Late Pleistocene faunal community patterns disrupted by Holocene human impacts

Barry W. Brook[1*], S. Kathleen Lyons[2], Benjamin E. Carter[3,6], William Gearty[2,4], Orlin S. Todorov[5,6], Zach Aandahl & John Alroy[6]

* Corresponding Author. e: barry.brook@utas.edu.au, ph: +61 3 6226 2655

[1] School of Natural Sciences, University of Tasmania, Hobart, TAS 7001, Australia

[2] School of Biological Sciences, University of Nebraska-Lincoln, Lincoln, NE 68588, USA

[3] Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109-1005, USA

[4] Open Source Program Office, Syracuse University, Syracuse, NY 13203, USA

[5] Tasmanian Institute of Agriculture, College of Sciences and Engineering, University of Tasmania, Launceston, TAS 7250, Australia

[6] School of Natural Sciences, Macquarie University, NSW 2109, Australia

Abstract

We analysed fossil mammal assemblages from over 350 Late Pleistocene and Holocene sites worldwide to test whether human activities, such as agriculture, domestication and intensified land use, restructured global patterns of mammal co-occurrence. Using presence-absence data, we contrasted a novel iterative ‘chase clustering’ method, which is compositionally driven, against a traditional spatially constrained Ward’s clustering approach. Both methods recovered continental-scale groupings in the Pleistocene consistent with known biogeographic boundaries. In the Holocene, however, domestication and agricultural expansion disrupted these historical patterns, generating novel clusters unbound by geography and traditional evolutionary lineages. Faunal turnover at the local scale varied substantially across regions, being especially pronounced in the Americas, whereas other areas showed relative stability. Even moderate expansion of domesticates altered how communities grouped, highlighting their disproportionate ecological influence. Our findings demonstrate that human-driven niche modification, beyond earlier megafaunal extinctions, profoundly reshaped mammal communities on a global scale. Recognising these anthropogenic legacies provides essential context for anticipating how current and future human pressures might further transform biodiversity.

Keywords

Biogeography, Mammal communities, Domestication, Agriculture, Faunal turnover, Clustering, Zooarchaeology, Holocene

1. Background

Technological intensification of human activities has shaped terrestrial ecosystems for millennia [1]. As the Late Pleistocene transitioned into the Holocene, people refined hunting strategies and began cultivating crops, ushering in a new era of landscape modification [2, 3]. Domestication soon followed, reorganising ecosystems and reducing biodiversity and ecological complexity, by favouring a handful of livestock and crop species over once-dominant wild taxa [4]. Although much attention has focused on the extinction of large-bodied 'megafauna' in the late Quaternary on multiple continents and islands [5, 6], humans continued to reshape habitats in other ways, often through clearing forests, irrigation, and converting wildlands into farmland or pasture [7].

These ecological transformations did not unfold uniformly in location or magnitude. Some regions lost iconic megafaunal species (e.g., giant sloths in the Americas or large marsupials in Australia), whereas others retained a greater diversity of native herbivores and predators [8]. Climatic warming, occurring simultaneously, further complicated these spatial patterns [9]. Moreover, domesticates spread unevenly, with pastoral and agricultural economies arising at different times across Eurasia, Africa, and the Americas [10]. As a result, faunal communities that once showed distinct geographic patterns could have converged in composition if domestic species spread widely, or they might have become more fragmented in composition if local factors took precedence.

Despite many case studies, the global-scale reorganisation of mammal communities across this transition remains largely underexplored [11]. Previous work has typically focused on single continents or specific taxonomic groups [6, 12, 13], leaving open the question of whether the Holocene, once domesticates entered the scene, ushered in fundamentally new community structures. Standard clustering methods often confuse geographic proximity with ecological similarity, obscuring whether Holocene assemblages maintained historical biogeographic boundaries or represented genuinely new groupings. Recent findings suggest that even small changes in species co-occurrences can ripple across entire ecosystems [14], implying that the replacement of wild fauna with domesticates might leave a large biogeographic footprint. Such findings imply that even limited changes, such as substituting wild fauna with domesticates, could significantly alter ecological relationships and community structure.

Here, we address this gap by examining a diverse suite of Pleistocene and Holocene fossil mammal assemblages from multiple continents and islands. We focus on presence-absence data to avoid the uneven biases of zooarchaeological counts (e.g., due to differential preservation or excavation methods). To uncover patterns, we apply two complementary clustering techniques: a spatially constrained hierarchical method [15] and a new unconstrained 'chase clustering' algorithm. The first approach anchors clusters partly in geographic distance, while the second relies only on compositional similarity, permitting regions to emerge or dissolve in unexpected ways. By comparing Late Pleistocene and Holocene groupings, we test whether domesticates and intensified land use disrupted earlier faunal boundaries.

We hypothesise that intensified land use and domestication amplified human impacts well beyond the megafaunal extinctions of the late Quaternary. We expect to see shifts in how sites group together, particularly in regions where livestock arrived early and spread widely. Departures from historical, pre-domestication faunal boundaries would provide evidence for agriculture's role in reshaping mammal community structure.

2. Material and Methods

(a) Study System and Data Collection

We sourced the zooarchaeological data on mammals assembled for the Ecological Register (http://ecoregister.org/?a=downloadForm), downloaded on 6 June 2023. These come from peer-reviewed sources and museum records, with a broad global coverage and a focus on sites with comprehensive counts of identifiable faunal remains [e.g., 3, 16, 17-19]. We prioritised studies with radiometric or stratigraphic dating to ensure robust chronological control. Papers reporting only minimum number of individuals or vague specimen tallies were excluded, as were poorly dated sites. Despite these efforts, Holocene sites largely represent human-generated assemblages (e.g., middens), potentially underestimating species richness. However, because our primary goal was to investigate biogeographic patterns rather than raw diversity, this limitation does not undermine our core questions. In total, the Pleistocene dataset contained 475 mammal species from 191 sites, whereas the Holocene dataset covered 350 species from 206 sites, including 12 domesticates. Variables for analysis were species occupancy (presence/absence) at each site and geographical coordinates for each assemblage. For further details, see Supplementary Information.

