

Research



Cite this article: Toivonen J, Fortelius M, Žliobaitė I. 2022 Do species factories exist? Detecting exceptional patterns of evolution in the mammalian fossil record. *Proc. R. Soc. B* **289**: 20212294.
<https://doi.org/10.1098/rspb.2021.2294>

Received: 20 October 2021
 Accepted: 21 February 2022

Subject Category:

Palaeobiology

Subject Areas:

palaeontology, evolution

Keywords:

macroevolution, net primary productivity, mammals, fossil record, NOW database, logistic regression

Author for correspondence:

Indrė Žliobaitė
 e-mail: indre.zliobaite@helsinki.fi

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.5882156>.

Do species factories exist? Detecting exceptional patterns of evolution in the mammalian fossil record

Jaakko Toivonen¹, Mikael Fortelius^{2,3} and Indrė Žliobaitė^{1,3}

¹Department of Computer Science, and ²Department of Geosciences and Geography, University of Helsinki, Helsinki, Finland

³The Finnish Museum of Natural History, Helsinki, Finland

JT, 0000-0002-5444-8877; MF, 0000-0002-4851-783X; IŽ, 0000-0003-2427-5407

A species factory refers to the source that gives rise to an exceptionally large number of species. However, what is it exactly: a place, a time or a combination of places, times and environmental conditions, remains unclear. Here we search for species factories computationally, for which we develop statistical approaches to detect origination, extinction and sorting hotspots in space and time in the fossil record. Using data on European Late Cenozoic mammals, we analyse where, how and how often species factories occur, and how they potentially relate to the dynamics of environmental conditions. We find that in the Early Miocene origination hotspots tend to be located in areas with relatively low estimated net primary productivity. Our pilot study shows that species first occurring in origination hotspots tend to have a longer average longevity and a larger geographical range than other species, thus emphasizing the evolutionary importance of the species factories.

1. Introduction

While it is clear that rates of evolution have varied in the past [1], in which way they vary remains at the centre of palaeobiology research. Some places or times tend to give rise to more new species than others. To explain this Vermeij & Dietl [2] proposed the majority rule, which posits that the source of new or recovering populations is to be found primarily in productive and large environments. While not tested with data, intuitively, their hypothesis makes sense. Leaving aside complications due to dependence on carrying capacity [3] or diversity [4], environments that cover larger areas should accommodate more species [5] and thus, under fixed rates of evolution, could potentially produce more new species. The more new species are produced, the higher the chances are for them to spread and become common. How else could it be?

The concept of the species factory [6] describes quite the opposite phenomenon, where rather than originating from productive environments, many successful species originate from harsh or marginal environments [7,8]. Three examples of species factories in the mammalian record, the latest Middle Miocene origins of the Pikermian palaeobiome in the Sub-Parathethyan province [7], the Pliocene origin of Pleistocene megafauna in the Tibetan Plateau [9] and the rise of arid-adapted fauna in the Plio-Pleistocene of the Eastern African Turkana Basin [8], all point to trends of environmental change. The key condition is that marginal environments that are initially rare later become widespread, following persistent directional climate change, such as cooling or drying.

Here we ask whether some places, times, or combinations of places, times and types of environment give rise to abnormal numbers of new species, and if so, can we define and detect such hotspots in the mammalian fossil record computationally? We further ask whether species factories are just about luck, getting started ahead in environments that will later become widespread,

or whether species factories could arise for other reasons. To answer this we search for exceptional patterns in the fossil record and analyse their contexts.

Even though the fossil record will remain incomplete [10,11], as more fossil data are accumulated and analysed, a more comprehensive picture emerges. In recent decades, fossil data from many sources have been put together to form large-scale fossil databases [12]. The availability of such vast resources has accelerated the development and usage of computational methods to analyse them, for instance, for tracking faunal communities and their dynamics [13–19], estimating biochronology of fossil localities [20–22], or reconstruction of their paleoenvironments [23–28].

Various techniques have been developed to make inferences on diversification rates that take the incomplete and biased sampling of the fossil record into account, such as sub-sampling [29,30] or capture–mark–recapture methods [31–36]. While most of the studies have concentrated on how origination and extinction rates have changed over time, spatial aspects are gaining attention as well, especially with availability of dedicated analysis toolboxes [37].

For analysing species factories our task is to detect anomalies in macroevolutionary patterns over space and time. We statistically detect exceptional origination, extinction and sorting patterns accounting for spatial resolution of the fossil record. By sorting we mean appearance of species that originated elsewhere and local extinctions. While many analyses of the mammalian record operate at the genus level [29,30,32,38], and there are good arguments for such choice [39], here we deliberately analyse the mammalian fossil record at the species level in order to reason about *species* factory phenomena. Our computational approach is thus tuned to work with species-level data that often lack precise taxonomic identification.

Our data analysis focuses on the Late Cenozoic record of mammalian species occurrences in Europe, reported in the NOW database of fossil mammals [40]. We flag origination, extinction and sorting hotspots in this region and analyse them in the context of palaeoclimatic conditions. For reconstructing palaeoclimatic conditions, we use existing dental ecometric models [27] that rely on composition of plant eating mammal communities and characteristics of their teeth.

In order to detect anomalies, we need baselines to quantify what patterns are normally expected, and we need to do it as locally as possible in time and space. For this purpose, we train a logistic regression model that gives us an expected number of occurrences of new species in a given place and time depending on sampling intensities in the current and the previous time unit. Then, we ask statistically, what is the probability to find this many or more occurrences of new species at any given locality in that time interval by chance. If that probability is low, we take that as a potential indicator of an exceptionally high origination rate at this place and time, and flag such a locality as a potential species factory. This approach extends naturally to flagging exceptional patterns of extinction and species sorting in space and time, which we analyse in the context of palaeoenvironmental reconstructions.

2. Study region and fossil data

We retrieved a public version of fossil data from the NOW database [40] on 26 April 2020. The data contained information on species occurrences at localities, stratigraphic context of

localities, their age estimates, taxonomic affiliation of species and their ecological characteristics.

We delimited our analysis to Europe, similar to previous studies [41], as within -25° to 40° of longitude and latitude above 35° . However, we referred to the global data to determine globally first and non-first occurrences of species, as well as times of their global extinction.

We selected the localities falling within the age range of European Land Mammal ages (known as MN units) [42,43], covering age range from 23 Ma to 1.9 Ma [44] (MN1–MN17), we also included two informal MQ units that covered the time from 1.9 Ma to 0.01 Ma (MQ18–MQ19). All time bounds that we used are listed in the electronic supplementary material. We used MN and MQ units as time bins for the analysis. The analysis of species factories was on bins MN2–MQ18, while the first and the last bin was only used for determining the first and the last observed occurrences.

The selected snapshot of data contained 48 318 rows of data for the global dataset, that is 48 318 occurrences of 9692 unique species over 5360 localities. The spatially restricted dataset (only localities with longitude between -25° and 40° and latitude above 35°) included 22 939 occurrences of 2937 unique species over 2822 localities.

