Article

Impacts of speciation and extinction measured by an evolutionary decay clock

https://doi.org/10.1038/s41586-020-3003-4

Received: 21 May 2020

Accepted: 16 September 2020

Published online: 09 December 2020

Check for updates

Jennifer F. Hoyal Cuthill[1,2,3,4]✉, Nicholas Guttenberg[3,5,6] & Graham E. Budd[7]

The hypothesis that destructive mass extinctions enable creative evolutionary radiations (creative destruction) is central to classic concepts of macroevolution[1,2]. However, the relative impacts of extinction and radiation on the co-occurrence of species have not been directly quantitatively compared across the Phanerozoic eon. Here we apply machine learning to generate a spatial embedding (multidimensional ordination) of the temporal co-occurrence structure of the Phanerozoic fossil record, covering 1,273,254 occurrences in the Paleobiology Database for 171,231 embedded species. This facilitates the simultaneous comparison of macroevolutionary disruptions, using measures independent of secular diversity trends. Among the 5% most significant periods of disruption, we identify the 'big five' mass extinction events[2], seven additional mass extinctions, two combined mass extinction–radiation events and 15 mass radiations. In contrast to narratives that emphasize post-extinction radiations[1,3], we find that the proportionally most comparable mass radiations and extinctions (such as the Cambrian explosion and the end-Permian mass extinction) are typically decoupled in time, refuting any direct causal relationship between them. Moreover, in addition to extinctions[4], evolutionary radiations themselves cause evolutionary decay (modelled co-occurrence probability and shared fraction of species between times approaching zero), a concept that we describe as destructive creation. A direct test of the time to over-threshold macroevolutionary decay[4] (shared fraction of species between two times ≤ 0.1), counted by the decay clock, reveals saw-toothed fluctuations around a Phanerozoic mean of 18.6 million years. As the Quaternary period began at a below-average decay-clock time of 11 million years, modern extinctions further increase life's decay-clock debt.

The destructive effects of extinction, especially mass extinction events, include the direct elimination of up to approximately 75% of living species[3], resulting in the decay of evolutionary and ecological communities[3,4] and potential ecosystem collapse[5]. However, major creative[6] impacts have also been hypothesized to result via vacation of ecological niches[4], post-extinction diversification[7], altered evolutionary trajectories[3,8] and shifts in the dominance of particular clades, including our own[3,5,6,9]. We group such latter hypotheses under the concept of evolutionary creative destruction. In the weak sense, this predicts that extinctions have often enabled subsequent diversifications[1]. In the stronger sense, the hypothesis of creative destruction can be expressed as a causative necessity: that major radiations require prior mass extinctions[1,3,5,10]. Recently, however, classic narratives of mass extinction, replacement and recovery have been called into question by complicating factors such as evidence for significant diversification predating a proposed enabling extinction[11] and protracted extinctions[12], as well as debates on the effects and rates of mass versus background extinction[2]. In addition, extinction and radiation may theoretically be more or less decoupled in time[10]. On one hand, new groups might radiate without a preceding decrease in diversity (pure evolutionary creation). On the other hand, biological groups lost in mass extinctions might not be replaced, either immediately or at all—for example, because of the temporary[1,2,4] or permanent elimination of the ecological niche that they represent (pure evolutionary destruction). Furthermore, we propose that the evolutionary radiation of one group may itself cause evolutionary decay (the dilution by origination, or erosion by extinction[4], of pre-existing communities), a concept that we describe, conversely, as destructive creation. However, the relative evolutionary impacts, balance and timing of radiation and extinction have not previously been quantitatively tested. These fundamental knowledge gaps affect assessments and predictions of the impacts of recent extinctions and of recovery potential, which require quantitative baselines from historical diversification and extinction data[3].

Machine learning of time structure in the fossil record

Our machine learning embedding method (Supplementary Computer Code 1, Extended Data Fig. 1a) allocates every fossil species a location in

[1]Institute of Analytics and Data Science, University of Essex, Colchester, UK. [2]School of Life Sciences, University of Essex, Colchester, UK. [3]Earth–Life Science Institute, Tokyo Institute of Technology, Tokyo, Japan. [4]Department of Earth Sciences, University of Cambridge, Cambridge, UK. [5]Cross Labs, Cross Compass Ltd, Tokyo, Japan. [6]GoodAI, Prague, Czech Republic. [7]Department of Earth Sciences, Palaeobiology Programme, Uppsala University, Uppsala, Sweden. ✉e-mail: j.hoyal-cuthill@essex.ac.uk

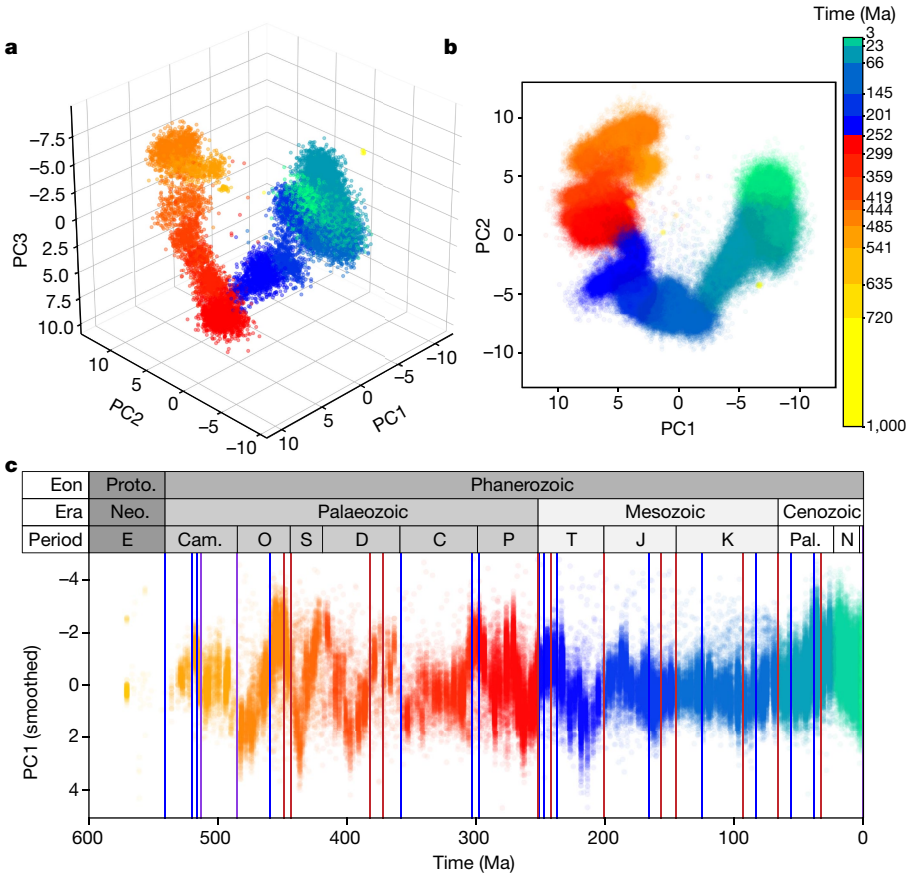


Fig. 1 | Time structure of the fossil record. a, First three principal component (PC1–PC3) axes from a 16-dimensional machine-learnt spatial embedding in which distance represents the probability of temporal co-occurrence (see Methods). **b**, First two PC axes. Points: $n = 171,231$ fossil species, occurring from 1,000 to 0 Ma (complete data set). Colours indicate geological periods: for example, red–blue, Permian–Triassic. **c**, First PC axis after moving-average smoothing, highlighting temporal shifts in co-occurrence structure (vertical movements, either up or down), independent of secular changes in diversity ($n = 171,173$ species, 600–0 Ma). Vertical lines, 5% most significant times of fractional species turnover (Fig. 3, Table 1): mass extinctions (red), mass radiations (blue) and mixed mass extinction–radiations (magenta). Proto., Proterozoic; Neo., Neoproterozoic; E, Ediacaran; Cam., Cambrian; O, Ordovician; S, Silurian; D, Devonian; C, Carboniferous; P, Permian; T, Triassic; J, Jurassic; K, Cretaceous; Pal., Palaeogene; N, Neogene.

a multidimensional spatial embedding, in which proximity represents the probability of temporal co-occurrence (the probability assigned by the machine learning model to whether species are observed to co-occur in time; see Methods). This optimizes, over the global record of species occurrences, the relative spatial position of each species, such that species that overlapped in time are close together and those that never coexisted are far apart. This enables the visualization of the time structure of species co-occurrences and reveals major disturbances in the history of life. Co-occurrence of fossil species was defined at relatively small time increments of 1 million years (Myr), enabling exploitation of the full temporal resolution of raw occurrence data (which aids the detection of evolutionary phenomena[3,13]). Sets of coexisting species are the fundamental constituents of any evolutionary biota, which may persist (to a greater or lesser extent) at one or more taxonomic levels[9,14–16]. A set of coexisting species is also the maximal set for possible ecological interactions, as co-occurrence in time is necessary (though not in itself sufficient[17]) for ecological interaction. Therefore, temporal co-occurrence probability also provides an evolutionarily (and therefore ecologically) meaningful distance measure between fossil species that facilitates analyses of the persistence versus decay of co-occurrences. The machine-learnt distances are then related to exhaustively calculated measures of species occurrence across time (shared species fraction between compared times) and proportionate extinction[2] versus origination[18]. In concert, these measures provide insights into the relative impacts and timing of extinction and radiation, independent of background trends in diversity (computer simulations, Extended Data Fig. 1b–g).