(b) Data Processing

Data from all sources were synthesised into a standardised format for subsequent analysis. For each site, we recorded latitude and longitude, and for this analysis used reduced the inventory data to the presence– absence of each mammal species, after filtering out sites with fewer than five species to avoid unstable compositional signals. Final matrices for Pleistocene and Holocene epochs had rows representing species and columns as sites. We also constructed a coordinate file indicating the precise location of each site in decimal degrees (WGS84 reference system) and computed geographic distances between sites from these coordinates.

(c) Chase Clustering Algorithm

We developed a novel unconstrained 'chase clustering' method specifically to test whether spatial patterns in mammal communities emerge naturally from compositional similarity alone, without imposing geographic structure. The algorithm's logic is illustrated in Fig. 1. Clusters were initiated ('seeded') using sites with distinct species compositions, prioritised according to their overall species richness. A 'chase matrix' was constructed from pairwise comparisons of species presence-absence across sites. In turn, remaining sites were 'chased' into the clusters they most resembled. We performed up to 10,000 random shuffles, each reassigning sites to improve a separation-based fit metric that penalises overly fragmented or trivial clusters. A regularisation parameter (γ) was tuned to balance cluster fragmentation and compositional coherence, by choosing several values and seeing which best matched the continents for the Pleistocene data (see Results). We repeated this for cluster sizes ranging from 2 to 15 (our imposed ceiling), with 250 random starts per size, thereby refining the cluster structure in a manner reminiscent of simulated annealing [20]. This procedure identified robust groupings driven purely by shared species composition.

(d) Constrained Hierarchical Clustering

For a comparison, we also applied a modified version of Ward's hierarchical clustering [21] that integrates both compositional and geographic information [15]. Unlike chase clustering, the hierarchical approach requires specifying the number of clusters beforehand. This method computes two dissimilarity matrices: one based on species occupancy (Euclidean distance), and a second derived from great-circle distances among sites. A mixing parameter, alpha (α), then weights these matrices in a convex combination. Low α values stress composition; higher values stress geographic proximity. The approach iteratively merges sites to minimise within-cluster variance across both matrices. This was implemented using `hclustgeo` (`ClustGeo` package) with number of clusters set to the chase-clustering result. To select α, we ran increments of 0.01 between 0-1 and chose the value maximising variance explanation (inertia). The resulting dendrogram was partitioned into clusters matching the number identified via chase clustering, to compare global patterns between the two epochs.

(e) Composition of wild versus domesticate species

For our main Holocene comparison, we excluded all domesticated taxa to focus on shifts within wild mammalian communities. We identified domesticates using established zooarchaeological criteria [18], removing them from the Holocene dataset to create a matrix directly comparable to the Pleistocene. This dataset is referred to as Holocene Wild. We then did a separate analysis using the full matrix (including domesticates, termed Holocene All) to measure how much these newly introduced species altered cluster structure, by comparing site assignments using the adjusted Rand index (ARI) [22]. This highlighted the extent to which domesticates reshaped Holocene biogeographic patterns beyond the changes observed in wild fauna.

(f) Turnover from Pleistocene to Holocene

For 34 locations with paired Late Pleistocene and all Holocene records, we measured turnover directly at approximately the same locality (i.e., grouped sites within a 50 km radius). Turnover was quantified using the Jaccard distance, calculated as species gained or lost between epochs, divided by the total species observed across both periods [23]. We then mapped these turnover values to visualise which regions exhibited the greatest compositional shifts over time.

(g) Implementation and Reproducibility

All analyses were conducted using R v4.4.2 [24], employing packages `dplyr` for data manipulation, `ggplot2` for visualisation, and `ClustGeo` for constrained clustering. Custom R functions were written for chase clustering.

The R scripts for data import, filtering, clustering, and visualisation are provided on GitHub (see Data Accessibility), ensuring reproducibility. A detailed description of chase clustering, and a comparison with Ward's geographically constrained hierarchical approach, is given in the Supplementary Material.

3. Results

We tested the regularisation parameter (γ) ranging from 0 to -0.20 to fine-tune the chase clustering for the Pleistocene data. At $\gamma = 0$, only two clusters emerged (Australia versus the rest of the world), whereas $\gamma = -0.05$ yielded four broad continental groupings. Further decreases in γ continued isolating unique regional sites (e.g., southern South America, New Guinea). At $\gamma = -0.10$, six coherent clusters captured major continental differences while avoiding excessive fragmentation. Beyond $\gamma = -0.15$, solutions became highly subdivided (up to 12 clusters). We therefore selected $\gamma = -0.10$ as it produced a stable partition with meaningful biogeographic structure for the Pleistocene, and then also used this value for the Holocene analysis.

Comparisons between Ward’s hierarchical method and chase clustering reveal differing global patterns across the Pleistocene, Holocene Wild, and Holocene All datasets (Fig. 2). Ward clustering consistently distinguishes the Americas, whereas Africa remains grouped closely with Eurasia. Australia and Southeast Asia separate from the Palaearctic in the Pleistocene, but they unify more in the Holocene, and Europe exhibits greater internal partitioning during the Pleistocene. Notably, the Ward clusters do not change when domesticates are included (Holocene Wild vs. All). By contrast, chase clustering yields partially overlapping but more flexible partitions. In the Pleistocene, the Americas again stand apart, while Africa, Europe, and Asia form a broad ensemble, with distinctive smaller clusters highlighting unique regional faunas in India, New Guinea, and Australia. In the Holocene Wild dataset, Africa detaches from Eurasia, and

North-South American sites show more similarity. Inclusion of domesticates (Holocene All) further reshapes clusters, uniting some previously separate regions (e.g., Africa), but also creating highly localised clusters (e.g., in Brazil and Java), highlighting how domestication homogenises some regions yet fragments others. The proportion of domesticated species relative to the site's total richness is shown in Figure S1.

