pollen-10ka	Blois	Po	NA	2528	21	7	72	251	10000	500	1000	50	2554.072
NorthAm-pollen-11ka	Blois	Po	NA	2466	17	2	71	205	11000	500	1000	50	2554.072
NorthAm-pollen-12ka	Blois	Po	NA	2542	13	1	72	155	12000	500	1000	50	2554.072
NorthAm-pollen-13ka	Blois	Po	NA	2542	12	2	72	117	13000	500	1000	50	2532.103
NorthAm-pollen-14ka	Blois	Po	NA	2200	8	3	67	76	14000	500	1000	50	2532.103
NorthAm-pollen-15ka	Blois	Po	NA	1946	3	4	63	50	15000	500	1000	50	2532.103
NorthAm-pollen-16ka	Blois	Po	NA	1826	2	2	61	39	16000	500	1000	50	2532.103
NorthAm-pollen-17ka	Blois	Po	NA	1480	1	4	55	30	17000	500	1000	50	2532.103
NorthAm-pollen-18ka	Blois	Po	NA	1324	1	1	52	21	18000	500	1000	50	2532.103
NorthAm-pollen-19ka	Blois	Po	NA	1484	1	0	55	16	19000	500	1000	50	2531.385
NorthAm-pollen-20ka	Blois	Po	NA	1481	0	4	55	15	20000	500	1000	50	1481.922
NorthAm-pollen-21ka	Blois	Po	NA	1176	0	0	49	13	21000	500	1000	50	2134.789

Numbers of aggregated versus segregated pairs and spatial and temporal scale of the ETE data sets included in this analysis. M, mammals; PI, macroplants; Po, pollen. AF, Africa; EA, Eurasia; NA, North America. #Rand, the number of taxon pairs that were not significantly different from random. #Agg, the number of significant taxon pairs that were aggregated. #Seg, the number of significant taxon pairs that were segregated. #Spp, the number of species in the data set. #Sites, the number of sites in the data set. Age (yr) is the midpoint age of the data set. Temp Grain (yr), temporal grain in years or the average amount of time encompassed by a site in the data set. Temp Extent (yr), the maximum amount of time encompassed by a data set. Spat Grain (km), the average distance from a site that fossils were transported. Spat Extent (km), the maximum linear distance encompassed by the data set.

Extended Data Table 2 | Effect of taxonomic resolution

Dataset	% Aggregations for Species	% Aggregations for Genera	Direction of difference when lumping
WI Overstory Plants-1950	0.531	0.800	Increase
WI Overstory Plants-2000	0.622	0.773	Increase
Eurasia mammals MN02	0	0	No change
Eurasia mammals MN03	0.815	0.732	Decrease
Eurasia mammals MN04	1	0.75	Decrease
Eurasia mammals MN05	0.662	0.608	Decrease
Eurasia mammals MN06	1	0.75	Decrease
Eurasia mammals MN07	0.882	0.864	Increase
Eurasia mammals MN09	1	0.75	Decrease
Eurasia mammals MN10	0.667	0.857	Increase
Eurasia mammals MN11	NA	1	Increase
Eurasia mammals MN14	NA	NA	No change
Eurasia mammals MN15	1	1	No change
Eurasia mammals MN16	0.806	1	Increase
Eurasia mammals MN17	0.818	0.619	Decrease
Africa mammals - historical	0.298	0.028	Decrease
Africa mammals – modern	0.382	0.031	Decrease
Great Basin Rodents Cryptic	0.661	0.500	Decrease

Change in the proportion of significant pairs when data sets are analysed at the species and genus levels. If lower taxonomic resolution of fossil data sets is driving the pattern of increased aggregations in the fossil data, we would expect to see increases in the percentage of aggregations when data are analysed at the genus level. Instead, most data sets show a decrease in the percentage of aggregated pairs. Only 6 of the 18 data sets analysed at multiple taxonomic resolutions show the expected increase. Nine show a decrease and three show no change. One data set (Great Basin Rodents Cryptic) was analysed at the species level and then taxonomically resolved with genetic data to include cryptic species. For that data set only, '% aggregations for genera' corresponds to the data set with cryptic species lumped and '% aggregations for species' corresponds to the data set with cryptic species resolved.