Data describing dental traits of the species came from the NOW database as well. We used the relative molar crown height (hypsodonty) and the number of longitudinal cutting edges (lophs) recorded at the species level. All the occurrences from species of the following orders were included in the dental-based environmental estimations: Perissodactyla, Artiodactyla, Primates, Proboscidea and Hyracoidea. Hypsodonty in NOW is reported as categories 'bra', 'mes', 'hyp' and 'hys', corresponding to brachydont, mesodont, hypsodont and hypselodont. We transformed this into an ordinal variable assigning values 1 and 2 for 'bra' and 'mes', respectively, and 3 for both 'hyp' and 'hys' following [25]. Longitudinal loph count in NOW ranged from 0 to 3 (many), but following [27], from which the predictive models come, we capped the number of longitudinal lophs at 2.

3. Computational methods

Our main methodological task is to define what makes the patterns of origination exceptional. The main challenge is varying sizes of the number of species at localities in time and space, which may be an artefact due to uneven sampling of fossil finds or it may be a genuine variation in diversity due to varying carrying capacity of the environment, or both. Rather than assuming and modelling sampling process for the purpose of data 'correction', we aim to directly model expectations for exceptional and non-exceptional patterns of occurrences along with any variations in sampling that there may be.

(a) Detecting exceptional origination (species factories)

Our task is to detect areas in space and time that show exceptionally high amounts of first occurrences. We take this as a potential sign of origination hotspots (species factories).

All else being constant we would expect to see high numbers of first occurrences in areas and times from which we have many samples from. Looking at the ratio of first to all occurrences does not quite solve the problem of exaggerated number of first occurrences. Exaggeration would be the highest if we proceed from a very poorly sampled time bin

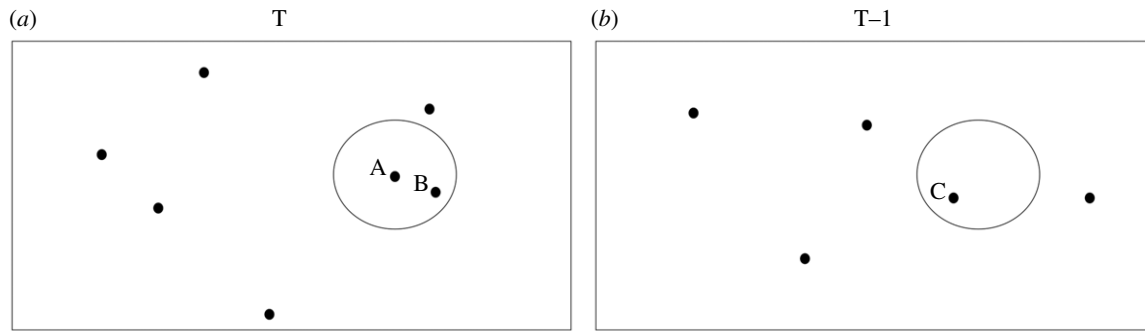


Figure 1. A toy example on spatial weighting: (a) assumed geographical area of interest at the current time unit, (b) the same geographical area at the previous time unit. A is the focal point (a potential species factory) for which the regression is made. Dots indicate nearby localities with fossil occurrences. A circle of a fixed radius (e.g. 500 km) is drawn around the focal locality A in figure (a) with another circle drawn in figure (b) at the same place geographically.

to a very intensively sampled time bin. Fossil data will never be complete enough for this not to happen at all, but the uneven sample sizes at different localities may cause us to miss even more new species than on average. Sample sizes may differ from locality to locality not only due to uneven preservation or recovery, but also due to different productivities of environments. Less productive environments would generally accommodate fewer species. In addition, the number of occurrences may differ due to differences in time averaging, that is, some localities might have accumulated organismal remains for longer time than others.

Existing statistical methods for correcting for unequal sampling, including capture–mark–recapture [45–49] do not take into consideration potential variations in diversity due to varying productivity of environments. In other words, they require productivity of the environment (and thus organismal diversity) to be constant over time and attribute all differences in diversity to uneven sampling. This does not suit our purpose, where we aim to analyse species factories in the context of environmental change.

Here we make a pilot attempt to circumvent this challenge. When determining the baseline origination threshold we aim to model transitions rather than absolute states. Our model assumes that the magnitude of environmental change trends are comparable across time units, despite absolute differences in productivity. We rely on the assumptions that the carrying capacity is similar locally geographically and that it changes in a similar way over time for localities that are close to each other geographically. Thus, for each locality of interest in time and space we fit an individual localized model, which captures changes in nominal diversity taking into account its neighbouring localities. The main assumption behind this approach is that differences in diversity of spatially nearby localities are due to sampling, but differences across time units over local mean diversity is due to changes in the carrying capacity. With these spatially explicit models we aim to predict the number of first occurrences to be observed given the sample size, no matter whether the number of occurrences is large or small due to sampling, productivity or time averaging.

We model the expectation for the number of first occurrences as a function of sampling over consecutive time bins. We model the number of events, not rates. For this, we train a logistic regression model on locally weighted data. Logistic regression is a classical statistical model that internally has a linear regression, which is topped up by a sigmoid function such that the output of the model is a probability estimate between 0 and 1. The logistic regression

has two input variables representing sampling intensities ‘before’ and ‘now’ (for last occurrences these would be ‘now’ and ‘after’) and one target variable denoting whether the occurrence is a first occurrence of a species in time.

We weight occurrences spatially such that in the regression modeling occurrences that are close to our focal point are weighted more than occurrences that are far away. Figure 1 illustrates the spatial treatment. For this toy example, let us assume that locality A has four occurrences with one first occurrence, locality B has three occurrences and locality C (in time unit $T - 1$) has five occurrences. Distance from locality A to B is 400 km and C is 350 km away from the centre of the circle (locality A). Then, there are $4 + (1 - 400/500) \cdot 3 = 4.6$ weighted occurrences at time unit T and $(1 - 350/500) \cdot 5 = 1.5$ weighted occurrences at time unit $T - 1$. We add one row (data point) for each occurrence in locality A to the overall regression model. In this case, we add four identical rows of [1.5, 4.6] as inputs and we add the vector [1, 0, 0, 0] to the overall target vector. Similar procedure is repeated for every locality in every time unit starting from MN2 so that every occurrence is included once in the dataset on which the model is fit.

The logistic regression gives us a way to estimate the probability that a given occurrence (with given sampling densities) is a first occurrence and then, given how many weighted occurrences a locality has, we can estimate how many weighted first occurrences to expect at any given grid point. Then, we estimate the probability of observing as many or more weighted first occurrences as counted at the focal grid point using the binomial distribution. Further details on modelling data and methodological choices are described in electronic supplementary material, the appendix.

(b) Generalization to exceptional extinction patterns and species sorting

Similar principles as we described for identifying exceptional origination hotspots can be applied to analyse exceptional patterns of extinction and sorting. For extinctions, instead of first occurrences one should estimate the probability of observing as many or more last occurrences. Otherwise, the process is analogous to origination, but when creating the regression dataset, we look at occurrences in the current and next time unit (instead of current and previous), and we use occurrences within time units MN1–MQ18.

Detecting exceptional species sorting patterns (immigration and local extinction) requires an additional data processing

step. We look at a geographical area around the focal point with a given radius in two consecutive time units ('current' and 'previous' for immigration, and 'current' and 'subsequent' for local extinction). For both time units, we list all the unique species observed within the focal area. Then, we remove all species from those two lists that either are first seen in the second time unit or last seen in the first time unit, i.e. we only keep species that actually exist somewhere in the world in both time units. Comparing these two lists now give us the lower bound on which species were present in the focal area in both time units and which may have become locally extinct or immigrated into it.