The analyses are based on global fossil occurrences (finds of fossil species from given times and geographical locations) publicly available in the Paleobiology Database (PBDB), comprising 1,273,254 occurrences for 171,231 species in the complete data set. After strict data screening to include only those occurrences classified to the species and phylum level, the data set included 665,590 occurrences for 137,779 species. The data set covers a broad taxonomic sample of 64 animal, plant and protist phyla and extends from the Neoproterozoic eon to the recent past, with unbroken Phanerozoic data coverage from 532 million years ago (Ma) in the Cambrian period to today (0 Ma).

These analyses permit new quantitative tests of both longstanding and new hypotheses in macroevolution, including (1) simultaneous comparison of the scale and pattern of macroevolutionary disruptions across the Phanerozoic fossil record, (2) quantitative assessment of the relative balance and timing of mass radiations and extinctions from 580 Ma to the present, (3) direct tests of the hypothesis of constant evolutionary decay[4] and (4) investigation of the corresponding impacts of extinction and radiation on macroevolutionary decay versus persistence.

Time structure of the fossil record

We initially visualized the temporal co-occurrence structure of the fossil record, as represented by our multidimensional machine-learnt spatial embedding, by using principal component analysis (PCA) to generate lower-dimensional projections from the full 16-dimensional embedding (Fig. 1). The spatial embedding method takes temporal co-occurrence structure, usually exclusively a property of groups of species[13,14,19], and translates it into an optimal embedding location for each individual species. This facilitates the simultaneous representation of the pattern of overlaps and separations between species time ranges in the fossil record (the time structure of species co-occurrences). Here, evolutionary restructuring events during the history of life are visible as shifts in species co-occurrence structure in spatial embedding projections to three, two or one dimensions (Fig. 1; PCA-explained variance: axis 1, 26%; axis 2, 15%; axis 3, 10%). By contrast, a simpler method applying PCA directly to vectors of species time occurrences recovers a coarse time

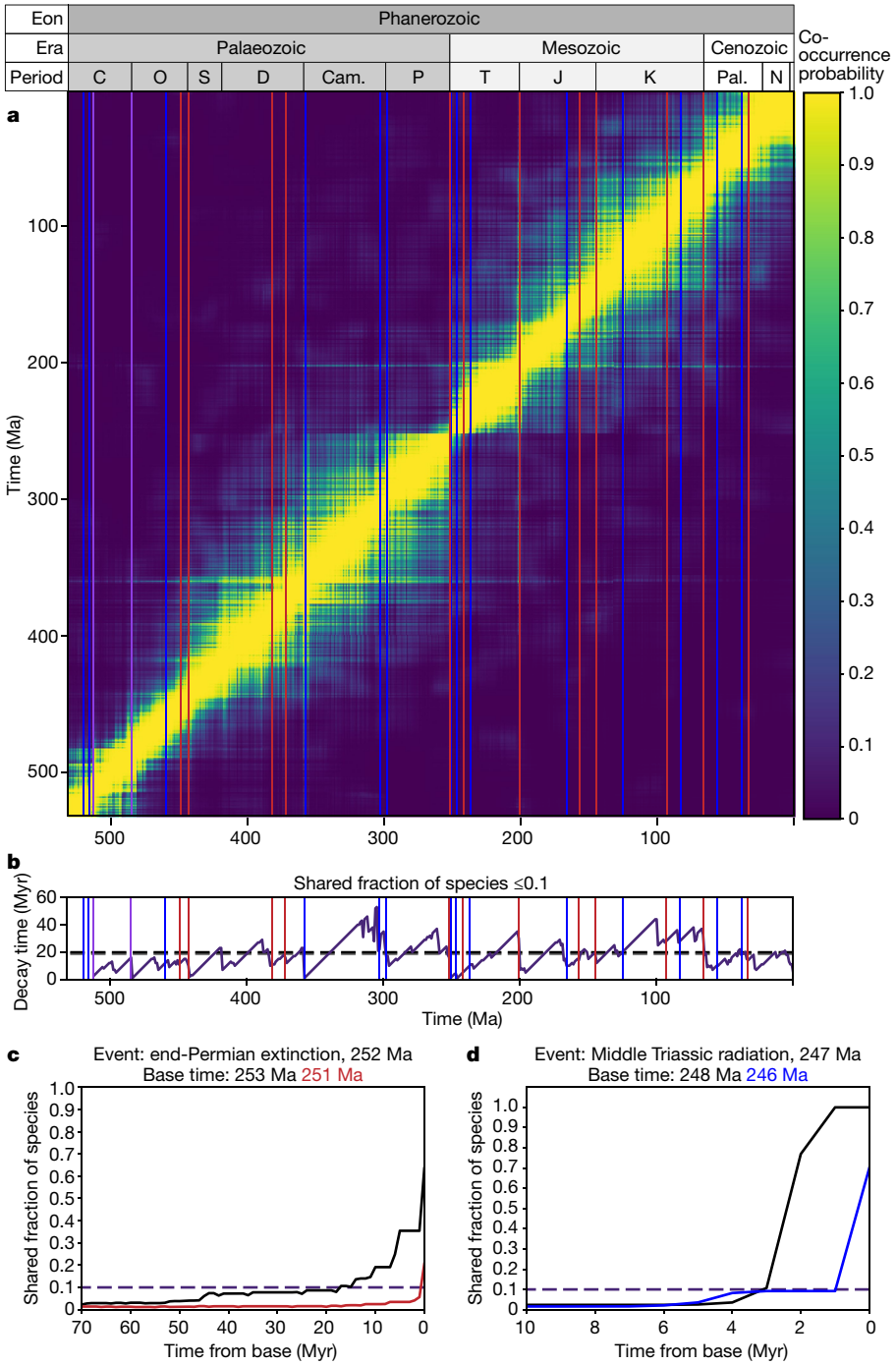


Fig. 2 | Macroevolutionary decay. a, Heatmap in which colour represents the mean probability of temporal co-occurrence between species occurring at compared times (complete data set, all pairwise time comparisons, 1-Myr increments, 531–0 Ma, $n = 532$ times) calculated from distances in the machine learning spatial embedding. **b**, Time to over-threshold evolutionary decay, when the fraction of species shared between a base time and its preceding times falls to 0.1 (taxonomically screened data set). Horizontal dashed lines indicate mean time to decay (grey) and maximum range among the 90% shortest species ranges (black). Vertical lines indicate 5% most significant mass extinctions (red), mass radiations (blue) and mixed mass extinction–radiation events (magenta) (Fig. 3, Extended Data Fig. 4). **c**, **d**, Examples of major disturbance events at which the rate of evolutionary decay rapidly increased: end-Permian mass extinction (**c**) and subsequent Middle Triassic mass radiation (**d**).

structure but not major evolutionary events (Supplementary Computer Code 5). In a test of the robustness of our embedding approach, 80% bootstrap data subsamples (Supplementary Computer Code 6) showed local stability of relative embedding positions across 18 retrained replicates (Extended Data Fig. 2a).