Extended Data Table 3 | Proportion of aggregated pairs across critical intervals

Taxon	Event	Place	Trend in % Positive Associations	p value
Large Mammals	Pleistocene-Holocene Transition	South Africa	Increase	0.007
Mammals	Pleistocene-Holocene Transition	North America	Decrease	<0.001
Pollen	Pleistocene-Holocene Transition	North America	Decrease	0.0008
Plants	PETM	North America	No Change	0.556
Pollen	K-Pg	North America	No Change	0.853
Pollen	Permian Crises	Greenland	No significant pairs	na

Significance of change in positive versus negative associations across critical intervals.

Landscapes carved by humans, as seen on the California-Arizona border, may disrupt species associations.

PHOTO: GOOGLE IMAGES

PALEOECOLOGY

Human impacts on ecosystems began thousands of years ago

Early humans broke up existing plant and animal networks, perhaps boosting extinction risks

By **Elizabeth Pennisi**

Silicon Valley entrepreneurs boast about how disruptive their creations are, from cellphones to Uber. But in a global sense, humans' most disruptive technologies may have been born millennia ago, when our ancestors began to plant crops and cross the globe. Just in the past 150 years, humans have taken such a toll on Earth's status quo that some researchers say we have ushered in a new geological time period: the Anthropocene. But scientists are now suggesting a new dimension to human impacts, one that began many thousands of years ago. Long before the Industrial Age, humans had broken up relationships among plants and animals that had been stable for millions of years, S. Kathleen Lyons, a paleoecologist at the Smithsonian National Museum of Natural History in Washington, D.C., and her colleagues report this week in *Nature*.

The work, based on an analysis of species distributions, "gives us an idea of how far we've already come in changing the planet," says Anthony Barnosky, a paleobiologist at the University of California (UC), Berkeley.

The paper also suggests that extinction isn't the only marker of ecological disruption. It and other recent findings "show how major shifts in the diversity of life are not only quantified by the number of species, but by changes in entire ecological communities," says Nick Haddad, an environmental scientist at North Carolina State University in Raleigh. "They elevate the disastrous consequences of the sixth mass extinction"—the one happening now.

Researchers have debated when human impacts became dramatic enough to mark the start of the Anthropocene. Some pinpointed the start of the Industrial Age, with its accelerated population growth, habitat and species loss, and rising carbon dioxide levels. Others have suggested that human-caused disruption began much earlier, when we became superpredators, able to take all kinds of prey, or when we began farming (*Science*, 7 October 2011, p. 32).

The new data imply an early start to the human-dominated era. As part of the Evolution of Terrestrial Ecosystems Program at the Smithsonian Institution, paleontologists and ecologists pooled 172 data sets on the species present in specific regions and time periods, mostly in North America, spanning the past 300 million years. Lyons and her colleagues assessed how tightly linked species were in these assemblages, analyzing how often about 350,000 pairs of species appeared together, as compared with the clustering seen in a randomly generated assemblage. Pairs of species that clumped together more often than by chance presumably shared resources, and the persistence of these relationships likely stabilizes a community, Lyons says.

Since 300 million years ago, an average of 64% of the nonrandom pairs appeared together and qualified as aggregations, they report. In North America, for example, the now-extinct dire wolf (famous pets in the *Game of Thrones* TV series) and giant ground sloth persisted as a pair, likely connected as predator and prey; bog lemmings and chipmunks were also found together, likely because of common habitat preferences. But a big shift occurred, roughly about 6000 years ago and certainly within the last 10,000 years: Only 37% of the nonrandom pairs of species living now are aggregations. At first, even Lyons's team doubted this result. "It took 2 years to convince ourselves that [the result] was real," she recalls.

Others now are confident as well. "A strength of the paper is the diversity of data sets (mammals, pollen, and plant macrofossils)," emailed Jennifer Marlon, a Yale University environmental scientist. Agrees team member Jessica Blois, a paleoecologist at UC Merced: "The pattern isn't seen in just one system or one time, but emerges across both plants and animals." Even across some ancient mass extinctions, such as that at the end of the Cretaceous Period, species tended to aggregate more than they do now, although the fossil data weren't detailed enough to clearly reveal ecosystem structure during such extinctions, Lyons says.