When creating the regression dataset to evaluate local extinction, we add a row for each species that is listed in the first time unit and include the sum of weighted occurrences in the first time unit and the second time unit as inputs of the model, and a one or a zero as the target, depending whether the species went locally extinct or not. Here, local extinction refers to species that were present in the focal area in the first time unit but not in the second, while they were observed somewhere else in the world at that time bin. The data for the immigration are prepared analogously. Similarly to origination and extinction, we fit a logistic regression with two input variables, weighted occurrences in the focal area in the two consecutive time units, and an intercept to the data.

(c) Estimation of environmental contexts

Dental characteristics of plant eating mammal communities can be used to infer palaeoenvironmental conditions [25]. Here, we use the regression model presented by Liu *et al.* [27] that uses dental crown height (hypsodonty, HYP) and the count of longitudinal cutting edges (lophs, LOP) of molar teeth within large plant eating mammal communities as inputs to estimate the primary productivity (NPP) of the environment. As any predictions about the past, these estimates are coarse and come with uncertainties.

We weigh observations spatially the same way as described in figure 1 in attaining average distance weighted HYP and LOP values for every grid point (where HYP and LOP scores are available in NOW database). HYP values are ordinal, roughly corresponding to the ratio scale. LOP values are counts. Taking the average over HYP and LOP is meaningful biologically and computationally.

(d) Implementation

The data and analysis pipeline is available on GitHub (https://github.com/jaksticks/NOW_codes). The code is in Python. We used the statsmodels package for Python to perform all the regression analyses. The analysis is scripted in Jupyter notebooks.

For our main analysis, we used a radius of 500 km for the spatial aggregation. We also performed the same analyses using a 100 km radius, the corresponding complementary plots are available in the code repository.

4. Results

(a) Palaeoenvironments

Figure 2 maps estimates of environmental productivity. We observe that Early Miocene eastern Europe (Turkey in particular) begins to drop in NPP before the rest of Europe. This becomes

particularly evident in MN9 and MN10. By MN11 and MN12 most of Europe has also dropped considerably in NPP. From MN13 to MN16 there appears to be a fairly clear division between more lush environments (higher estimated NPP) in central or northern central Europe, whereas western and eastern Mediterranean remain arid (low in estimated NPP). From MN17 onwards we observe further overall aridification.

Figures 3 and 4 depict identified origination, extinction and sorting hotspots. Figures of origination, extinction, immigration and local extinction separately are given in electronic supplementary material.

In the earlier part of the Miocene, up to and including MN10, origination hotspots are primarily on the more arid side of the environmental gradient (figure 5a). Extinction hotspots appear quite dynamic throughout (figure 5b). Meanwhile, the estimated productivity for sorting hotspots is mostly in par with the mean for non-hotspots up to and including MN7–8 (figure 5c,d). Immigration hotspots show quite wild environmental fluctuations between MN9 and MN12, and then up to MN17 they tend to be found on the more lush side compared to non-hotspot areas. Local extinction also shows much more variation (compared to non-hotspots) after and including MN10 than it did prior to that time bin.

(b) Faunal dynamics in the Early Miocene

The Early Miocene (MN2–MN4) displays widespread origination hotspots occurring (figure 3). While this may be partially due to modest representation of older records in Europe in the NOW data, the identified hotspots in these Early Miocene times extend far beyond the region with increasingly dense sampling (electronic supplementary material, figure S12). In particular, central Europe as well as Turkey and Greece continuously appear as hotspots, most prominently in MN3. At the same time bin (MN 3) central Europe appears as an immigration hotspot (figure 4). The map of productivity estimates suggests that during MN3 central Europe might have had slightly more productive environment than the surrounding areas and also more productive than in the previous time step. Turkey and Greece lack large mammal data to display environmental estimates for MN3, but the territory appears much more productive than the European context in MN4 and also partially lights up as an origination hotspot.

The red area in MN2 in the Iberian peninsula does not have central localities (no black dots), but lights up red due to neighbouring localities within 250 km (500 km radius) showing exceptionally many new species. Perhaps in future work this could be remedied with an exponentially decaying weight scheme for the occurrences or by cutting out grid points that are too far away from any localities.

Figure 5 compares environmental estimates at identified hotspots against regular areas over time. We see that overall statistically in the Early Miocene origination hotspots tend to occur at the more arid end of the environment, thus the visual pattern in central Europe and Turkey makes an exception. An interesting observation emerges when comparing trends, we can see that the estimated productivity around origination and extinction hotspots steadily increases over MN2–MN4 against more or less steady level of productivity for regular sites at the same time period. At the same time, immigration patterns show the reverse—immigration hotspots stay at the steady level while regular spots show an