Of the 350 Holocene species in our dataset, 12 were domesticated mammals, appearing in roughly half of the global sites (110 of 206, with a mean of 1.47 domesticates per site). Comparing chase clustering solutions generated from only wild species versus those generated from all species (i.e., including domesticates) yielded an ARI = 0.433. While this indicates moderate concordance, it clearly shows that domesticates significantly restructured Holocene community patterns.

Across the 34 aggregated (multi-site) localities common to both epochs, turnover from Pleistocene to Holocene varied considerably (Fig. 3). Regions like New Guinea and Sri Lanka experienced minimal turnover, whereas western Europe and eastern Africa exhibited moderate levels of turnover. In contrast, the Americas, Australia, eastern Europe, and parts of Asia, including Ethiopia and southern Africa, experienced the highest turnover. This geographic heterogeneity suggests that anthropogenic (e.g., farming, intensified land use) and environmental drivers (post-glacial climate change) affected mammal communities unevenly across the globe.

4. Discussion

Our modelling revealed that Holocene mammal communities diverged in several important ways from those of the Late Pleistocene, underscoring the lasting impacts of intensified land use and domestication associated with farming and other human activities. Clusters in the Pleistocene data often aligned with broad continental regions, consistent with climatic gradients and physical barriers such as mountain ranges and oceans [7, 8]. In contrast, Holocene groupings showed fluid boundaries under unconstrained chase clustering, indicating that domestic animals and other human impacts disrupted faunal communities beyond climate-driven expectations.

Including domesticated species strongly influenced chase cluster assignments. Domesticates spread widely, creating new compositional similarities among geographically distant regions—for instance, mammal communities in Europe and Africa converged due to the adoption of domesticated species originating in the Middle East [4, 10]. When we removed domesticates from Holocene analyses, some regions regained a resemblance to Pleistocene patterns, implying that a subset of introduced taxa had an outsized role in determining site similarity. This finding aligns with Tóth et al. [14], who showed that even moderate changes in species interactions can reshape macro-scale co-occurrence patterns. Domesticates spread across multiple continents, creating new overlaps among sites separated by large distances. This effect would remain hidden in methods that assume strictly contiguous ranges or that place strong emphasis on geographical adjacency. Spatially constrained Ward's clustering misses these effects, illustrating how anthropogenic factors can transcend traditional biogeographic boundaries, as well as underscoring the utility of the new chase algorithm to detect non-spatially contiguous similarities.

At the same time, these transformations varied by region. Turnover analyses identified stable regions (e.g., Oceania) and others undergoing significant shifts (e.g., the Americas, Eurasia). Differences in the timing and intensity of agriculture may explain some of this heterogeneity, but other factors, like pre-existing environmental conditions or local cultural practices, also likely mattered [11, 12]. In the Americas, substantial Late-Pleistocene megafaunal losses were followed by regional domestication of native mammals, notably llamas and alpacas in South America, reshaping faunal communities long before European livestock introductions [13, 25]. Elsewhere, indigenous mammalian domesticates (e.g., cattle, sheep, pigs) spread unevenly, producing patchier community compositions.

Despite strong domestication signals, our analysis has inherent limitations. We used site-occupancy data, which can mask subtle shifts in species abundance. Zooarchaeological records can be biased by human food preferences or differential bone preservation. However, the clear clustering of Holocene African faunas, even when domestic species are excluded, suggests the observed pattern is not merely an artifact of midden use or preservation bias [8]. Nevertheless, these data are among the most comprehensive currently available, and our analysis focuses on large-scale compositional shifts rather than fine-grained population dynamics. Future studies using abundance data, environmental proxies, or additional taxa (e.g., birds, plants) could further refine these patterns. More complete chronological information would also help disentangle the relative contributions of climate, hunting, and agriculture.

Broadly, these findings underscore how human-driven extinctions, agriculture, and resource extraction profoundly reshaped mammal community structures [2, 12]. Modern biogeographic patterns have deep historical roots shaped by ancestral choices about species cultivation, trade, and conservation [3]. Such perspective may guide current conservation efforts by reminding us that ecosystems are neither fixed nor solely climate-driven but reflect ongoing interactions between people and the rest of the biosphere. How we manage these interactions today will determine whether mammal communities become resilient or increasingly destabilised.

Ethics. Ethical approval was not required.

Data accessibility. All data and R code required for full replication are provided at https://github.com/bwbrook/chase-clustering, archived and citable via Zenodo (DOI: 10.5281/zenodo.15054872).

Authors' contributions. All authors co-designed the study. B.E.C. and J.A. collected the site-assemblage data. B.W.B. wrote the R code, did the analyses and prepared the display items. J.A. devised the chase-clustering algorithm; B.W.B. implemented the R function. B.W.B. wrote the paper and all authors provided input into revisions. All authors agree to be held accountable for the content.

Competing interests. None to declare.

Funding. Research and workshop support was provided to J.A., B.W.B. and S.K.L. from the Australian Research Council (grant DP210101324).

References

[1] Smil, V. 2011 Harvesting the biosphere: the human impact. *Population and Development Review* **37**, 613-636. (doi:10.1111/j.1728-4457.2011.00450.x).