Marked effects on temporal co-occurrence structure are apparent during episodes of both diversification and extinction. For example, the end-Permian mass extinction (the 'great dying') corresponds to a major breakpoint in co-occurrence among species occurring before and after the boundary between the Palaeozoic and Mesozoic eras (red-to-blue transition, Fig. 1). All of our analyses recovered this end-Permian mass extinction as the most significant restructuring event in the continuous Phanerozoic fossil record and the most marked break with preceding times (Fig. 2a, Extended Data Fig. 2b–e), as further described below. However, major restructuring events are also identified during episodes of diversification[14].

Balance between radiation and extinction

Attempts to characterize macroevolution have often focused on mass extinctions and subsequent ecological replacements, including implicit causative hypotheses of creative destruction, which assume that large-scale radiations require preceding mass extinctions[1,3,5]. However, comparisons of proportionate origination[18] versus extinction[2] at 1-Myr increments through the Phanerozoic eon (Supplementary Computer Code 2) indicate that evolutionary destruction and creation have been almost perfectly balanced, with a full continuum of events occurring between these extremes (Fig. 3, Extended Data Fig. 3). All of the big five

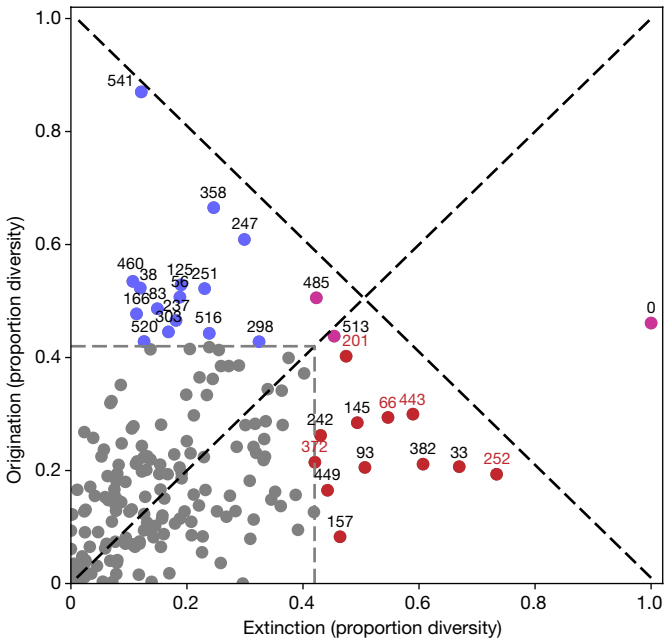


Fig. 3 | Balance between mass radiation and extinction. Species origination versus extinction, as a proportion of total diversity within the time window, at 1-Myr increments from 600 to 0 Ma. Data points: $n = 222$ times at which any species enter or exit the fossil record (taxonomically screened data set). Labelled times: 30 (5%) most significant event times from 600 to 0 Ma (corresponding to a species entry or exit threshold of >42%, grey square). Colours: magenta, both extinction and origination above threshold (mass extinction–radiation); red, extinction only (mass extinction; red labels denote the 'big five' mass extinction events[2]); blue, origination only (mass radiation).

mass extinction events previously identified based on drops in raw[2] or subsampled[20] diversity were among the 5% most significant times of evolutionary disruption identified here. However, among those most significant disruption times, we additionally identified seven other mass extinctions; fifteen comparable-scale diversifications, which we therefore call the mass radiations; and two combined mass extinction–radiation events (Fig. 3, Table 1). From either side of this continuum it is therefore possible to identify mirror events (which we also call, in reference to the Red Queen hypothesis[4], looking-glass events), defined as those showing the most closely reversed proportions of species entering or exiting the fossil record (Fig. 3, Table 1). For example, the most extreme mass radiation is the signal of the Cambrian explosion at 541 Ma, at which 87% of species enter the record and 12% leave. The closest mirror to this is the end-Permian mass extinction, which saw 73% extinction but also 19% origination within a 1-Myr window.

This analysis shows that the most comparable mass radiations and extinctions (for example, mirror events among the 5% most significant disruption times, Table 1) are in general temporally decoupled, strongly arguing against an immediate causal connection between them. In particular, the proportionately most extreme mass extinctions were, necessarily, not accompanied by a radiation of comparable scope within the same 1-Myr time window (Table 1). Nor are the mass extinctions generally observed to be closely followed by a mirroring mass radiation (Pearson correlation $r = 0.20$, $P = 0.295$, Shapiro–Wilk $W = 0.934$), which would be predicted by hypotheses of vacation of niches and direct replacement, for example[1,3,10]. Instead, the events in Phanerozoic history that have created proportionately the most diversity (including mass radiations at the beginning of the Cambrian, Carboniferous, Late Ordovician and early Cretaceous) have generally occurred at times that were widely separated from the mass extinction events (Table 1, Extended Data Fig. 4). The most extreme of these mass radiations are the Cambrian explosion (from 541 Ma)[18,21], in which species representing many animal phyla first appear, and the beginning of the Carboniferous period (358 Ma), in which a signal of major terrestrialization is evident in both plant and animal speciations (Extended Data Fig. 5). Therefore, the proportionately largest radiations arguably occurred not after ecological niches were vacated by extinctions of comparable scale[1,3,10] but when life exploited new realms of opportunity[10,18,21,22]. One notable exception to this temporal decoupling of mass extinction and radiation is the end-Permian mass extinction at 252 Ma, which was followed closely[2,18,20] by two significant radiation events at 251 and 247 Ma. Mapping of these mass turnover events, evident from proportionate extinction or origination, onto the visual output from our machine-learnt spatial embedding showed that they were associated with major shifts in species co-occurrence structure (Figs. 1 and 2, Extended Data Fig. 2c, e).

Macroevolutionary decay

Visualization of all possible time-to-time distances (Fig. 2) generally shows a trail of high, and then decaying, co-occurrence probabilities. This trail extends from a given base time back to those earlier times in which existing species remain comparatively closely located within our multidimensional spatial embedding. Its fall-off represents the process of macroevolutionary turnover, over which the probability of species co-occurrence falls to a very low level. Across the Phanerozoic, the exhaustively calculated fraction of fossil species shared between any two times (which is closely conceptually related to the co-occurrence probability, but here has the additional advantage of non-heuristic value calculation) falls below 0.1 in a mean of 18.6 Myr (taxonomically screened species data set, standard deviation (s.d.) = 9.84, median = 17 Myr). This decay rate results from the distribution of species occurrence times and ranges, which in aggregate comprise the fossil record (90% ranges ≤ 19.8 Myr, median = 6.5 Myr; additional summary statistics, Extended Data Fig. 6a, b). The fraction of species shared between times falls below 0.5 in a mean of 4.4 Myr (s.d. = 3.1); therefore, this represents the relative half-life of species occurrences. A lower threshold of 0.05 is reached at a mean of 30.6 Myr (s.d. = 14.9). For comparison against the shared fraction, the probability of species co-occurrence across compared times (calculated from the mean time-to-time embedding distance, Fig. 2) falls below 0.1 in a mean of 30.4 Ma for the complete data set, and similarly below 32.5 Myr after strict taxonomic screening. Therefore, on average for a time series, by approximately 19 Myr after it starts, proportionally very few to none of the species that exist will be those that were present at the beginning. Conversely, by this time the existing species will, on average, be entirely new.