The loss of species aggregations was not seen during other periods of major climate changes. But by 6000 years ago, humans were migrating around the world; in North America, some had started to settle down and practice agriculture. If the loosening of ecosystem associations is a uniquely human effect, then "the processes that are going on now with humans in the picture are fundamentally different than the processes that have gone on before," says Erle Ellis, an ecologist at the University of Maryland, Baltimore County. "There are going to be a lot of people thinking about this for a while." ■

relatively large volume of space between the companion star and the black hole, which allows a large disk to form. But the supply of infalling matter from the companion star is insufficient to fill such a large disk with a steady flow. Without a steady flow, the accretion rate becomes unstable and can fluctuate violently (Fig. 1). These fluctuations, in turn, trigger oscillating emissions of energetic X-ray photons near the black hole, which then light up the whole disk with the observed pulsating visible effects.

But the authors show that this explanation requires the disk to be very large, close to its maximum possible size. Moreover, the X-ray oscillations that they observed from V404 Cygni are much stronger than the visible-light ones. These puzzling facts will need to be accounted for. How, and whether, the jet of the black hole tracks these oscillations is also yet to be determined. The proposed parallels between the observed oscillations and those of GRS 1915+105 will undoubtedly be investigated in detail in the future. This will help researchers to understand the above issues in light of the wealth of supporting observations currently being analysed by astronomers the world over.

Black-hole outbursts are unpredictable and some can be two weeks or even shorter in duration, so worldwide coordination and round-the-clock monitoring is essential if we are to understand the physics of these extreme events. This becomes particularly challenging when coordinating observations between space telescopes and those on the ground. The outburst of V404 Cygni last year invigorated the efforts of black-hole astronomers to tackle these challenges, with at least one conference dedicated entirely to this theme. Amateurs can also play a key part in this effort. Kimura and colleagues gathered data from many small telescopes, some with optical elements only 20 centimetres in diameter, showing that, in astronomy, size is not necessarily what matters; collaboration does. ■

Poshak Gandhi *is in the Department of Physics & Astronomy, University of Southampton, Highfield, Southampton SO17 1BJ, UK.*
e-mail: p.gandhi@soton.ac.uk

1. Bowyer, S, Byram, E. T., Chubb, T. A. & Friedman, H. *Science* **147,** 394–398 (1965).
2. Webster, B. L. & Murdin, P. *Nature* **235,** 37 (1972).
3. Kimura, M. *et al. Nature* **529,** 54–58 (2015)
4. Fender, R. P. & Belloni, T. *Annu. Rev. Astron. Astrophys.* **42,** 317 (2004).
5. Casares, J., Charles, P. A. & Naylor, T. *Nature* **355,** 614–617 (1992).
6. Khargharia, J., Froning, C. S. & Robinson, E. L. *Astrophys. J.* **716,** 1105 (2010).
7. Miller-Jones, J. C. A. *et al. Astrophys. J.* **706,** L230–L234 (2009).

ECOLOGY

Different worlds

Patterns of species association reveal that terrestrial plant and animal communities today are structured differently from communities spanning the 300 million years that preceded large-scale human activity. SEE LETTER P.80

GREGORY P. DIETL

The British author L. P. Hartley wrote in one of his best-known novels, *The Go-Between*, that "The past is a foreign country: they do things differently there." This poignant imagery of remoteness from the past captures the essence of an emerging global consciousness. Human hegemony over nature has become so pervasive and profound that it is quite possible that we have created a world that has little or no precedent — in ecological parlance, it has no analogue. On page 80 of this issue, Lyons *et al.*[1] detail a compelling case that this extraordinary situation is an undeniable reality for the rules that govern how plant and animal communities are structured.

The authors assembled data on the presence and absence of terrestrial plant and animal taxa for 80 fossil and modern assemblages in North America, Africa and Eurasia, spanning the past 300 million years. Using a statistical approach that was designed to compare occurrence data against a randomized 'null' assemblage, they quantified the fraction of species pairs in each assemblage that deviated from random expectations about where they should be found. Species pairs meeting this criterion provide valuable insight into the ecological processes that structure communities[2,3].

In modern communities, most species pairs show random co-occurrence, but those that are non-randomly associated are typically segregated — that is, they tend to co-occur less frequently than would be expected by chance[2] (Fig. 1). Lyons *et al.* wanted to know whether the fossil record is consistent with this pattern of species segregation. The headline finding is that the pattern of co-occurrence dominating modern communities departs sharply from that of the past. As in modern communities, co-occurrence of most pairs was random. But unlike in modern communities, the non-random associations were dominanted by aggregated species pairs, which co-occur more frequently than would be expected by chance[2] (Fig. 1). This dominance of aggregated pairs persisted with little change for more than 300 million years on different contintents and across diverse taxa, until about 6,000 years ago, when the sharp transition to the segregated co-occurrence pattern began.