[2] Ellis, E.C., Gauthier, N., Goldewijk, K.K., Bird, R.B., Boivin, N., Diaz, S., Fuller, D.Q., Gill, J.L., Kaplan, J.O., Kingston, N., et al. 2021 People have shaped most of terrestrial nature for at least 12,000 years. *Proceedings of the National Academy of Sciences USA* **118**. (doi:10.1073/pnas.2023483118).

[3] Stephens, L. & Fuller, D. & Boivin, N. & Rick, T. & Gauthier, N. & Kay, A. & Marwick, B. & Armstrong, C.G.D. & Barton, C.M. & Denham, T., et al. 2019 Archaeological assessment reveals Earth's early transformation through land use. *Science* **365**, 897-902. (doi:10.1126/science.aax1192).

[4] Ahmad, H.I., Ahmad, M.J., Jabbir, F., Ahmar, S., Ahmad, N., Elokil, A.A. & Chen, J.P. 2020 The domestication makeup: evolution, survival, and challenges. *Frontiers in Ecology and Evolution* **8**. (doi:10.3389/fevo.2020.00103).

[5] Cooke, R., Gearty, W., Chapman, A.S.A., Dunic, J., Edgar, G.J., Lefcheck, J.S., Rilov, G., McClain, C.R., Stuart-Smith, R.D., Lyons, S.K., et al. 2022 Anthropogenic disruptions to longstanding patterns of trophic-size structure in vertebrates. *Nature Ecology & Evolution* **6**, 684-+. (doi:10.1038/s41559-022-01726-x).

[6] Smith, F.A., Smith, E.A.E., Villaseñor, A., Tomé, C.P., Lyons, S.K. & Newsome, S.D. 2022 Late Pleistocene megafauna extinction leads to missing pieces of ecological space in a North American mammal community. *Proceedings of the National Academy of Sciences USA* **119**. (doi:10.1073/pnas.2115015119).

[7] Boivin, N.L., Zeder, M.A., Fuller, D.Q., Crowther, A., Larson, G., Erlandson, J.M., Denham, T. & Petraglia, M.D. 2016 Ecological consequences of human niche construction: Examining long-term anthropogenic shaping of global species distributions. *Proceedings of the National Academy of Sciences USA* **113**, 6388-6396. (doi:10.1073/pnas.1525200113).

[8] Turvey, S.T. & Fritz, S.A. 2011 The ghosts of mammals past: biological and geographical patterns of global mammalian extinction across the Holocene. *Philosophical Transactions of the Royal Society B - Biological Sciences* **366**, 2564-2576. (doi:10.1098/rstb.2011.0020).

[9] Lyons, S.K., Miller, J.H., Fraser, D., Smith, F.A., Boyer, A., Lindsey, E. & Mychajliw, A.M. 2016 The changing role of mammal life histories in Late Quaternary extinction vulnerability on continents and islands. *Biology Letters* **12**, 20160342. (doi:10.1098/rsbl.2016.0342).

[10] Larson, G. & Fuller, D.Q. 2014 The evolution of animal domestication. In *Annual Review of Ecology, Evolution, and Systematics* (ed. D.J. Futuyma), pp. 115-136.

[11] Rueda, M., Gonzalez-Suárez, M. & Revilla, E. 2024 Global biogeographical regions reveal a signal of past human impacts. *Ecography* **2024**. (doi:10.1111/ecog.06762).

[12] Hatfield, J.H., Davis, K.E. & Thomas, C.D. 2022 Lost, gained, and regained functional and phylogenetic diversity of European mammals since 8000 years ago. *Global Change Biology* **28**, 5283-5293. (doi:10.1111/gcb.16316).

[13] Lyons, S.K., Amatangelo, K.L., Behrenstneyer, A.K., Bercovici, A., Blois, J.L., Davis, M., DiMichele, W.A., Du, A., Eronen, J.T., Faith, J.T., et al. 2016 Holocene shifts in the assembly of plant and animal communities implicate human impacts. *Nature* **529**, 80-U183. (doi:10.1038/nature16447).

[14] Toth, A.B., Lyons, S.K., Barr, W.A., Behrensmeyer, A.K., Blois, J.L., Bobe, R., Davis, M., Du, A., Eronen, J.T., Faith, J.T., et al. 2019 Reorganization of surviving mammal communities after the end-Pleistocene megafaunal extinction. *Science* **365**, 1305-1308. (doi:10.1126/science.aaw1605).

[15] Chavent, M., Kuentz-Simonet, V., Labenne, A. & Saracco, J. 2018 ClustGeo: an R package for hierarchical clustering with spatial constraints. *Computational Statistics* **33**, 1799-1822. (doi:10.1007/s00180-018-0791-1).

[16] Ament, J.M., Carbone, C., Crees, J.J., Freeman, R. & Turvey, S.T. 2023 Anthropogenic predictors of varying Holocene occurrence for Europe's large mammal fauna. *Biology Letters* **19**. (doi:10.1098/rsbl.2022.0578).

[17] Davoli, M., Kuemmerle, T., Monsarrat, S., Crees, J., Cristiano, A., Pacifici, M. & Svenning, J.C. 2024 Recent sociocultural changes reverse the long-term trend of declining habitat availability for large wild mammals in Europe. *Diversity and Distributions* **30**. (doi:10.1111/ddi.13921).

[18] Larson, G., Piperno, D.R., Allaby, R.G., Purugganan, M.D., Andersson, L., Arroyo-Kalin, M., Barton, L., Vigueira, C.C., Denham, T., Dobney, K., et al. 2014 Current perspectives and the future of domestication studies. *Proceedings of the National Academy of Sciences USA* **111**, 6139-6146. (doi:10.1073/pnas.1323964111).

[19] Carter, B.E. & Alroy, J. 2024 Energy use of modern terrestrial large mammal communities mirrors Late Pleistocene megafaunal extinctions. *Frontiers of Biogeography* **16**, e62724. (doi:10.21425/F5FBG62724).