Across the Phanerozoic as a whole, this time to over-threshold evolutionary decay fluctuates around an approximately constant mean (Fig. 2). This equilibrium level has been consistently returned to over Phanerozoic history despite secular diversity increases during this period[20] (from which our measures of co-occurrence structure are largely independent, Extended Data Figs. 1b–g, 6b, 7e). Based on constant extinction probability estimates for taxa of different ages, Van Valen predicted that the effective environment[4] (ecological[23] setting) of a given species would tend to deteriorate at a constant rate (the Red Queen hypothesis)[4]. The measures of species co-occurrence calculated here provide a direct estimate of the decay rate of macroevolutionary structure, which we call the decay clock. The decay clock counts the time to over-threshold evolutionary decay, which is here defined as the time (looking back from each base time, Fig. 2b–d) at which the shared fraction of species (or co-occurrence probability) approaches zero (specifically, falls to 0.1). As the global set of co-occurring species is the arena within which all ecological interactions must take place, the decay clock shows how this maximal ecological envelope decays or persists over time. Our results demonstrate that the global Phanerozoic biota has indeed decayed over an equilibrium average of 19 Myr (Fig. 2b). However, rather than remaining flat (as might be

Table 1 | Looking-glass events in macroevolution

Event						Mirror event			
Event rank	**Time (Ma)**	**Classification**	**Event unit**	**Extinctions (%)**	**Originations (%)**	**Time (Ma)**	**Classification**	**Extinctions (%)**	**Originations (%)**
1	541	Mass radiation	Cambrian start	12	87	252	Mass extinction	73	19
2	358		Carboniferous start	25	67	33		67	21
3	247		Middle Triassic start	30	61	443		59	30
4	460		Late Ordovician start*	11	53	157		46	8
5	125		Aptian stage start	19	53	93		51	21
6	38		Priabonian stage start	12	52	157		46	8
7	251		Triassic start	23	52	93		51	21
8	56		Eocene start	19	51	93		51	21
9	83		Campanian stage start	15	49	449		44	17
10	166		Callovian stage start	11	48	157		46	8
11	237		Late Triassic start	18	47	449		44	17
12	303		Gzhelian stage start*	17	45	449		44	17
13	516		Nangaoian stage start*	24	44	242		43	26
14	520		Atdabanian stage start*	13	43	449		44	17
15	298		Permian start	32	43	242		43	26
1	485	Mass extinction–radiation	Ordovician start	42	51	201	Mass extinction	47	40
2	513		Middle Cambrian start*	45	44	485	Mass extinction–radiation	42	51
1	252	Mass extinction	**Permian end**	73	19	358	Mass radiation	25	67
2	33		Eocene end	67	21	358		25	67
3	382		Middle Devonian end	61	21	358		25	67
4	443		**Ordovician end**	59	30	247		30	61
5	66		**Cretaceous end**	55	29	247		30	61
6	93		Cenomanian stage end	51	21	56		19	51
7	145		Jurassic end	49	28	251		23	52
8	201		**Triassic end**	47	40	485	Mass extinction–radiation	42	51
9	157		Oxfordian stage end	46	8	166	Mass radiation	11	48
10	449		Blackriveran stage end**	44	17	303		17	45
11	242		Anisian stage end*	43	26	516		24	44
12	372		**Late Devonian**	42	21	516		24	44

Top 5% fractional species turnover times (n = 29 event times, present 0 Ma excluded) in the Phanerozoic fossil record and their closest mirrors. Mirror events have opposite dominance of species origination versus extinction and closest reversed magnitudes (closest points in mirroring of Fig. 3 across the identity line). Bold rank numbers denote the 9 most extreme events (top 5% of 222 identified turnover events); bold names of events denote the 'big five' mass extinction events[2]. Relevant stratigraphic unit names, dates and corresponding references are those used in the PBDB: *ref. [30]; **refs. [31,32].

expected from a consideration only of the mean or maximum species range, Fig. 2b), we show that macroevolutionary decay is characterized by dynamic fluctuations around this long-term average as species co-occurrence structure is periodically disturbed and then gradually recovers continuity.

At times of major evolutionary disruption during the Phanerozoic (Fig. 3), the normal chains of species co-occurrences have been broken, leading to sudden discontinuities (Figs. 1 and 2), in which the probability that any existing species co-occurred with species from any preceding time fell to exceptionally low levels at an exceptionally rapid rate (Fig. 2). Notably, the great majority of species that have lived at any time from 251 Ma onwards did not occur before the end-Permian mass extinction, or co-occur with any species that existed in the preceding Palaeozoic era. Consequently, there was a considerable increase in the rate of macroevolutionary decay at the end of the Permian period (Fig. 2c, Extended Data Fig. 6c, d), with a drop to a shared species fraction of 0.1 by 1 Myr after this extinction event (reaching 0.1 before 253 Ma, 19 times faster than the Phanerozoic mean). As time goes on, after each such disturbance event, the decay-clock time can increase only gradually in each 1 Myr that >10% of a given biota has persisted. This highlights an inherent time asymmetry in macroevolutionary disturbance and recovery, in that the decay clock can be reset rapidly but can count up only year by year between disturbances. Comparatively long intervals between major disturbance events are therefore characterized by long-term persistence of evolutionary biotas (the flip side of evolutionary decay), such as those during the Carboniferous and mid-Cretaceous (Fig. 2).

The concept of evolutionary decay was originally formulated in relation to extinctions[4] (conceptual diagram, Extended Data Fig. 7a–c). Extinctions themselves erode a given community by removing original members[3]. However, we show that evolutionary radiations also cause comparable decay by diluting a pre-existing species set, thereby decreasing the co-occurrence probability and the fraction of species shared with times preceding a radiation event (Fig. 2, Extended Data Figs. 2b–e, 6e, f, 7a–c). In this sense, mass radiations (Fig. 3, Table 1) can be as destructive to existing species sets (and potentially, therefore, to the ecological communities within this maximal envelope) as major extinction events. Consequently, the decay clock has been periodically reset throughout Phanerozoic history by both extinctions and radiations (Fig. 2). Although this destructive aspect of evolutionary

radiation may initially appear counterintuitive (because radiations necessarily create new species), recent biogeography presents numerous examples of the major ecological disruptions that can result from the appearance within an existing community of new invasive species[10]. The analyses conducted here show that disturbances resulting from the evolution of new species have occurred periodically, sometimes on a huge scale, throughout Phanerozoic history (Fig. 2). Those species present at the onset of a mass radiation experienced influxes of new species generating up to 87% of total standing diversity (Fig. 3), with the most extreme example occurring at the Ediacaran–Cambrian transition. Mass radiations have therefore represented disruptions to the prior biota[9,14] at scales comparable to, and in some cases exceeding, those of the mass extinctions (Figs. 2 and 3).

There has been considerable interest in trends in diversity and extinction across Phanerozoic history, including effects of marine versus terrestrial settings[24], biotic[4] versus abiotic[24] extinction triggers, and trends[9,25] and periodicities[26,27] in extinction magnitude (all of which have been subject to scientific debate). Our analysis provides an overview of the relative dynamics of diversity over time that takes into account all events recorded by the pattern of species occurrences (not solely extinctions or their largest or best-known subset). Contrary to some previous results obtained using other measures of diversity or taxonomic levels (for example, number or percentage of families going extinct within a time interval[25,27,28]), the species-level measures, calculated here, do not show significant declines throughout the Phanerozoic either in the intensity of disruptions to co-occurrence structure or in the proportional origination or extinction levels (statistics, Extended Data Fig. 7e).

Three major disturbance events in the Eocene epoch of the Palaeogene period are particularly relevant to the establishment of the modern ecosystem, including two mass radiations at the start of the epoch and the latter Priabonian stage, as well as a mass extinction at the Eocene–Oligocene transition approximately 33 Ma (Fig. 3, Extended Data Fig. 4). Subsequently (although they fall outside the 5% most significant times of disturbance), events within the two most recent geological periods, the Neogene and Quaternary, show moderate to high levels of disturbance (Fig. 2; detail, Extended Data Fig. 7d), with fractional species turnover greater than 30% (within the top 11% of 600 analysed times and top 30% of 222 times of identified turnover, Extended Data Fig. 3). These events include radiations at approximately 28, 23 and 20 Ma (with originations ≥30%). They also include extinctions at approximately 15, 5 and 2 Ma associated with climate change at the end-Miocene (5.3 Ma) and Neogene–Quaternary transitions (2.58 Ma)[18,29], which, although moderate when compared against the entire scope of Phanerozoic history[2], are formidable from a modern conservation perspective[10] (with species extinction of ≥30%). Because macroevolutionary disturbances can reset the decay clock, these recent extinction events resulted in rapid evolutionary decay (Fig. 2; detail, Extended Data Fig. 7d). Consequently, diversity entered the Quaternary period with an already below-average decay-clock time of approximately 11 Myr. From that point, the decay clock would therefore be expected to take a minimum of 8 Myr, in the absence of large-scale disturbance, to count up to the Phanerozoic mean. Based on the historical processes identified here, modern extinctions and originations are likewise predicted to erase the connections to the past that are measured by the decay clock. Each modern extinction therefore represents a step towards macroevolutionary decay that further increases the time required to recover to the long-term equilibrium of species persistence

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-3003-4.