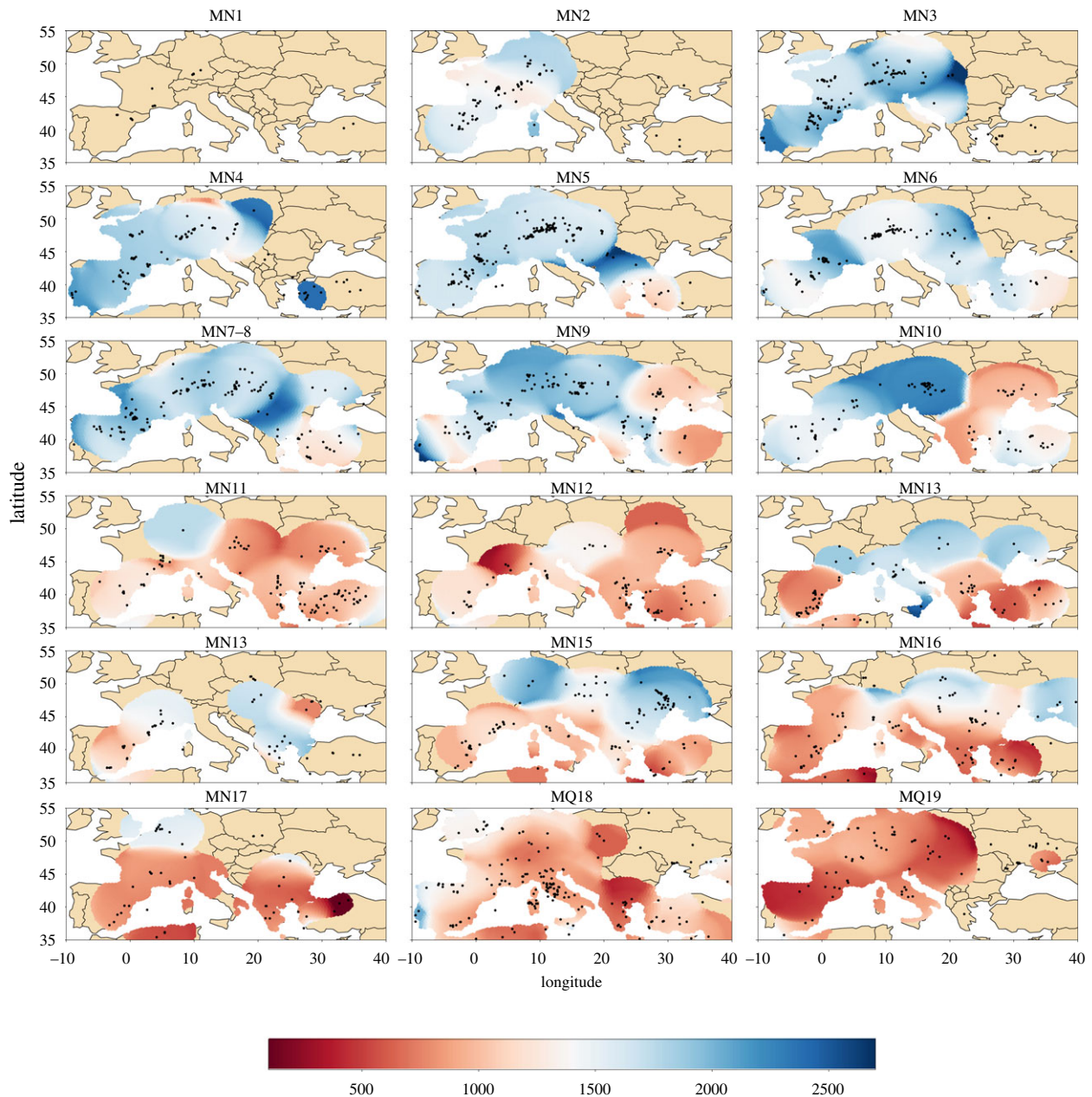


Figure 2. Net primary productivity (with units $\text{g Cm}^{-2} \text{yr}^{-1}$) estimated from dental traits of large plan eating mammal community data. Blue indicates areas with high estimated productivity and red indicates areas with low productivity. All spatial grid points with at least two weighted occurrence is displayed. Black dots indicate the location of all fossil localities designated for the given time unit.

upward trend. Intuitively, this makes sense, origination and extinction hotspots seem to be at the frontier of environmental change, whereas immigration hotspots are those places that carry on the previous conditions while globally the environment changes; in other words, refugia.

(c) Faunal dynamics in the Middle Miocene

The Middle Miocene appears as time of stability in terms of origination and extinction. One prominent area is northeast Spain, which appears as an origination hotspot already since MN4 and carries onwards as a hotspot all the way through MN7–8. This might be related to a known diachrony of the MN system, with the temporal time correlates of MN units 4–6 being consistently younger in Spain than in the eastern Mediterranean or central Europe, so that some actually later-occurring taxa might appear to occur ahead of their time in Spain [50]. Despite this offset, Turkey also displays

recurring origination hotspots in the same time frame, most prominently in MN5. Thus it would clearly be premature to attribute these patterns to diachrony alone.

While the majority of Europe remains relatively stable in terms of origination, many more hotspots of sorting appear, which are also dynamically intermixed during this period. Migration patterns that start in MN4 continue through MN5—geographically central Europe acts as a source area, while southeast and (later) southwest act as sink hotspots. It seems this could be a natural succession of the Early Miocene species factory in central Europe; the ‘manufactured’ species now move away. Environmentally, we indeed see (figure 2) that the directions of dispersal preserve slightly higher productivity in MN5, especially the southwest.

In MN6 and even more so in MN7–8 the directions of species sorting hotspots reverse—the central north becomes the immigration hotspot. However, there is no clear environmental contrast in association with this. The box plots

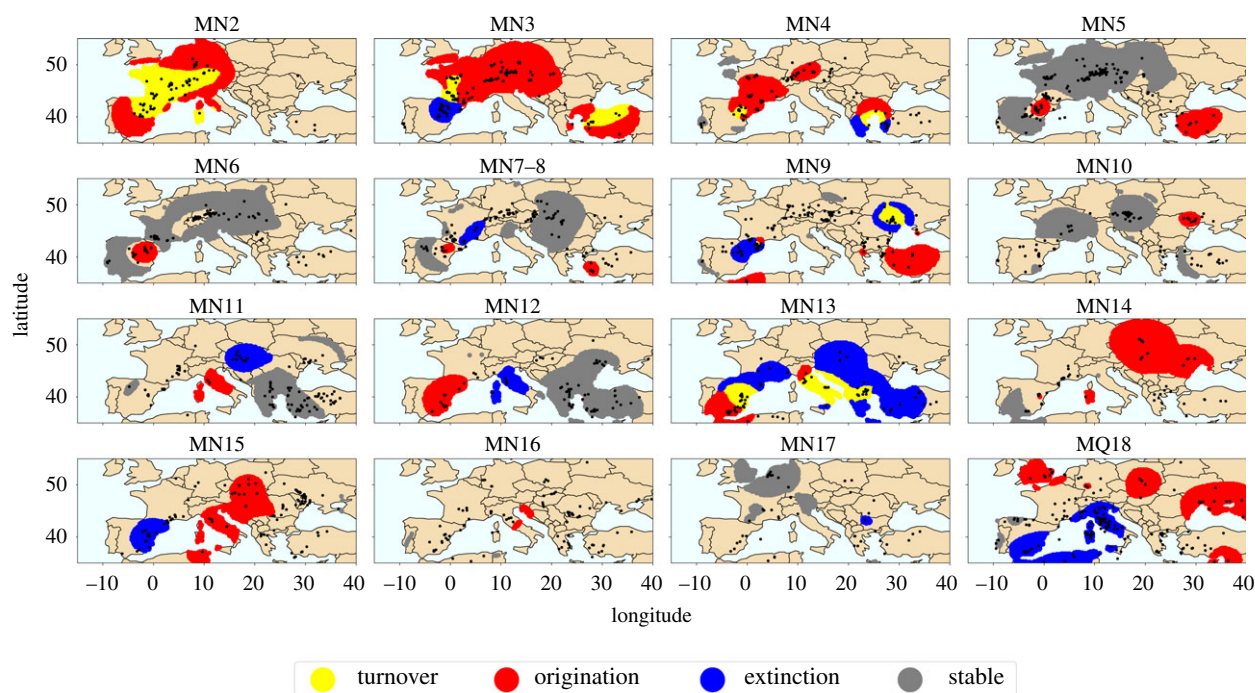


Figure 3. Diversity patterns. Red indicates areas with statistically significant ($p < 0.05$) number of weighted first occurrences (origination). Blue indicates areas with statistically significant ($p < 0.05$) number of weighted last occurrences (extinction). Yellow indicates areas with statistically significant ($p < 0.05$) number of weighted first and last occurrences (turnover). Grey indicates areas with statistically small ($p > 0.95$) number of weighted first and last occurrences (stable). All spatial grid points with at least one weighted occurrence is displayed. Black dots indicate the location of all fossil localities designated for the given time unit.