After running a battery of tests to ascertain that this temporal trend was not an artefact,

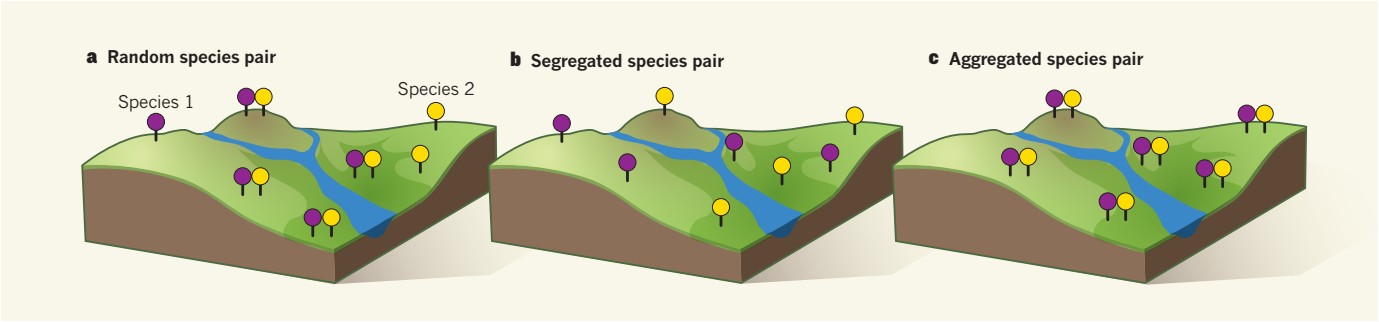


Figure 1 | Species associations. In a terrestrial ecological community, any two species may occur randomly (**a**) at locations in a landscape. Alternatively, species pairs may be non-randomly associated, in which case they can be either segregated (**b**), meaning that the two species co-occur less frequently than would be expected by chance, or aggregated (**c**), meaning that they co-occur more frequently than expected by chance. Among non-randomly associated pairs, Lyons *et al.*[1] document a shift from a dominance of aggregated pairs before the expansion of human populations to the segregated pattern typically seen today. (Figure adapted from ref. 7.)

the authors speculate that an expanding human population may explain why species co-occurrence patterns are so different today. The shift was most obvious in North American assemblages (where the most occurrence data were available) and coincided with the inexorable spread of agriculture in this region. The authors propose that habitat fragmentation and limitations on species dispersal associated with land use were probably the main engines driving the shift. The structure of plant and animal terrestrial communities would never be the same again.

This interpretation is sure to attract fervent debate and lead to further research to confirm the pattern and disentangle the proposed mechanisms involved. The tension between the distant past and the familiar present that the study highlights, however, has an underlying implication that may not be as obvious. If the past is different from the present (in this case not in the immanent processes that were operating, but in their frequency), its applicability to our current societal need to anticipate ecological changes and design adaptation measures — a goal that Lyons *et al.* acknowledge is a priority — is not immediately manifest. There is no easy way around this tension. At stake is whether we can reliably use the past as a guide to an uncertain, anthropogenically modified future.

A small cadre of voices argues that a human-dominated present limits the use of the past as a key to unlocking the future[4]. In this view, the world we live in today, and the immediate future that our grandchildren will inherit, has no analogue in the geological past. As a consequence, referencing 'natural experiments' in the distant past as a guide to predict what might happen, now or in the future, is a flawed strategy. Out is the use of uniformitarianism[5] as a guiding principle, and in is a new kind of 'post-normal' science[6]. Lyons and colleagues' study of human impacts on community-assembly rules, at least as implicated by species co-occurrence patterns, seems to embody evidence for this no-analogue world.

A more optimistic view of this tension between the past and present — one that acknowledges that processes change and interact in complex ways over time, whether human action is involved or not — is that it poses a challenge for how we select analogues from the past to gain insight into future conditions. Lyons and colleagues' finding is a stark reminder that analogue selection often over-stresses likenesses at the expense of differences. However, small and unknowable differences in starting points may overwhelm the signal of the likenesses, making analogue selection a risky business. To use the past as a guide, we must select from the dense fabric of likenesses and differences that was its contingent state at a moment in time, and apply only those particular events and conditions relevant to our present needs.