1. Simpson, G. G. *Tempo and Mode in Evolution* (Columbia Univ. Press, 1944).
2. Raup, D. M. The role of extinction in evolution. *Proc. Natl Acad. Sci. USA* **91**, 6758–6763 (1994).
3. Hull, P. M., Darroch, S. A. F. & Erwin, D. H. Rarity in mass extinctions and the future of ecosystems. *Nature* **528**, 345–351 (2015).
4. Van Valen, L. A new evolutionary law. *Evol. Theory* **1**, 1–30 (1973).
5. Jablonski, D. Extinctions: a paleontological perspective. *Science* **253**, 754–757 (1991).
6. Jablonski, D. Lessons from the past: evolutionary impacts of mass extinctions. *Proc. Natl Acad. Sci. USA* **98**, 5393–5398 (2001).
7. Budd, G. E. & Mann, R. P. History is written by the victors: the effect of the push of the past on the fossil record. *Evolution* **72**, 2276–2291 (2018).
8. Lehman, J. & Miikkulainen, R. Extinction events can accelerate evolution. *PLoS One* **10**, e0132886 (2015).
9. Sepkoski, J. J. A kinetic model of Phanerozoic taxonomic diversity. III. Post-Paleozoic families and mass extinctions. *Paleobiology* **10**, 246–267 (1984).
10. Stroud, J. T. & Losos, J. B. Ecological opportunity and adaptive radiation. *Annu. Rev. Ecol. Evol. Syst.* **47**, 507–532 (2016).
11. Field, D. J., Benito, J., Chen, A., Jagt, J. W. M. & Ksepka, D. T. Late Cretaceous neornithine from Europe illuminates the origins of crown birds. *Nature* **579**, 397–401 (2020).
12. Wood, R. et al. Integrated records of environmental change and evolution challenge the Cambrian Explosion. *Nat. Ecol. Evol.* **3**, 528–538 (2019).
13. Fan, J. X. et al. A high-resolution summary of Cambrian to Early Triassic marine invertebrate biodiversity. *Science* **367**, 272–277 (2020).
14. Muscente, A. D. et al. Quantifying ecological impacts of mass extinctions with network analysis of fossil communities. *Proc. Natl Acad. Sci. USA* **115**, 5217–5222 (2018).
15. Alroy, J. Are Sepkoski's evolutionary faunas dynamically coherent? *Evol. Ecol. Res.* **6**, 1–32 (2004).
16. Brett, C. E., Ivany, L. C. & Schopf, K. M. Coordinated stasis: an overview. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **127**, 1–20 (1996).
17. Blanchet, F. G., Cazelles, K. & Gravel, D. Co-occurrence is not evidence of ecological interactions. *Ecol. Lett.* **23**, 1050–1063 (2020).
18. Sepkoski, J. J., Jr. Rates of speciation in the fossil record. *Phil. Trans. R. Soc. Lond. B* **353**, 315–326 (1998).
19. Sadler, P. M. Quantitative biostratigraphy—achieving finer resolution in global correlation. *Annu. Rev. Earth Planet. Sci.* **32**, 187–213 (2004).
20. Alroy, J. et al. Phanerozoic trends in the global diversity of marine invertebrates. *Science* **321**, 97–100 (2008).
21. Na, L. & Kiessling, W. Diversity partitioning during the Cambrian radiation. *Proc. Natl Acad. Sci. USA* **112**, 4702–4706 (2015).
22. Kearsey, T. I. et al. The terrestrial landscapes of tetrapod evolution in earliest Carboniferous seasonal wetlands of SE Scotland. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **457**, 52–69 (2016).
23. Van Valen, L. Adaptive zones and the orders of mammals. *Evolution* **25**, 420–428 (1971).
24. Benton, M. J. The Red Queen and the Court Jester: species diversity and the role of biotic and abiotic factors through time. *Science* **323**, 728–732 (2009).
25. Newman, M. E. J. & Eble, G. J. Decline in extinction rates and scale invariance in the fossil record. *Paleobiology* **25**, 434–439 (1999).
26. Fischer, A. G. & Arthur, M. A. in *Deep Water Carbonate Environments* (eds Cook, H. E. & Enos, P. E.) 10–50 (Society of Economic Paleontologists and Mineralogists, 1977).
27. Raup, D. M. & Sepkoski, J. J. Jr. Periodicity of extinctions in the geologic past. *Proc. Natl Acad. Sci. USA* **81**, 801–805 (1984).
28. Gilinsky, N. L. Volatility and the Phanerozoic decline of background extinction intensity. *Paleobiology* **20**, 445–458 (1994).
29. Pimiento, C. et al. The Pliocene marine megafauna extinction and its impact on functional diversity. *Nat. Ecol. Evol.* **1**, 1100–1106 (2017).
30. Gradstein, F. M., Ogg, J. G. & Smith, A. G. *A Geologic Time Scale 2004* (Cambridge Univ. Press, 2004).
31. Ross, R. J., Adler, F. J., Amsden, T. W., Bergstrom, D. & Bergström, S. M. *The Ordovician System in the United States: Correlation Chart and Explanatory Note* (International Union of Geological Scientists, 1982).
32. Walker, J. D., Geissman, J. W., Bowring, S. A. & Babcock, L. E. The Geological Society of America geologic time scale. *Geol. Soc. Am. Bull.* **125**, 259–272 (2013).

Methods

Palaeobiological data

The raw data for our analyses were temporal occurrences of fossil species publicly recorded in the Paleobiology Database (PBDB). These raw data are time ranges (intervals in the geologic timescale[32]) at which a fossil taxon (for example, species) was observed to occur. A given taxon (for example, species) present in the database may therefore be represented by one, or more than one, observed occurrence at one, or more than one, time interval.

Recorded occurrences of fossil species, from the Neoproterozoic to the present, were downloaded from the PBDB using the temporal overlap interval of 1,000–0 Ma, with all default output plus taxonomic classification. Analyses were conducted at the fundamental taxonomic level of species to avoid the potential for complicating factors of taxonomic occupancy that may result from the use of higher taxonomic ranks[24,28,33]. PBDB data were therefore downloaded and analysed at two levels of resolution of the taxonomic classification[34], as follows. (1) A taxonomically more inclusive data set that used unique species names as the IDs for analyses but with PBDB taxonomic resolution set to genus. This allowed the inclusion of some fossil occurrence records that are only classified to the level of genus (for example, an identified name such as *Acaste* sp.). This gave a total of 1,273,254 fossil occurrences for 171,231 species. (2) A taxonomically more exclusive data set screened to include only occurrences with an accepted name classified to species rank and with a specified phylum name. This gave a total of 665,590 fossil occurrences for 137,779 species. More relaxed taxonomic restrictions therefore resulted in 48% more fossil occurrence data for machine learning, whereas more strict taxonomic restrictions ensured uniform classification to species and phylum level. Principal results were then compared between the two data sets to determine any effects from these different data-screening protocols. This comparison showed that the main results were similar for the two data sets. Specifically, the rank orders of the magnitude of evolutionary disruptions at 1-Myr intervals were shown to be significantly correlated between the two alternative data sets (Spearman's rank-order correlation: fraction of shared occurrences $r = 0.3755$, $P = 2.9752 \times 10^{-19}$; embedding distances $r = 0.0960$, $P = 0.0268$). The top 20% times of evolutionary restructuring identified were also found to have an overlap across the two data sets of 75% for the machine learning (ML) spatial-embedding method and 92% for fractional turnover. Therefore, results from both data sets are reported in the main text, with ML visualizations in the main figures showing the complete data set while additional results, for example shared fractions of strictly taxonomically screened species, are reported in the text and Extended Data figures.

We note that we have not attempted to further process the PBDB raw data to correct for any dating uncertainties or preservation bias (see, for example, ref. [29]). Future work—for example, focusing on specific events—might consider incorporating additional data processing steps. However, the events that we identify can be verified against previously recovered patterns of extinction and radiation[2,20,29], suggesting that at the level of our analysis any data inconsistencies have not been sufficient to obscure events of evolutionary interest.

For comparison with the new metrics generated in this study, standard diversity statistics were calculated using the PBDB Navigator. These were the number of genera and families sampled in geological-stage time bins.