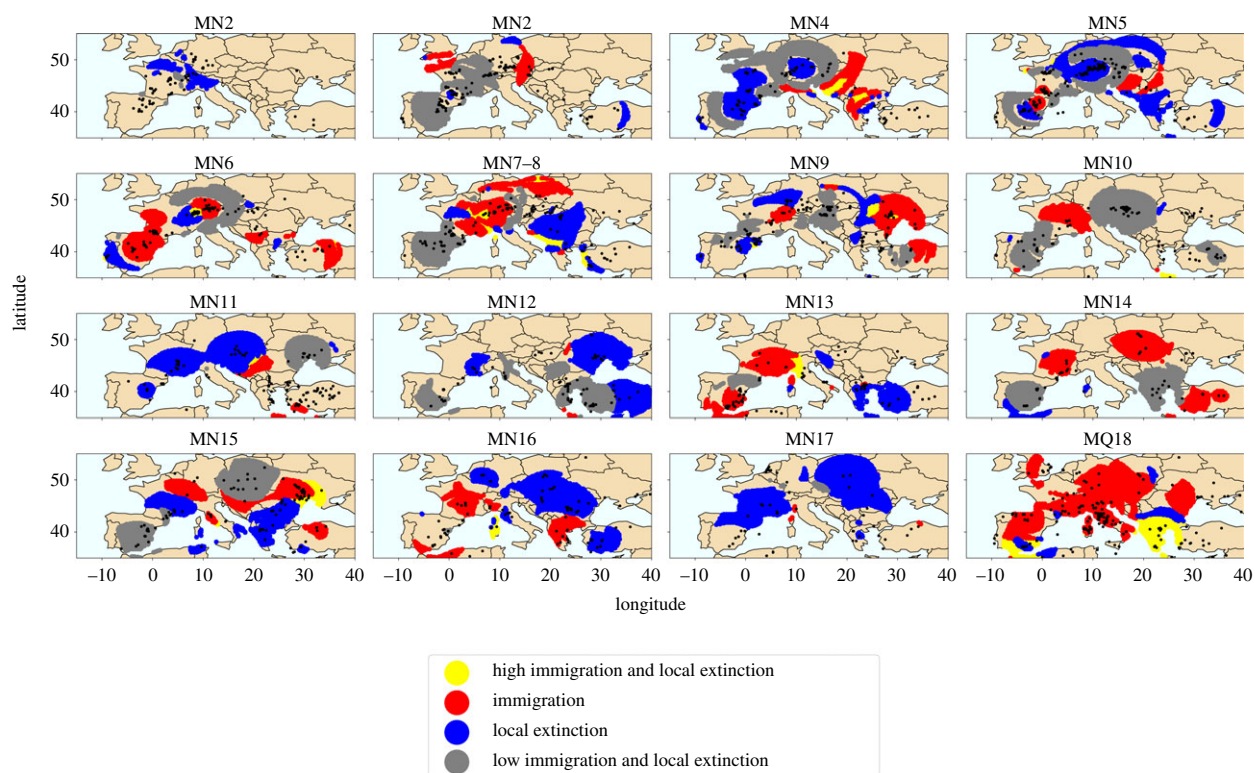


Figure 4. Sorting patterns (local appearance and local extinction). Red indicates areas with statistically significant ($p < 0.05$) immigration. Blue indicates areas with statistically significant ($p < 0.05$) local extinction. Yellow indicates areas with statistically significant ($p < 0.05$) immigration and local extinction together. Grey indicates areas with statistically small ($p > 0.95$) immigration and local extinction together. All spatial grid points with at least one weighted occurrence in the focal, preceding and subsequent time unit is displayed. Black dots indicate the location of all fossil localities designated for the given time unit.

comparing environmental conditions of sorting hotspots with non-hotspots are more dynamic than in the Early Miocene, the conditions progress to more arid in MN6, but then go back to the Early Miocene level in MN7–8.

Overall, distribution of conditions of sorting hotspots against background conditions strongly suggest that species sorting tracks stable conditions. It does not mean that conditions need to remain stable at the same place.

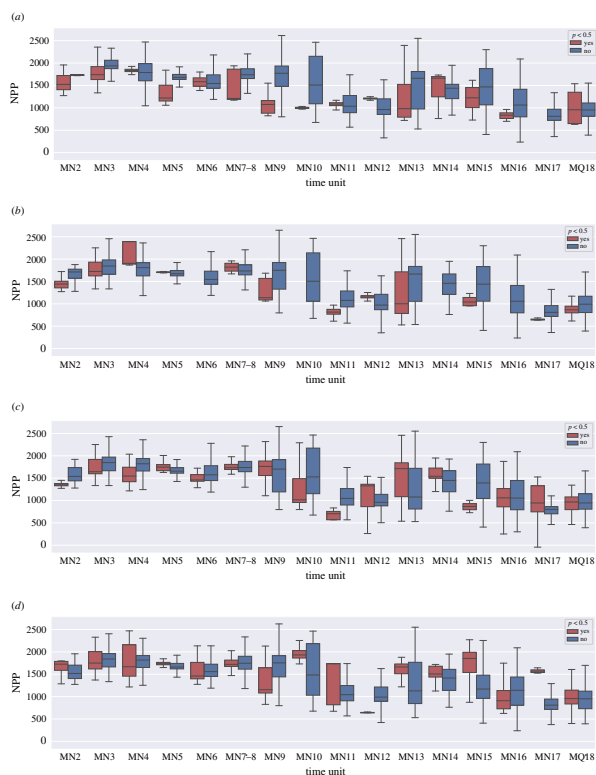


Figure 5. Comparison of origination, extinction and species sorting hotspots versus non-hotspots with respect to net primary productivity (NPP). In each plot, two boxplots are shown for each time unit: red for a given hotspot type and blue for corresponding non-hotspots. The calculations include all spatial grid points, which have at least two weighted occurrences with the relevant HYP and LOP data available. (a) Speciation. (b) Extinction. (c) Local extinction. (d) Immigration.

We rather see that when many sorting hotspots occur (like in the Middle Miocene here), those hotspots, especially the immigration hotspots, conserve the environmental conditions of previous times, acting as refugia [51].

(d) Faunal dynamics in the Late Miocene

The known Vallesian pattern is seen in MN9 western Europe, but this appears to be a local, rather than a continent-wide event, which is in line with recent research [52]. Eastern Europe, however, does exhibit another extinction hotspot in MN9 (figure 3). Both Spain and eastern Europe spots show as notably more arid than the rest during the time period (figure 2). At the same time Turkey, which also shows as extremely arid in MN9, appears as an origination hotspot. After several time units of stability, associated with higher productivity estimates during MN10–MN12, in MN13 Turkey displays large local extinction and permanent extinction hotspots, which might be interpreted as biodiversity sinks. The estimates show that in MN13 Turkey is extremely arid along with Spain, while the northern territories across Europe return to more productive environments.

Overall, the Late Miocene has the most a notable dip in estimated aridification in MN11 and MN12 (figure 6). In MN11, both local and permanent extinction hotspots by a wide margin environmentally harsher than the rest of the localities (figure 5). Those might be interpreted as sink areas. Those highly prominent local and permanent extinction hotspots in MN11 cover large areas in central-southern

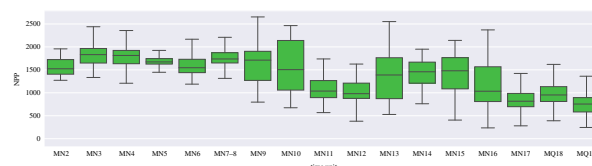


Figure 6. Net primary productivity over time estimated from dental traits of large plant eating mammal communities. The calculations include all spatial grid points, which have at least two weighted occurrences with the relevant HYP and LOP data available. (Online version in colour.)

Europe (figures 4 and 3). Italy at the same time acts as an origination hotspot, quickly becoming an extinction hotspot in the following MN12 time unit. This probably reflects the presence of insular, highly endemic faunas in Italy at this time.

The origination hotspot trend in the Late Miocene is similar to the pattern observed in the Early Miocene. Origination hotspots start the most arid as compared to the background conditions and steadily increase in productivity while the background conditions follow the European v-shaped trend in average aridity. In other words, it seems that the defining feature of species factories to be exceptional environmentally and represent conditions that soon will become wide spread broadly holds. In the Late Miocene, species factories start arid and then become like the rest. It is more difficult to make clear conclusions from immigration contrasts at the same period, but at least in MN10, which is the first notable time step towards more arid conditions, origination hotspots are notably more arid than the background, while immigration hotspots are notably less arid than the background (figure 5).