[20] Kirkpatrick, S., Gelatt, C.D.J. & Vecchi, M.P. 1983 Optimization by simulated annealing. *Science* **220**, 671-680. (doi:10.1126/science.220.4598.671).

[21] Murtagh, F. & Legendre, P. 2014 Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *Journal of Classification* **31**, 274-295. (doi:10.1007/s00357-014-9161-z).

[22] Hubert, L. & Arabie, P. 1985 Comparing partitions. *Journal of Classification* **2**, 193-218. (doi:10.1007/BF01908075).

[23] Jaccard, P. 1901 Étude comparative de la distribution florale dans une portion des Alpes et du Jura. *Bulletin de la Société Vaudoise des Sciences Naturelles* **37**, 547-579.

[24] R Core Team. 2025 R: A language and environment for statistical computing. Available at: https://www.R-project.org

[25] Hedberg, C.P., Lyons, S.K. & Smith, F.A. 2022 The hidden legacy of megafaunal extinction: Loss of functional diversity and resilience over the Late Quaternary at Hall's Cave. *Global Ecology and Biogeography* **31**, 294-307. (doi:10.1111/geb.13428).

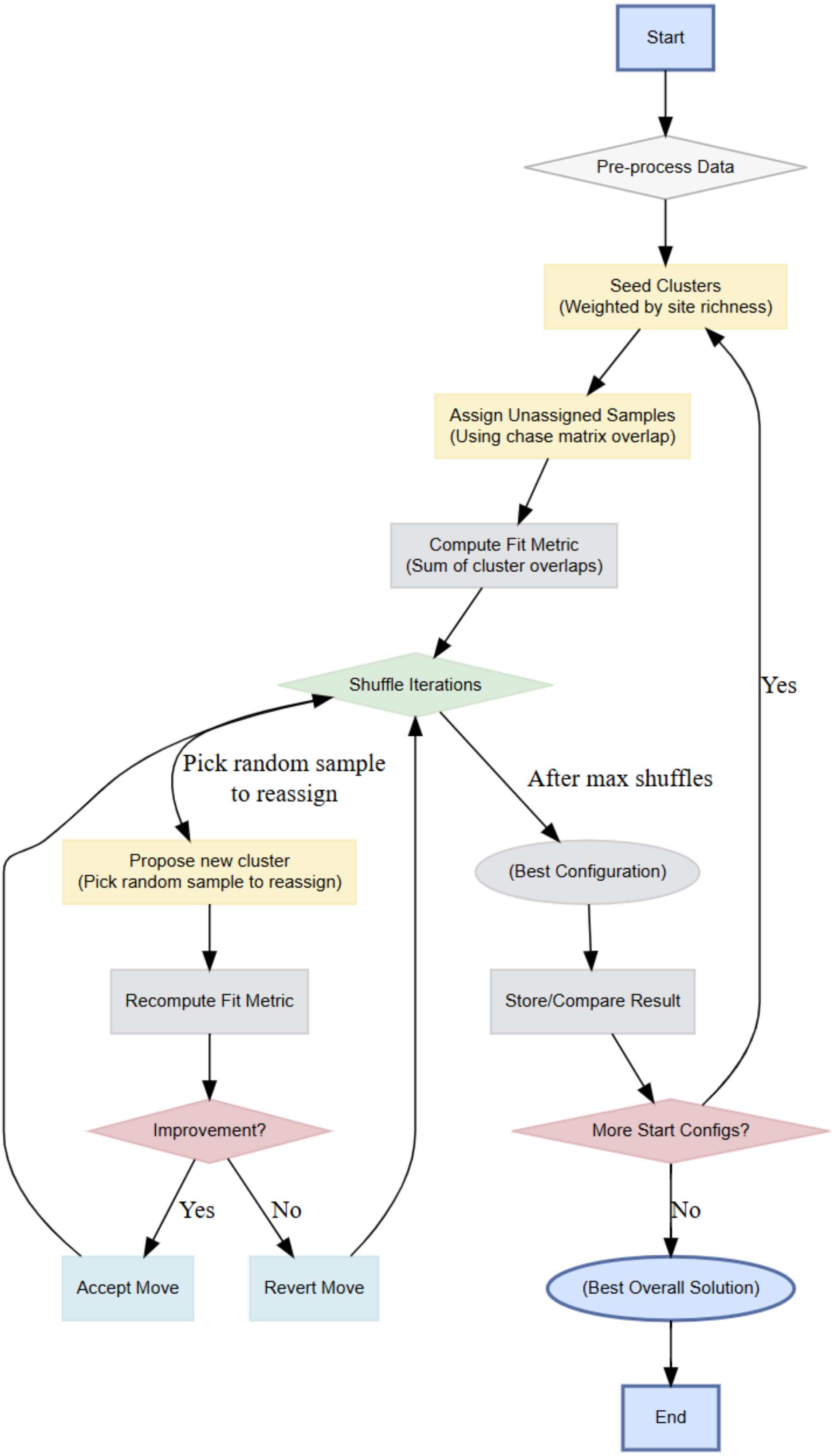
Figure Legends

Figure 1. Schematic of the ‘chase clustering’ algorithm, which uses compositional overlap in ecological communities to identify and refine cluster assignments. First, sites with insufficient data are filtered out, then clusters are seeded from weighted samples. Unassigned sites join their most similar cluster. The algorithm shuffles assignments iteratively to maximise compositional coherence, retaining the best solution after multiple starts.

Figure 2. Global cluster maps comparing Ward’s hierarchical method (left columns) and chase clustering (right) for three data subsets: Late Pleistocene (top row), Holocene Wild species only (middle), and Holocene all species (bottom); colours represent distinct clusters of sites but are otherwise arbitrary. Ward’s clusters remain relatively stable due to geographic constraints, whereas chase clustering partitions vary more flexibly with compositional shifts, especially after domesticates are introduced.