Machine learning

A new ML spatial embedding method was applied to the raw data of recorded occurrences of 171,231 fossil species in time (ML methods summary figure, Extended Data Fig. 1a). Geographical coordinates of fossil finds, which are also present in the PBDB, were not used in our ML method. Our ML method embeds fossil species within a multidimensional space (with 16 dimensions) in which inter-species distance represents their probability of temporal co-occurrence (definition, equation (1) below). Co-occurrence for a given pair of fossil species was identified based on temporally overlapping observed occurrences, a standard criterion for coexistence in time[19]. This method thereby takes high-dimensional data (the temporal occurrences of species in the fossil record) and projects it into a low-dimensional space that aims to preserve key aspects of that high-dimensional data (specifically the probability of species co-occurrence). Our method falls within a wider class of ML embedding methods. Existing ML embedding methods include, for example, non-metric multidimensional scaling[35], *t*-distributed stochastic neighbour embedding (t-SNE)[36], the word2vec[37] algorithm that embeds words (in that case in a vector space) and triplet-trained neural networks[38,39]. ML embedding methods may use a variety of ML optimization methods (for example, here, gradient descent[40]) and specific optimization functions (here, co-occurrence probability) to place (ordinate) points (for example, here, representing fossil species) within a multidimensional space. Some such embedding methods may additionally be linked with neural network methods and/or data classification steps (for example, triplet networks[38,39]). However, we note that this is not necessarily the case, and the specific method used here is not a neural network method, nor does it involve data classification, or the learning of a trained model that aims to generalize to new data (and may therefore be subject to associated methodological problems such as model overfitting on the training data set[41]). Rather, the specific aim of the ML method used here is solely to embed all training data according to the specific optimization function used (co-occurrence probability). Therefore, the meaning of proximity within our embedding is easily interpretable (as co-occurrence probability) and comparable to exhaustively calculated measures (see brute-force methods below). This is in contrast to some other multidimensional ordination methods, including ML methods such as the word2vec algorithm[42], in which the reason for proximity within a constructed space may be difficult to interpret.

The dimensionality of the embedding space (16 dimensions) was arbitrarily chosen in order to project the high-dimensional raw data to a comparatively low number of dimensions (a basic aim of dimensionality reduction techniques), while allowing a sufficiently large number of dimensions for the capture of biologically interesting structure in the data.

The ML spatial embedding was generated using a Python program (Supplementary Computer Code 1) implementing the following procedure. Each fossil species (which can have multiple observed occurrences in the database) is given a 16-dimensional embedding x (which is randomly initialized). We train the embedding over 50,000 training iterations (epochs). Within each training epoch, we train the embedding via gradient descent on a succession of batches (a method used in many current ML applications to optimize model parameter values[40]). Each batch consists of 20,000 examples. An example is constructed by first picking a random time window. A random time window is selected rather than a random fossil occurrence because randomizing by time window normalizes for variations in diversity over time. After a time window has been selected, a random occurrence is picked (whose species has embedding x_1) from that time window. We then randomly select whether this example will be a co-occurrence (or non-co-occurrence), with 50% probability. If a co-occurrence has been selected, we select another random occurrence from that time window (whose species has embedding x_2). If a co-occurrence has not been selected, we pick another random time window, pick a random occurrence from that time window and ensure that it does not co-occur with x_1. We then calculate the Euclidean distance (d) between x_1 and x_2 and interpret that as a probabilistic prediction of co-occurrence:

$$\begin{aligned} p(x_1, x_2) &= \mathrm{sigmoid}(a - d(x_1, x_2)) \\ &= p(x_1, x_2) = 1/(1 + \exp(-(a - d(x_1, x_2)))) \end{aligned} \quad (1)$$

where a is a learned parameter of the model, observed during ML to be 11.994 for the complete data set (and 12.5998 for the taxonomically screened data set).

The learnt parameter a can then be entered into equation (1) to convert a learnt embedding distance d to a corresponding co-occurrence probability.

We train the embeddings and the parameter a to minimize the binary cross entropy:

$$L = E[-y \log(p(x_1, x_2)) - (1-y)\log(1-p(x_1, x_2))] \quad (2)$$

where p is the probability assigned by the model that the two given species co-occur and y is the ground-truth label (1 when the species co-occur and 0 when they do not).

We used the Adam optimizer with a learning rate of 10^{-2} for 50,000 batches.

The length of the ML training time (measured in number of training epochs) for each data set (real or simulated) was assessed visually and statistically using visualization tools provided in Supplementary Computer Code 1–3. These tools allow visualization of the training error as training proceeds, PCA visualization of the output embedding and statistical assessment (by visualization and Pearson correlation) of behaviour of the embedding under simulated secular increases in diversity (linear or exponential).

Comparison of ML spatial embedding to pre-existing methods

This method of ML spatial embedding has some commonalities with previous methods for analysing biological abundance, diversity and temporal co-occurrence, including co-occurrence diversity assessment[13,19] and network analysis[14] (for example, utilization of species co-occurrence information), as well as non-metric multidimensional scaling[35] (for example, representation of inter-species variation within multidimensional spaces), but it has additional advantages for evolutionary analyses over time. These methodological advantages include (i) the meaning of inter-taxon distances (probability of species co-occurrence); (ii) consequent opportunities to perform new quantitative tests of macroevolutionary hypotheses; (iii) provision of human-readable data visualizations, facilitating new data-driven insights; (iv) robustness to potential problems of data sampling, crucially including secular variations in fossil preservation potential through time (which show complex relationships with palaeo-diversity that may impact detection or interpretation of evolutionary trends[43]), and (v) capacity to analyse macroevolutionary structure across continuous time series at any specified time increment (for example, 1 Myr). This is in contrast to, for example, standard within-bin diversity counting in comparatively large, discrete time bins (for example, geological stages, which are on the order of tens of millions of years in length), for which increasing bin size is known to impact detection of evolutionary phenomena[13].

Comparison of ML spatial embedding to alternative methods

For comparison to the ML embedding method (described above), a simpler method was implemented (Supplementary Computer Code 5) that applied PCA directly to vectors of the times at which fossil species were observed to occur. This method first takes the raw fossil occurrence data and encodes this as an array of time vectors. Here, each species has one vector of times at which it is recorded to occur (1) or not occur (0) according to the raw observed occurrences. The method then applies a PCA directly to these time vectors so that each fossil species is placed into a PCA projection with 16 components (comparable with our main ML embedding method, which uses 16 dimensions for the embedding space). Graphical output and code to generate this are provided as Supplementary Computer Code 5.

Validation of ML fossil embeddings

The measures of macroevolutionary disruption used in this study were designed to be independent of background trends in diversity (which have themselves been extensively investigated using other methods such as raw diversity analysis[2] and diversity subsampling[20]). The measures used here are therefore normalized for diversity. Diversity normalization is performed for the exhaustively calculated shared fraction of species between times by using overall diversity as the denominator (see Methods section below for further details). Diversity normalization was also incorporated into the ML spatial embedding method, for example by initially sampling data from times rather than species to avoid excessive weight from high diversity times. However, variation in diversity through time might potentially have unforeseen impacts on the ML process and outputs, which are in general highly data driven. Therefore, in order to validate our ML methods for further evolutionary analyses, we used computer simulations to test the sensitivity of the generated measures to changes in co-occurrence structure versus secular variation in diversity (Supplementary Computer Code 3). We show, using computer simulated data with a known distribution (linear or exponential diversity increase, Extended Data Fig. 1b–g), that co-occurrence-based spatial embedding allows the generation of comparative measures that are sensitive to shifts in species co-occurrence but are comparatively unaffected by background trends in diversity (which could themselves occur due either to genuine changes in biodiversity or to sampling variation). Specifically, given appropriate ML training time, Pearson correlation indicated no significant correlation between a simulated linear diversity increase and the mean embedding distance between species simulated at successive times ($r = 0.1311$, $P = 0.1936$, Extended Data Fig. 1b–d). A simulated exponential diversity increase produced a weak, although significant, negative trend across successive times ($r = -0.2761$, $P = 7.58 \times 10^{-5}$, Extended Data Fig. 1e–g), which can be removed by subtraction of the mean embedding path.

Additional exhaustive calculations of the shared fraction of fossil species between time windows facilitated further validation of, and comparison with, the ML spatial embeddings (Extended Data Fig. 2b–e), as well as additional evolutionary analyses. Bootstrap analyses (Supplementary Computer Code 6, details below) were used to test whether the ML methods were methodologically and statistically robust across multiple subsamples of the fossil occurrence data set (given its size and properties).