(e) Faunal dynamics in the Pliocene and Pleistocene

The final stage of our analysis firstly (MN14–MN16) display highly polarized environments, continuing the pattern from MN13—productive in the north and more arid in the south, and then at the latest time stages (MN17–MQ18) everything becomes quite arid. In MN14, the productive northeast acts as a large origination hotspot, to an extent continuing into MN15 and shifting south in MN16 (figure 3). Again, MN16 is just before everything gets more arid and at that unit origination hotspots are in harsher conditions than the rest of the localities (figure 5). At the same time of polarized Europe MN14 and MN15 show notable immigration hotspots in central Europe (figure 4) that appear notably more productive in terms of NPP than the background conditions (figure 5). In MN16, when the last margins of the productive north begin to disappear central Europe and especially central eastern Europe become one huge local extinction hotspot (figure 4). While the productivity maps look like the Black Sea shores in MN16 might have acted as a refugium, the question is why species so massively go locally extinct and why is this not becoming a permanent extinction hotspot neither in MN16 nor in MN 17 (figure 3) when the last green colour disappears from NPP estimates (figure 2).

The final periods of the analysis are as quiet as they are dramatic. MN17 is extremely stable in terms of origination and extinction, but shows massive local extinction hotspots. An open question is in what other places they continuing after local extinctions. MQ18 shows extinction hotspots in the south and several large origination hotspots in the

north, as well as massive immigration hotspots. Could those that left in MN17 now be getting back?

(f) Species longevity and occupied area

Considering only species that had a first and a last occurrence within MN2–MQ18 in the study area, we find that the average age of origination hotspot species (species that had at least one first occurrence in an origination hotspot) is 3.5 Myr, whereas non-hotspot species (species without any first occurrence in an origination hotspot) only had an average observed lifespan of 2.5 Myr. If we for a moment disregard species that occurred in a single time unit, then the corresponding average longevity of the remaining species would be 5.0 and 4.1 Myr, respectively.

We further estimated the geographical ranges of each species over their lifetime. Specifically, for each species and time unit, if a given species had three or more occurrence locations, we calculated the area of the convex hull of those locations. When only two locations existed for a given species and time unit, we calculated the distance between those locations and assumed a 10 km wide rectangle between those two locations for which we calculated the corresponding area. If there was only one location for a given species and time unit, we assumed the species range to be zero for that instance. We found that origination hotspot species occupied an average area of 135 000 km² and non-hotspot species occupied an average area of 86 000 km² over their longevity. Excluding single time unit species, we found that hotspot and non-hotspot species' average occupied areas were 229 000 km² and 177 000 km², respectively.

We also looked at average longevity and occupied areas for species that had a last occurrence in an extinction hotspot. We found that the mean longevity was 3.1 and 2.9 Myr for hotspot and non-hotspot species, respectively. Removing single time unit species, we found that the mean longevity for hotspot species was 4.6 Myr and for non-hotspot species it was 4.5 Myr. The mean occupied area for hotspot species was 176 000 km² and for non-hotspot species it was 68 000 km². Removing single time unit species, we found that the mean occupied area for hotspot species was 333 000 km² and for non-hotspot species it was 125 000 km².

5. Discussion

Species factory as an evolutionary concept has existed as part of the informal discourse of evolutionary palaeontologists for a long time [7,53], but apart from our earlier study in adaptive dynamics [54] appears not to have been defined explicitly or formalized computationally. The term has been used in a broad sense for the source area of exceptionally many species [6]. Empirically the concept has been applied in several contexts to explain the dynamics of mammalian faunas—the latest Middle Miocene origins of the Pikermian palaeobiome in the Sub-Parathethyan province [7], the Pliocene origin of Pleistocene megafauna in the Tibetan Plateau [9] and the rise of arid-adapted fauna in the Plio-Pleistocene of the eastern African Turkana Basin [8]. How prevalent the phenomenon is across the world and what macroevolutionary mechanisms may be behind it has remained unclear.

Our attempt to define and account for species factories and related concepts in a systematical way at a large spatial and temporal scale shows that the phenomenon can be

detected computationally. The phenomenon, of course depending on the thresholds for deciding what is exceptional, appears to be more common than the few anecdotal cases known previously. Based on the patterns resulting from our analysis, it seems that species factories are by and large a matter of being present in locations that represent conditions that are about to become widespread. Whether adaptations leading to these originations are somehow special or simply a matter of lucky placement remains an open question but from our pilot analysis we see that species from recognized species factories do tend to live longer and become more widespread than species on average. We hope follow up studies will address this question in more detail.

A refugium is the macroevolutionary counterpart of species factory. The term refers a location that supports relict populations of previously more widespread species. The concept has been commonly used to describe survival through and recovery from adverse conditions, such as glacial–interglacial cycles. Here we label as refugia areas that receive abnormally many survivors that have become locally extinct elsewhere. Such areas can be a long or short term refugia depending on what happens next. The absence of extinction hotspots along with refugia implies that a refugium was successful; the species did not come there to die, but to survive through bad times and carry on.

From this perspective, species sorting appears as a more complicated phenomenon than origination, not expected to follow any simple pattern. It draws on availability of immigrants, availability of dispersal routes, the absence of competitors and other such factors that are challenging to account for analytically.

The interpretation of the extinction hotspots is less clear, but from the perspective of species survival they could be either well-sampled places of bad luck at bad times, or they could be short-lived refugia. An open question concerns the large differences in species range between extinction hotspot and non-hotspot species. Why do common species appear there last before extinction, while rare species do not? Could this be a matter of broad versus local adaptation?

Another open question is what proportion of species originate from 'species factories'. While we showed that species factories exist, we could not within the methodology of our pilot analysis assess the prevalence of the factories. How common are they? Clearly, species factories cannot be as common a mechanism as the majority rule postulated by Vermeij & Dietl [2]. A species factory is an exception but not an unexpected one. It does require a certain configuration of environmental changes over time and space but over millions of years of environmental change that particular configuration will inevitably occur from time to time.

The computational methods put together in this paper allow us to find exceptional patterns from fossil occurrence data. To the extent of the coverage that the fossil record can give us, we expect these patterns to be robust with a natural level of noise present in the data. The tailored computational solution involved many design choices. As always, choices might have been different, the solution that we report here was driven by feasibility and transparency of the approach. We did not explicitly model sampling effects on diversity versus true variation in diversity due to changing carrying capacity. This is an open question for which we do not have resolutions to offer yet. Nonetheless, we hope that we captured both effects implicitly to a crude approximation

via locally weighted models. Rather than modelling sampling versus carrying capacity explicitly, the idea of the current study was to model how many events are expected assuming that the carrying capacity was similar locally geographically and did not change fast in time. In this case, we expected the localized regression approaches with local weighing to take care of potential unevenness in sampling.

The hotspots that we found open broader research directions for future: what is the community composition, what are the environmental conditions, are there common features between different hotspots? Do species from local extinction hotspots tend to move to immigration hotspots? Identifying and studying these different hotspot areas can help us learn more about important evolutionary and ecological mechanisms over geological time scales.