Figure 3. Regional-level faunal turnover from the Late Pleistocene to the Holocene at 34 multi-site locations. Domesticated species are included in Holocene locations. Circles represent proportional species turnover (losses or gains); warmer colours (red) indicate higher turnover, cooler colours (blue/purple) are lower turnover. The variability highlights regionally distinct responses to human and environmental pressures.

385 Figure 1



386

Figure 2

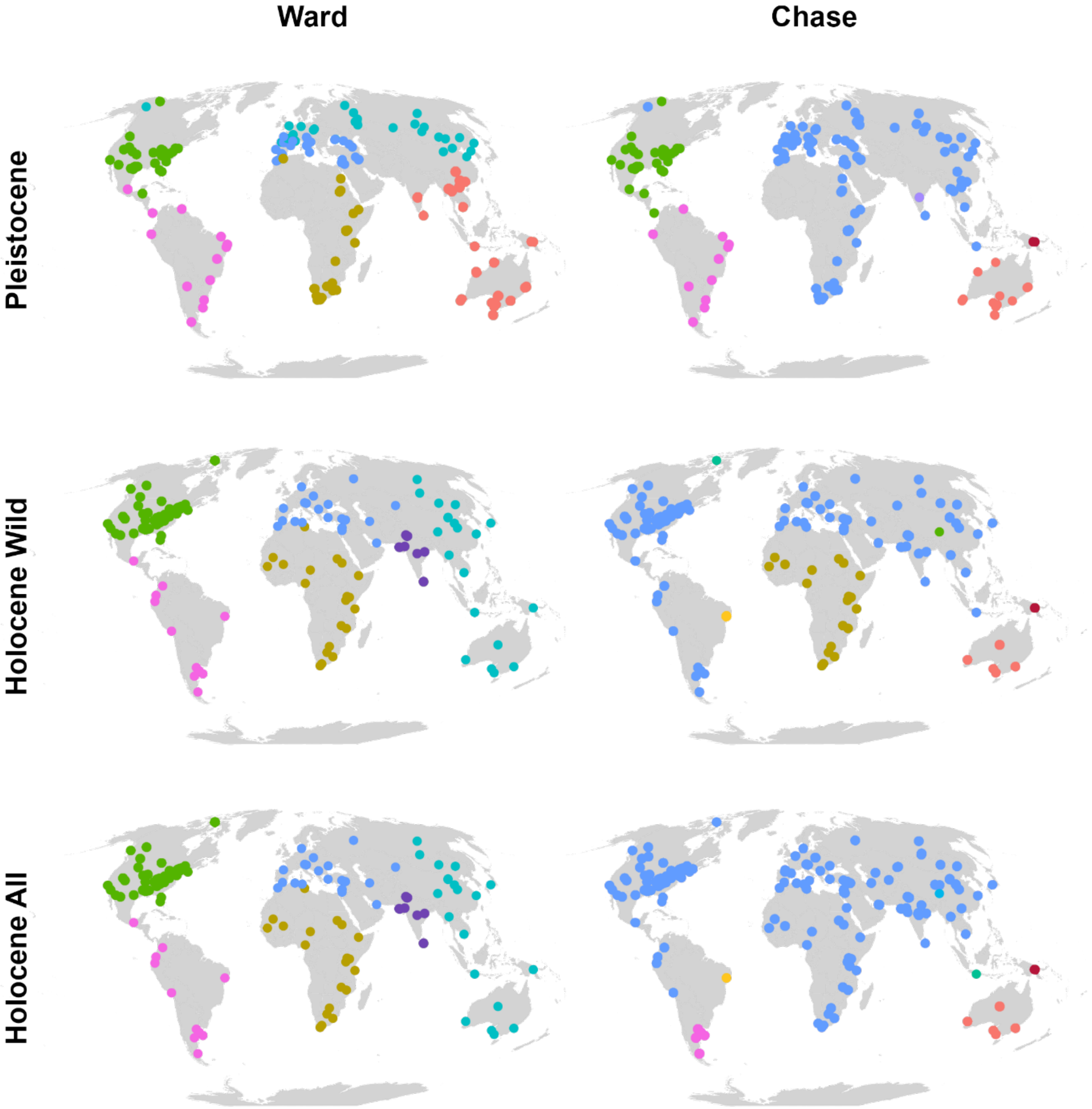
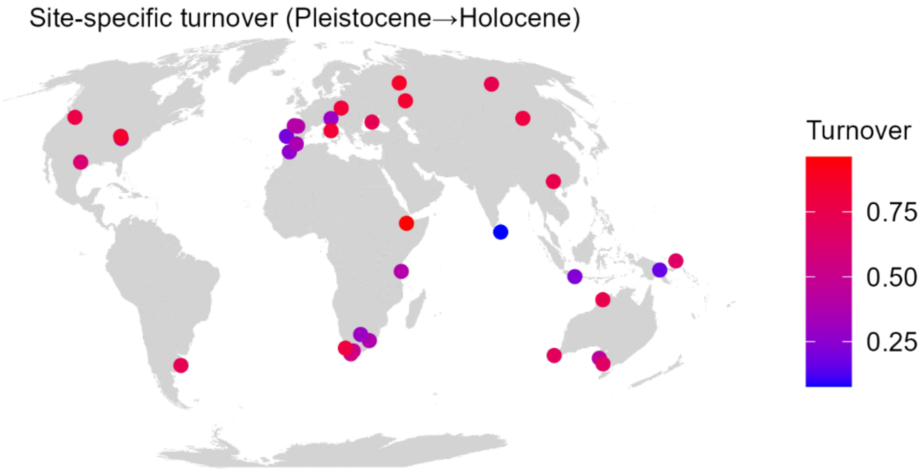
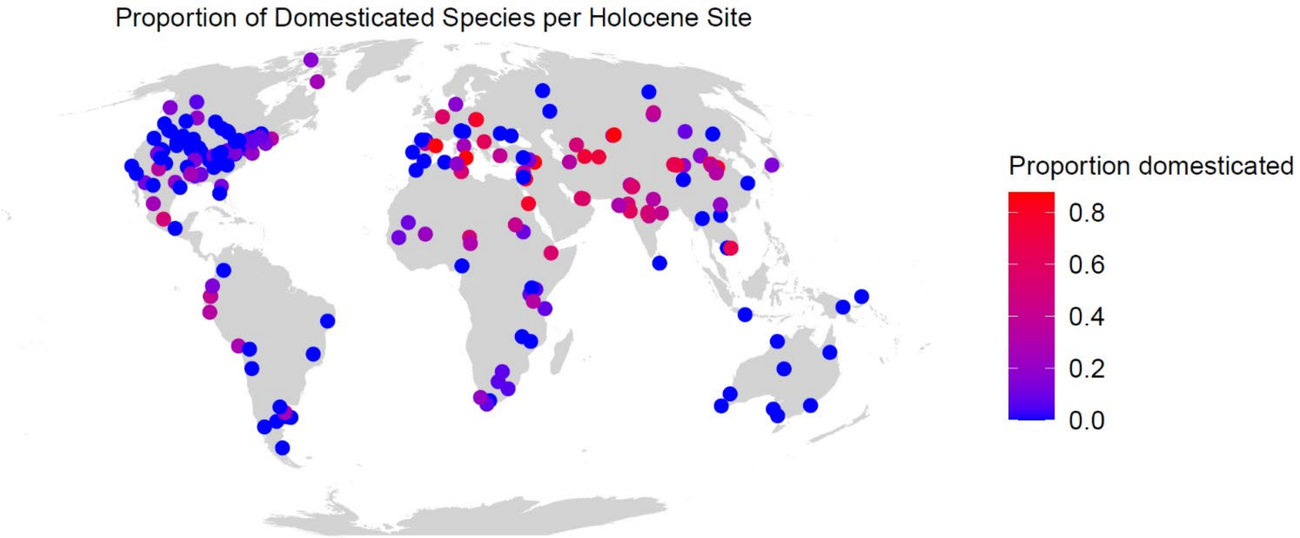