Brute-force computations of co-occurrence

For comparison with the ML spatial embedding distances, measures of proportionate species co-occurrence between times were calculated using a brute-force algorithm (Supplementary Computer Code 2), implementing the following procedure. For each time t_1, make an array of species occurrences at that time t_1. In this case, a given species is considered present at a given time t if t is within the time range of fossil occurrences of that species observed in the database ($t \geq t_{min}$ and $t \leq t_{max}$, where t_{min} is the minimum observed age of occurrence of the species and t_{max} is the maximum). For a compared time t_2, make an array of species occurrences. Calculate the fraction of occurrences that are shared between t_1 and t_2 (shared fraction = intersection/union). The fraction of species that were different is then calculated as the fractional symmetric difference = symmetric difference/union or 1 – shared fraction. If two compared times have exactly the same set of species existing, the shared fraction of species will equal 1. If either originations or extinctions occur, causing sets of species to differ between two compared times, the shared fraction of species between these times will fall. If the sets of species occurring at two compared times are entirely different, the shared fraction of species between times will equal zero.

The fraction of fossil species shared between any two times is closely conceptually related to the co-occurrence probability: both measure the extent and pattern of temporal co-occurrence (between times or between species across time, respectively), but they provide complementary advantages, respectively for the simultaneous visualization of co-occurrence structure (spatial embedding) versus the exhaustive calculation and simplicity of interpretation (shared fractions).

Drill plots and turnover event thresholding

Proportions of species originating versus going extinct at 1-Myr time increments were calculated and plotted (Fig. 3, Extended Data Fig. 4) using a Python program (Supplementary Computer Code 4). We present a new type of plot that we call drill plots (Extended Data Fig. 4) for focal times. These compare stratigraphic ranges of all species occurring within a 1-Myr time window from the focal time, vertically sorted into originations, extinctions and crossing ranges. Comparisons of event types in these analyses use threshold-based classification into three types: mass extinctions, mass radiations and mixed mass extinction–radiations. To classify events, the analyses first identify all turnover times, meaning times at which there are any speciations or extinctions observed in the data set, within 1 Myr (≤0.99 Myr) of the considered time (Supplementary Computer Code 4). We then calculate the proportions of the occurring species, within this time window, which are originating or going extinct. Each turnover event is then classified as to whether a selected threshold is exceeded by the proportion of extinctions only (in which case it is therefore classified as a mass extinction), radiations only (classified as a mass radiation) or both extinctions and radiations (classified as a mixed mass extinction–radiation).

The identification of turnover events in these analyses is therefore invariant to the entry/exit threshold used. What can potentially change with an increased threshold is the classification of these events as either a mass extinction, a mass radiation or a mixed event. Figures 3 and Extended Data Fig. 4 use a species entry/exit threshold of 42%, which was selected in order to highlight the most extreme 5% of turnover times, defined as the top 5% of the 600 times included in this analysis. 5% of the 600 included times equals 30, and the corresponding species entry/exit threshold of 42% is required to return 30 most extreme fractional turnover times. For comparison, Extended Data Fig. 3 shows a lower species entry/exit threshold of 30% which highlights a greater number of turnover times. This 30% threshold was selected as notable based on observation of the data, as this is the level above which all observed turnover events involved both extinction and origination. Choosing a higher entry/exit threshold (for example, >42%) for included times corresponds to reading off higher extinction/origination percentages from Fig. 3 to restrict consideration to a smaller number of turnover times. For example, another interesting threshold is the top 5% of the 222 identified times of turnover (out of 600 total times included in this analysis). This equals 11 times, which requires a 53% entry/exit threshold and returns the 10 most extreme times shown on Fig. 3 (with event classification unchanged except for 0 Ma, which does not pass the 53% entry threshold). A 52% entry/exit threshold returns the 13 most extreme times shown on Fig. 3.

Mirror (or looking-glass) events were identified, among the events classified using the extinction/origination threshold procedure described above. First, those events with % origination > % extinction were mirrored over the identity line (for example on Fig. 3, where % extinction = % origination), by temporarily swapping the x and y axes. The closest mirror events were then identified as those events from opposite halves of the original distribution that had the lowest Euclidean distance after mirroring. These mirror events are, therefore, those that are most comparable in scale but with opposite dominance of radiation versus extinction.

Comparison of brute-force co-occurrence measures to pre-existing methods

The shared fraction of fossil species between compared times (shared fraction = intersection(t_1,t_2)/union(t_1,t_2)) can be conceptually related (Extended Data Fig. 7a–c) to the fraction of surviving species (for example, survivor fraction = intersection(t_1,t_2)/t_1), a core concept of standard survivor analyses (for example, see ref. [4]). The main advantage, for the purposes of this study, of the co-occurrence measures used here (such as shared species fraction) is that these measures pick up the effect of any new species originations that have occurred between two compared times. This facilitates the comparison of the parallel effects of extinction and radiation within a unified measurement framework. It also facilitates time-symmetric comparisons: for example, measurement of the drop-off in shared fraction of species looking back in time from a given start time or event (Fig. 2c, d). More broadly, the shared species fraction between times also links mathematically to the ecological concept of spatial beta diversity (with beta diversity measures usually considering variation in species composition between spatial samples[44]).

Decay-clock calculations

The time-to-time average species co-occurrence probabilities from the ML analyses and exhaustively calculated fractions of species shared between times were each used to calculate the time to over-threshold decay in species co-occurrence (Supplementary Computer Code 2). For the time range in which there was continuous occurrence data in the data sets (0–532 Ma), this time to evolutionary decay was calculated for each base time, at 1-Myr increments, looking backwards in time, as follows. First, for each base time, a time series was considered that included all greater times within the total time range for this analysis (for example, for base time 252 Ma, the considered time series would be 253–532 Ma). Then, the values of the ML co-occurrence probability and fraction of shared species were extracted that compared the given base time to each time in the compared time series. The time taken, along the given time series, for co-occurrence to decay to the threshold value was then recorded. This is counted as the time vector position, such that a decay-clock time of 1 means that over-threshold decay has occurred after 1 and within 2 Myr. The mean of this decay value was then reported (as the average decay-clock time) across the considered times (0–532 Ma). A number of thresholds were used in this calculation. The main analyses use a decay threshold of 0.1, corresponding to ≤10% species shared between considered times. This threshold value of 0.1 was selected because it is a low-level cut-off that remains comparatively representative of species in aggregate (and so will not be driven, for example, by long-lived singleton species, as a cut-off of zero might be). For comparison, a threshold of 0.5 was also used, which represents a half-life for species co-occurrence, as well as a lower threshold of 0.05.

To give a worked example of the decay-clock calculation, consider base time 251 Ma (immediately after the end-Permian mass extinction at approximately 252 Ma). For the next few compared times, the fractions of species shared with the base time 251 Ma are for 251 Ma (identity), 1; 252 Ma, 0.21; 253 Ma, 0.06. For a threshold of 0.1, the decay-clock time for 251 Ma is therefore reported as 1 Myr because by 253 Ma (that is, within 2 Myr), fewer than 10% of species are shared with 251 Ma.

Geographical range of the analyses

Our analyses use all global fossil occurrences recorded in the PBDB and evaluate temporal co-occurrence only (equation (1)). Although it would be theoretically possible to extend our ML method to consider geographical locations (within an extended definition of co-occurrence), consideration of time alone has a number of advantages in the context of the present study. First, the examination of patterns of decay in co-occurrence through time has not previously been done, whereas ecological patterns in spatial structure have been extensively studied for example[21]. Second, by defining co-occurrence based solely on time (and not geographical location), we retain a close conceptual connection between our new ML distance measures and exhaustively calculated statistics on the proportion of species shared across times (as described above), which aids the validation and interpretation of the ML. Third, focusing purely on time provides an additional mathematical connection from these new statistics (machine learnt and exhaustively calculated) to fundamental measures of species survival (as described above and shown in Extended Data Fig. 7a–c).