6. Conclusion

We set out to identify species factories computationally. For that we developed a regionally informed statistical approach to detect exceptional places and times of species origination, extinction and sorting in the fossil record. Applying this approach to the mammalian fossil record of the Late Cenozoic we found that early hotspots were characterized by species adapted to environments that subsequently became more common, giving these hotspot species a potential selective advantage. Our pilot results also suggest that species

originating in hotspots had on average longer longevity and larger geographic ranges than other species.

Species factories thus appear to be a real phenomenon. Our analysis suggests that species factories can be found particularly in synch with gradual but persistent directional changes of environmental conditions. A species from the factory is thus a lucky species—appearing at the right time in the right place, and with the right set of evolutionary raw material for the new adaptations required.

Data accessibility. All scripts used in this study are openly accessible through <https://github.com/StochasticBiology/boolean-efflux.git>. The data are provided in electronic supplementary material [55].

Authors' contributions. J.T.: conceptualization, formal analysis, investigation, methodology, software, validation, visualization, writing—original draft, writing—review and editing; M.F.: conceptualization, data curation, investigation, methodology, validation, writing—review and editing; I.Ž.: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, supervision, validation, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Competing interests. We declare we have no competing interests.

Funding. This work was supported by the University of Helsinki startup grant and the Academy of Finland grant no. 1314803 (to I.Ž.).

Acknowledgements. We thank Giulia Varvara for comments on computational approaches to correcting for potential sampling biases in the fossil record. We thank Michael Mechenich for advice on spatial interpolation. We thank Raymond Bernor for helping to track down the history of the term 'species factory'.

References

- Raup D, Sepkoski J. 1982 Mass extinctions in the marine fossil record. *Science* **215**, 1501–1503. (doi:10.1126/science.215.4539.1501)
- Vermeij G, Dietl G. 2006 Majority rule: adaptation and the long-term dynamics of species. *Paleobiology* **32**, 173–178. (doi:10.1666/0094-8373(2006)32[173:MRAATL]2.0.CO;2)
- Marshall C, Quental T. 2016 The uncertain role of diversity dependence in species diversification and the need to incorporate time-varying carrying capacities. *Phil. Trans. R. Soc. B* **371**, 20150217. (doi:10.1098/rstb.2015.0217)
- Rabosky D. 2013 Diversity-dependence, ecological speciation, and the role of competition in macroevolution. *Ann. Rev. Ecol. Syst.* **44**, 481–502. (doi:10.1146/annurev-ecolsys-110512-135800)
- Rosenzweig M. 1995 *Species diversity in space and time*. New York, NY: Cambridge University Press.
- Fortelius M, Eronen J, Kaya F, Tang H, Raia P, Puolamaki K. 2014 Evolution of Neogene mammals in Eurasia: environmental forcing and biotic interactions. *Annu. Rev. Earth Planet. Sci.* **42**, 579–604. (doi:10.1146/annurev-earth-050212-124030)
- Bernor R, Koufos G, Woodburne M, Fortelius M. 1996 The evolutionary history and biochronology of European and Southwest Asian Late Miocene and Pliocene hipparionine horses. In *The evolution of Western Eurasian Neogene mammal faunas* (eds R Bernor, V Fahlbusch, W Mittmann), pp. 449–470. New York, NY: Columbia University Press.
- Fortelius M *et al.* 2016 An ecometric analysis of the fossil mammal record of the Turkana basin. *Phil. Trans. R. Soc. B* **371**, 20150232. (doi:10.1098/rstb.2015.0232)
- Deng T *et al.* 2011 Out of Tibet: Pliocene woolly rhino suggests high-plateau origin of ice age megaherbivores. *Science* **333**, 1285–1288. (doi:10.1126/science.1206594)
- Darwin Ch. 1871 *On the origin of species*, 6th edn. London, UK: John Murray.
- Raup D. 1972 Taxonomic diversity during the Phanerozoic. *Science* **177**, 1065–1071. (doi:10.1126/science.177.4054.1065)
- Uhen M *et al.* 2013 From card catalogs to computers: databases in vertebrate paleontology. *J. Vertebr. Paleontol.* **33**, 13–28. (doi:10.1080/02724634.2012.716114)
- Brocklehurst N, Frobisch J. 2018 The definition of bioregions in palaeontological studies of diversity and biogeography affects interpretations: Palaeozoic tetrapods as a case study. *Front. Earth Sci.* **6**, 200. (doi:10.3389/feart.2018.00200)
- Catena A, Croft D. 2020 What are the best modern analogs for ancient south american mammal communities? Evidence from ecological diversity analysis (EDA). *Palaeontol. Electronica* **23**, a03. (doi:10.26879/962)
- Eronen J, Mirzaie Ataabadi M, Micheels A, Karne A, Bernor R, Fortelius M. 2009 Distribution history and climatic controls of the Late Miocene Pliocene chronofauna. *Proc. Natl Acad. Sci. USA* **106**, 11 867–11 871. (doi:10.1073/pnas.0902598106)
- Kaya F, Bibi F, Zliobaite I, Eronen J, Tang H, Fortelius M. 2018 The rise and fall of the old world savannah fauna and the origins of the african savannah biome. *Nat. Ecol. Evol.* **2**, 241–250. (doi:10.1038/s41559-017-0414-1)
- McKinney M, Zachos L. 1986 Echinoids in biostratigraphy and paleoenvironmental reconstruction: a cluster analysis from the Eocene Gulf Coast (Ocala limestone). *PALAIOS* **1**, 420–423. (doi:10.2307/3514480)
- Mirzaie Ataabadi M, Liu L, Eronen J, Bernor R, Fortelius M. 2013 Continental scale patterns in Neogene mammal community evolution and biogeography: a Europe-Asia perspective. In *Fossil mammals of Asia: Neogene biostratigraphy and chronology* (eds X Wang, LJ Flynn, M Fortelius), pp. 629–655. New York, NY: Columbia University Press.
- Valentine J, Peddicord R. 1967 Evaluation of fossil assemblages by cluster analysis. *J. Paleontol.* **41**, 502–507.
- Fortelius M *et al.* 2006 Late Miocene and Pliocene large land mammals and climatic changes in Eurasia. *Palaeogeogr. Palaeoclimatol.*