Figure 3


Site-specific turnover (Pleistocene→Holocene)
Turnover
0.75
0.50
0.25

SUPPLEMENTARY MATERIAL

Figure S1. Geographic distribution of domesticated species in Holocene faunal communities. Points indicate Holocene fossil sites, coloured according to the proportion of domesticated species present relative to the site's total recorded mammalian species richness. Higher proportions indicate stronger influence from domesticated taxa. The Mollweide equal-area projection is used.



Supplementary Methods

Chase Clustering vs. Constrained Ward's Hierarchical Approach

Motivation and Conceptual Underpinnings

In the main paper, we introduced a custom 'chase clustering' algorithm as an alternative to a spatially constrained hierarchical approach. The need for an alternative arises from the question of whether strong geographic structure in fossil mammal communities is purely an artifact of spatial constraints or whether it can emerge solely from compositional patterns, i.e., shared occurrences of species. Many clustering frameworks, such as Ward's method with spatial weighting, explicitly incorporate geography, potentially obscuring the extent to which spatial signals arise naturally from the species data itself. The chase clustering algorithm directly addresses this question by discarding any spatial constraint: it relies solely on compositional overlap among sites yet still seeks well-separated clusters.

Our impetus was twofold: (1) to see if purely compositional clustering recapitulates known spatial groupings and (2) to provide a flexible iterative algorithm that can incorporate additional constraints or penalise degenerate solutions.

Ward's constrained method (as implemented in the `ClustGeo` R library) is a relatively standard approach for analysing spatially explicit data. It merges sites hierarchically, minimising within-cluster variance in a mixture of compositional and geographical distances. Adjusting the mixing parameter α allows the degree of spatial constraint to vary: high α emphasises geographic proximity, whereas α near zero prioritises compositional similarity. This approach is appealing for datasets in which spatial processes (e.g., dispersal limitation) structure communities. However, it presupposes that geography matters. If we want to test the emergent role of geography, an unconstrained method provides a natural baseline for comparison. We developed chase clustering precisely to test whether purely compositional data can reveal meaningful geographic structure without imposed spatial constraints.

Chase Clustering Algorithm: Detailed Rationale

Our chase clustering algorithm assigns sites to clusters based on compositional similarity alone. It works by 'pursuing' compositional similarity: each site is assigned to the cluster whose existing members share the greatest overlap in species. The novelty lies in how we track compositional overlap (via cross-products of site occupancy vectors) and in how the algorithm randomly shuffles assignments to avoid local optima. Conceptually, it borrows from simulated annealing and genetic algorithms in its repeated, stochastic improvement steps. The workflow is as follows:

1. *Data Filtering*. We first exclude sites that fail to meet a minimum species threshold. Because sites with only a few species can be very noisy (and trivially dissimilar from most other sites), filtering them reduces artifacts and computational overhead.

2. *Seeding Clusters.* Clusters are seeded by randomly selecting K distinct sites, weighted by species richness, ensuring initial clusters are compositionally distinct (i.e., sites with many species are more likely to be initial seeds). This step attempts to ensure that each cluster "seed" is compositionally distinct and not empty.

3. *Initial Assignment.* Remaining sites are initially assigned to clusters based on maximal compositional overlap (computed via the 'chase matrix'). If multiple clusters have equal similarity, assignment is randomly chosen among them.