Moving beyond this tension will require creative ways of thinking about how we use the distant past to improve our understanding of the present and our anticipation of the future, which may provide a ground for wiser action. Lyons and colleagues' study is an excellent entry point into thinking about this problem. ■

Gregory P. Dietl *is at the Paleontological Research Institution, Ithaca, New York 14850, USA and in the Department of Earth and Atmospheric Sciences, Cornell University, Ithaca.*
e-mail: gpd3@cornell.edu

1. Lyons, S. K. *et al. Nature* **529**, 80–83 (2015).
2. Gotelli, N. J. & Ulrich, W. *Oecologia* **162,** 463–477 (2010).
3. Connor, E. F., Collins, M. D. & Simberloff, D. *Ecology* **94,** 2403–2414 (2013).
4. Knight, J. & Harrison, S. *Anthropocene* **5,** 71–75 (2014).
5. Baker, V. R. *Anthropocene* **5,** 76–79 (2014).
6. Funtowicz, S. O. & Ravetz, J. R. *Futures* **25,** 739–755 (1993).
7. Blois, J. L. *et al. Ecography* **37,** 1095–1108 (2014).

This article was published online on 16 December 2015.

VIROLOGY

Host protein clips bird flu's wings in mammals

The polymerase enzyme from avian influenza A viruses does not function well in human cells. The protein ANP32A has been identified as the cellular factor mediating a major component of this host restriction. SEE LETTER P.101

ANICE C. LOWEN

Influenza A viruses circulate in diverse natural hosts, including mammalian and avian species. Yet transmission of these viruses between mammals and birds occurs only rarely, owing to host restriction: an influenza A virus that is adapted to an avian host typically does not grow well in a mammalian host, and vice versa. When such restrictions are overcome and an avian virus transmits to humans, a pandemic can occur. On page 101 of this issue, Long *et al.*[1] report a breakthrough in understanding the restriction of avian influenza viruses in mammals.

The protein PB2 is a necessary component of the influenza A polymerase enzyme complex, which copies the viral genome and thus is essential for viral replication. For many years, researchers have known that a specific domain of PB2, the 627 domain, is involved in host restriction[2]. H5N1 strains and other 'bird flu' viruses rapidly acquire mutations in this domain following transmission to humans or inoculation of mammals in the laboratory[3,4]. These mutations, in turn, greatly enhance the growth, virulence and transmission of avian influenza A viruses in mammals[5–8]. Yet despite intense effort, the host factors and mechanisms that limit the functionality of non-mutated avian-adapted PB2 proteins in mammalian cells[9–14] remained obscure.

Long *et al.* knew from previous work[15] that the avian-adapted PB2 did not work well in mammalian cells because of the absence of a factor that enhances polymerase activity in avian cells, rather than because of the presence of an inhibitory factor in mammalian cells. To identify the missing positive factor, the authors used a panel of hybrid hamster cell lines that each carried a different fragment of the chicken genome. They expressed an avian-adapted PB2 protein, along with its essential viral partner proteins, in each cell line and measured the activity of the viral polymerase (Fig. 1). Out of 53 hybrid cell lines tested, four showed robust activity of the avian-adapted polymerase complex. By identifying the chicken genes that were shared by these four cell lines, Long *et al.* narrowed their search for the positive avian factor to just 12 genes. Then, by expressing each of the candidate genes singly in mammalian cells, the authors found what they were looking for: chicken *ANP32A* is a single gene that enables an avian-adapted PB2 protein to function efficiently in mammalian cells.

Confirmation that ANP32A protein supports influenza-polymerase activity was obtained by decreasing the expression of ANP32A in cells. When levels were reduced in chicken cells, the activity of an avian-adapted viral polymerase decreased. Similarly, when expression of the mammalian version of ANP32A was reduced in human cells, a human-adapted viral polymerase was less active. Thus, ANP32A is crucial for influenza A virus replication in both birds and mammals, but avian-adapted polymerases work inefficiently with mammalian ANP32A. These findings indicate that the adaptive changes that influenza viruses acquire in the PB2 627 domain following transmission to mammals allow the viral polymerase to partner with mammalian ANP32A.

The researchers report that chicken and