- Palaeoecol.* **238**, 219–227. (doi:10.1016/j.palaeo.2006.03.042)
21. Puolamaki K, Fortelius M, Mannila H. 2006 Seriation in paleontological data using Markov chain Monte Carlo methods. *PLoS Comput. Biol.* **2**, e6. (doi:10.1371/journal.pcbi.0020006)
 22. Ukkonen A, Fortelius M, Mannila H. 2005 Finding partial orders from unordered 0-1 data. In *Proc. of the eleventh ACM SIGKDD Int. Conf. on knowledge discovery in data mining* (eds Grossman R, Bayardo R, Bennett KP), pp. 285–293. New York, NY: Association for Computing Machinery.
 23. Barr W. 2018 Ecomorphology: reconstructing cenozoic terrestrial environments and ecological communities. In *Methods in paleoecology: reconstructing Cenozoic terrestrial environments and ecological communities* (eds DA Croft, DF Su, SW Simpson), pp. 339–349. Berlin, Germany: Springer.
 24. Croft D. 2001 Cenozoic environmental change in south america as indicated by mammalian body size distributions (cenograms). *Diversity and Distributions* **7**, 271–287. (doi:10.1046/j.1366-9516.2001.00117.x)
 25. Fortelius M *et al.* 2002 Fossil mammals resolve regional patterns of eurasian climate change over 20 million years. *Evol. Ecol. Res.* **4**, 1005–1016.
 26. Hernandez-Fernandez M, Vrba E. 2006 Pliopleistocene climatic change in the Turkana basin (East Africa): evidence from large mammal faunas. *J. Hum. Evol.* **50**, 595–626. (doi:10.1016/j.jhevol.2005.11.004)
 27. Liu L, Puolamaki K, Eronen J, Ataabadi M, HERNESNIEMI E, Fortelius M. 2012 Dental functional traits of mammals resolve productivity in terrestrial ecosystems past and present. *Proc. R. Soc. B* **279**, 2793–2799. (doi:10.1098/rspb.2012.0211)
 28. Vermillion W, Polly P, Head J, Eronen J, Lawing A. 2018 Ecometrics: a trait-based approach to paleoclimate and paleoenvironmental reconstruction. In *Methods in paleoecology: reconstructing cenozoic terrestrial environments and ecological communities* (eds DA Croft, DF Su, SW Simpson), pp. 373–394. Berlin, Germany: Springer.
 29. Alroy J *et al.* 2001 Effects of sampling standardization on estimates of phanerozoic marine diversification. *Proc. Natl Acad. Sci. USA* **98**, 6261–6266. (doi:10.1073/pnas.111144698)
 30. Alroy J *et al.* 2008 Phanerozoic trends in the global diversity of marine invertebrates. *Science* **321**, 97–100. (doi:10.1126/science.1156963)
 31. Finarelli J, Liow L. 2016 Diversification histories for North American and Eurasian Carnivorans. *Biol. J. Linnean Soc.* **118**, 26–38. (doi:10.1111/bij.12777)
 32. Foote M. 2003 Origination and extinction through the phanerozoic: a new approach. *J. Geol.* **111**, 125–148. (doi:10.1086/345841)
 33. Liow L, Finarelli J. 2014 A dynamic global equilibrium in carnivoran diversification over 20 million years. *Proc. R. Soc. B* **281**, 20132312. (doi:10.1098/rspb.2013.2312)
 34. Liow L, Nichols J. 2010 Estimating rates and probabilities of origination and extinction using taxonomic occurrence data: capture-mark-recapture (CMR) approaches. *Paleontol. Soc. Pap.* **16**, 81–94. (doi:10.1017/S108933260001820)
 35. MacKenzie D, Nichols J, Lachman G, Droege S, Royle J, Langtimm C. 2002 Estimating site occupancy rates when detection probabilities are less than one. *Ecology* **83**, 2248–2255. (doi:10.1890/0012-9658(2002)083[2248:ESORWD]2.0.CO;2)
 36. MacKenzie D, Nichols J, Hines J, Knutson M, Franklin A. 2003 Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology* **84**, 2200–2207. (doi:10.1890/02-3090)
 37. Matzke N. 2013 Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Front. Biogeogr.* **5**, 242–248. (doi:10.21425/F55419694)
 38. Eronen J, Puolamaki K, Liu L, Lintulaakko K, Damuth J, Janis Ch, Fortelius M. 2010 Precipitation and large herbivorous mammals II: application to fossil data. *Evol. Ecol. Res.* **12**, 235–248.
 39. Eronen J, Evans A, Fortelius M, Jernvall J. 2011 Genera are often better than species for detecting evolutionary change in the fossil record: a reply to Salesa *et al.*. *Evolution* **65**, 1514–1516. (doi:10.1111/j.1558-5646.2011.01220.x)
 40. The NOW Community. New and old worlds database of fossil mammals (now). Licensed under CC BY 4.0, 2020. See <http://www.helsinki.fi/science/now/>.
 41. Jernvall J, Fortelius M. 2004 Maintenance of trophic structure in fossil mammal communities: site occupancy and taxon resilience. *Am. Nat.* **164**, 614–624. (doi:10.1086/424967)
 42. Mein P. 1989 Updating of mn zones. In *European Neogene mammal chronology* (eds E Lindsay, V Fahlbusch, P Mein), pp. 73–91. Berlin, Germany: Springer.
 43. Steininger F. 1999 Chronostratigraphy, geochronology and biochronology of the Miocene ‘European land mammal mega-zones (ELMMZ)’ and the Miocene ‘mammal zones (MN-zones)’. In *The Miocene land mammals of Europe*, pp. 9–24. Munich: Verlag Dr. Friedrich Pfeil.
 44. Hilgen F, Lourens L, Van Dam J. 2012 The Neogene period. In *The geologic time scale* (eds FM Gradstein, JG Ogg, MD Schmitz, GM Ogg), pp. 923–978. Boston, MA: Elsevier.
 45. Chao A, Hsieh T, Chazdon R, Colwell R, Gotelli N. 2015 Unveiling the species-rank abundance distribution by generalizing the good-turing sample coverage theory. *Ecology* **96**, 1189–1201. (doi:10.1890/14-0550.1)
 46. Close R, Evers S, Alroy J, Butler R. 2018 How should we estimate diversity in the fossil record? Testing richness estimators using sampling-standardised discovery curves. *Methods Ecol. Evol.* **9**, 1386–1400. (doi:10.1111/2041-210X.12987)
 47. Connolly S, Miller A. 2001 Joint estimation of sampling and turnover rates from fossil databases: capture-mark-recapture methods revisited. *Paleobiology* **27**, 751–767. (doi:10.1666/0094-8373(2001)027<0751:JEOSAT>2.0.CO;2)
 48. Foote M, Sepkoski J. 1999 Absolute measures of the completeness of the fossil record. *Nature* **398**, 415–417. (doi:10.1038/18872)
 49. Sanders H. 1968 Marine benthic diversity: a comparative study. *Am. Nat.* **102**, 243–282. (doi:10.1086/282541)
 50. van der Meulen A *et al.* 2011 Biostratigraphy or biochronology? lessons from the early and middle miocene small mammal events in europe. *Geobios* **44**, 309–321. (doi:10.1016/j.geobios.2010.11.004)
 51. Fortelius M, Zhang Z. 2006 An oasis in the desert? History of endemism and climate in the late Neogene of North China. *Paleontographica* **277**, 131–141. (doi:10.1127/pala/277/2006/131)
 52. Madern P, van de Put J, Casanovas-Vilar I, van den Hoek Ostende L. 2018 Iberian micromammals show local extent of vallesian crisis. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **496**, 18–31. (doi:10.1016/j.palaeo.2017.12.033)
 53. Fortelius M *et al.*. 1996 Provinciality, diversity, turnover, and paleoecology in land mammal faunas of the later Miocene of western Eurasia. In *The evolution of Western Eurasian Neogene mammal faunas* (eds RL Bernor, V Fahlbusch, W Mittmann), pp. 414–448. New York, NY: Columbia University Press.
 54. Fortelius M, Geritz S, Gyllenberg M, Raia P, Toivonen J. 2015 Modeling the population-level processes of biodiversity gain and loss at geological timescales. *Am. Nat.* **186**, 742–754. (doi:10.1086/683660)
 55. Toivonen J, Fortelius M, Žliobaitė I. 2022 Do species factories exist? Detecting exceptional patterns of evolution in the mammalian fossil record. Figshare.