4. *Shuffling and Improvement*. Sites are iteratively reassigned at random, with moves retained only if they improve (reduce) the total compositional overlap metric. This stochastic refinement continues for numerous iterations (e.g., 10,000) to identify optimal partitions.

In the version presented, each cluster is internally normalised by its total occupancy, and the pairwise overlaps between clusters are summed. High overlap indicates less separation; low overlap indicates better separation. The algorithm minimises the total overlap (smaller = better). To prevent degenerate solutions (e.g., clusters with zero occupancy), the algorithm imposes large penalties, ensuring clusters remain compositionally meaningful. Users can apply an additional penalty to regulate how many clusters form (e.g., penalising/rewarding K to avoid over-/under-splitting).

Comparison with Ward's Hierarchical Method

To clarify how chase clustering differs from Ward's constrained approach, we highlight the main distinctions:

Feature	**Chase Clustering (new algorithm)**	**Constrained Ward's (Clustgeo)**
Data Inputs	Species occupancy matrix only; site coordinates optional for *post hoc* mapping.	Two distance matrices: compositional (D0) and spatial (D1).
Spatial Constraints	None; purely compositional clusters. Geographic coordinates optional for mapping clusters post hoc.	Weighted by α ∈ [0,1]. Larger α → clusters formed primarily by minimizing spatial distance; smaller α → composition is emphasised.
Algorithm Structure	Iterative, stochastic reassignment ('shuffle')	Deterministic, bottom-up hierarchical merges. No random restarts; merging

	algorithm that explores multiple starting points.	steps determined by Ward criterion on combined distance.
Number of Clusters	Explored across a range (2 to 15). We track the best solution at each size (akin to k-means–style iterative search).	Usually chosen *a priori*. The user ‘cuts’ the dendrogram at a specified level. The same K can be tested under multiple α values.
Objective	Minimises overall cluster overlap, or a penalized metric if the user sets alpha > 0 for more clusters.	Minimises total within-cluster variance in a convex combination of compositional + spatial distances.
Computational Effort	Potentially high for large data, but crossprod speeds it up significantly; random restarts can be parallelised.	Relatively fast for moderate data sizes. Scaling each distance matrix to [0,1] is standard.
Output	A single ‘best’ assignment of sites to clusters for each K, plus a universal best across all K in the tested range.	A tree (dendrogram) with an assigned partition. Variation in α changes how much the tree respects geography vs. species composition.
Interpretation	Identifies purely compositional clusters, providing a baseline to assess if geography emerges naturally from species data.	Embeds an explicit geographic constraint, suitable if you already believe adjacency or distance should shape cluster membership.

When to Prefer Which?

Chase clustering is beneficial if you suspect compositional structure aligns with geography **BUT** seek evidence that this occurs spontaneously. It also allows for flexible modifications (e.g., ignoring or penalising zero-overlap clusters). However, it can be slower on very large datasets and lacks the neat hierarchical tree representation. Constrained Ward’s is suitable when spatial proximity is assumed critical, offering deterministic hierarchical structure and easily interpretable results. It yields a dendrogram and has a straightforward parameter α controlling how much weight is placed on geography versus composition. That said, it presupposes some role of space and does not allow the same freedom to ‘shuffle’ site memberships iteratively.

Practical Implications and Trade-Offs

Both methods can yield meaningful ecological insights, especially if used in tandem:

Chase: Strong compositional gradients appearing spontaneously suggest geography strongly shapes community composition, whereas weak spatial signals imply other drivers (land use, history).

Ward's: Clearly identifies geographically contiguous clusters and effectively explores compositional-spatial trade-offs by varying α.

A key practical difference is the ease of deciding on the number of clusters. Chase clustering tests a user-defined range (e.g., 2 to 15) and picks the best solution at each step (subject to a tuneable complexity penalisation), while Ward's approach typically demands a prior choice of K (e.g., K=7). One can combine the two as we did in the main paper: first let chase clustering suggest an optimal K purely from compositional data, then fix that K in Ward's approach to evaluate how increasing α (the spatial constraint) shifts the partition.

Algorithmic Details in Our Implementation

In our workflow, chase clustering is implemented with the following steps in the R scripts (see GitHub repository: bwbrook/chase-clustering):

1. *Matrix Preprocessing*: Columns (sites) with fewer than five species and empty species rows are removed prior to clustering.

2. *Chase Matrix*: Use `` `crossprod(dataMatrix)` `` to compute compositional overlap, creating a symmetric matrix of species co-occurrence sums between sites.

3. *Shuffling*: Loop over a large number (e.g., 10,000) of random reassignments, each time accepting the change only if it lowers the overlap-based metric.

4. *Penalties*: If overlap equals exactly zero (a degenerate cluster), we assign a large penalty (`1e6`) to disqualify such partitions.

5. *Range of Cluster Sizes*: We iterate from 2 up to a user-specified maximum number of clusters, storing the best solution for each. The 'universal best' is the solution that yields the lowest overall overlap among all tested cluster sizes.

For Ward's clustering, we rely on the `ClustGeo` package to compute a compositional distance matrix `` `D0` `` and a geographic distance matrix `` `D1` ``, then vary α to determine which combined distance best explains overall inertia. Finally, we fix a number of clusters (K) to make an apples-to-apples comparison with chase clustering results at that same K.

Conclusions and Recommendations

Chase clustering serves as an informative baseline, revealing if spatial structure emerges purely from species composition. Ward’s hierarchical clustering explicitly tests spatial hypotheses. In our study, comparing these methods reveals which aspects of faunal clustering might be driven by external geographic forcing versus internal compositional structure. We recommend combining both approaches in palaeoecological analyses: chase clustering to identify natural compositional structure, and spatially constrained Ward’s to test the significance of geography explicitly.

Code to reproduce both methods is available in the project's GitHub repository, along with the datasets used in this paper, allowing for complete replication.