Bootstrap analyses

To test whether the ML methods were methodologically and statistically robust across subsamples of the fossil occurrence data set, a bootstrap

procedure was implemented (Supplementary Computer Code 6). The ML embedding analysis was repeated over 18 bootstrap (technical) replicates (with an embedding run-time of 3 days on a GPU computer cluster), each sampling 80% of the 171,231 species from the complete data set. To analyse the stability of the embeddings across ML retraining on these bootstrap data samples, 60 reference fossils were randomly selected for comparison of embedding positions across the bootstrap replicates. These reference fossils were organized into triplets, each of which contained three members designated A, B and C. The distances in each learnt embedding between fossils A,B and A,C within each triplet were then compared across bootstrap replicates, using the mean differences and ratios between these distances and their standard deviations. In order to select reference fossils, 20 reference times were first randomly sampled from the total range of times (at 1-Myr increments) at which fossils were observed to occur in the complete data set. Reference fossils were sampled such that all three members of a given triplet were observed to occur within 30 Myr of a given reference time. This sampling process was used to ensure that compared fossils within a triplet occurred, relative to each other, within the time range over which the main analyses indicated an average co-occurrence probability above zero (with mean decay to co-occurrence probability ≤0.1 observed by 30 Myr for the complete data set). This is the approximate time range (average observed for the complete data set) over which we expect embedding distances to be comparatively tightly constrained by observed co-occurrences.

Statistical and visualization analyses

Further visualizations and statistical analyses were produced using the ML embedding distances and exhaustively calculated measures of species co-occurrence. Embedding distances and shared species fractions were compared between successive times at 1-Myr increments for the time interval over which there was continuous data coverage within the fossil occurrence data set (from 532 Ma, with numbers of species per time window of 5 Myr for the complete data set and 1 Myr for the strictly taxonomically screened data set). Time-to-time comparisons were conducted for all possible pairwise combinations of time windows of 1-Myr duration. Here, as above, the occurrence time for each species was summarized as the time-range midpoint across observed occurrences in the database.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Raw data are publicly available in the Paleobiology Database at https://paleobiodb.org. Additional source data for Figs. 1–3 are provided in the Dryad data repository (https://doi.org/10.5061/dryad.b8gtht79t). Additional data are provided as Extended Data Figs. 1–7.

Code availability

Custom computer code is provided as Supplementary Computer Code Code 1–6.

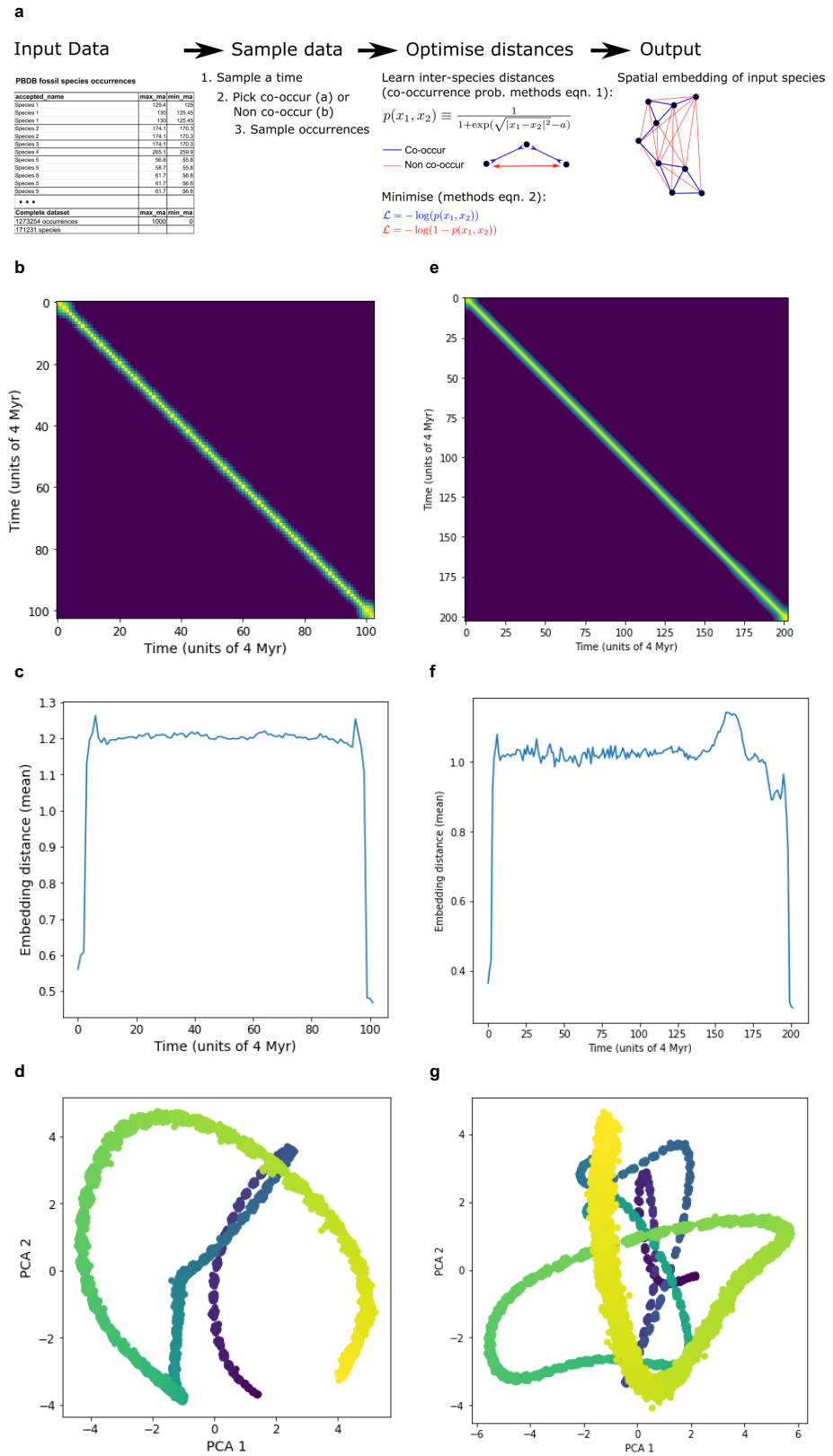
33. Gilinsky, N. L. & Bambach, R. K. Asymmetrical patterns of origination and extinction in higher taxa. *Paleobiology* **13**, 427–445 (1987).
34. Peters, S. E. & McClennen, M. The Paleobiology Database application programming interface. *Paleobiology* **42**, 1–7 (2016).
35. Caswell, B. A. & Frid, C. L. J. Learning from the past: functional ecology of marine benthos during eight million years of aperiodic hypoxia, lessons from the Late Jurassic. *Oikos* **122**, 1687–1699 (2013).
36. van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
37. Mikolov, T., Chen, K., Corrado, G. S. & Dean, J. A. Efficient estimation of word representations in vector space.. Preprint at https://arxiv.org/abs/1301.3781 (2013).
38. Schroff, F., Kalenichenko, D. & Philbin, J. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 815–823 (Institute of Electrical and Electronics, 2015).
39. Hoyal Cuthill, J. F., Guttenberg, N., Ledger, S., Crowther, R. & Huertas, B. Deep learning on butterfly phenotypes tests evolution's oldest mathematical model. *Sci. Adv.* **5**, eaaw4967 (2019).
40. Ruder, S. An overview of gradient descent optimization algorithms. Preprint at https://arxiv.org/abs/1609.04747 (2017).
41. Dietterich, T. Overfitting and undercomputing in machine learning. *ACM Comput. Surv.* **27**, 326–327 (1995).
42. Goldberg, Y. & Levy, O. word2vec explained: deriving Mikolov et al.'s negative-sampling word-embedding method. Preprint at https://arxiv.org/abs/1402.3722 (2014).
43. Heim, N. A. & Peters, S. E. Covariation in macrostratigraphic and macroevolutionary patterns in the marine record of North America. *Geol. Soc. Am. Bull.* **123**, 620–630 (2011).
44. Bacaro, G. & Ricotta, C. A spatially explicit measure of beta diversity. *Community Ecol.* **8**, 41–46 (2007).

Acknowledgements This research was supported by funding from an IADS Research Fellowship (J.F.H.C.), EON Research Fellowship (J.F.H.C.) at the Tokyo Institute of Technology (supported by a grant from the John Templeton Foundation), Earth–Life Science Institute Research Interactions Committee Visitor Fund (N.G. and J.F.H.C.) and Swedish Research Council (VR grant no. 2015-04726, G.E.B.). We thank S. Newman and L. Schalkwyk for computing time and support, S. Conway Morris and E. Mitchell for highly constructive comments on the manuscript.

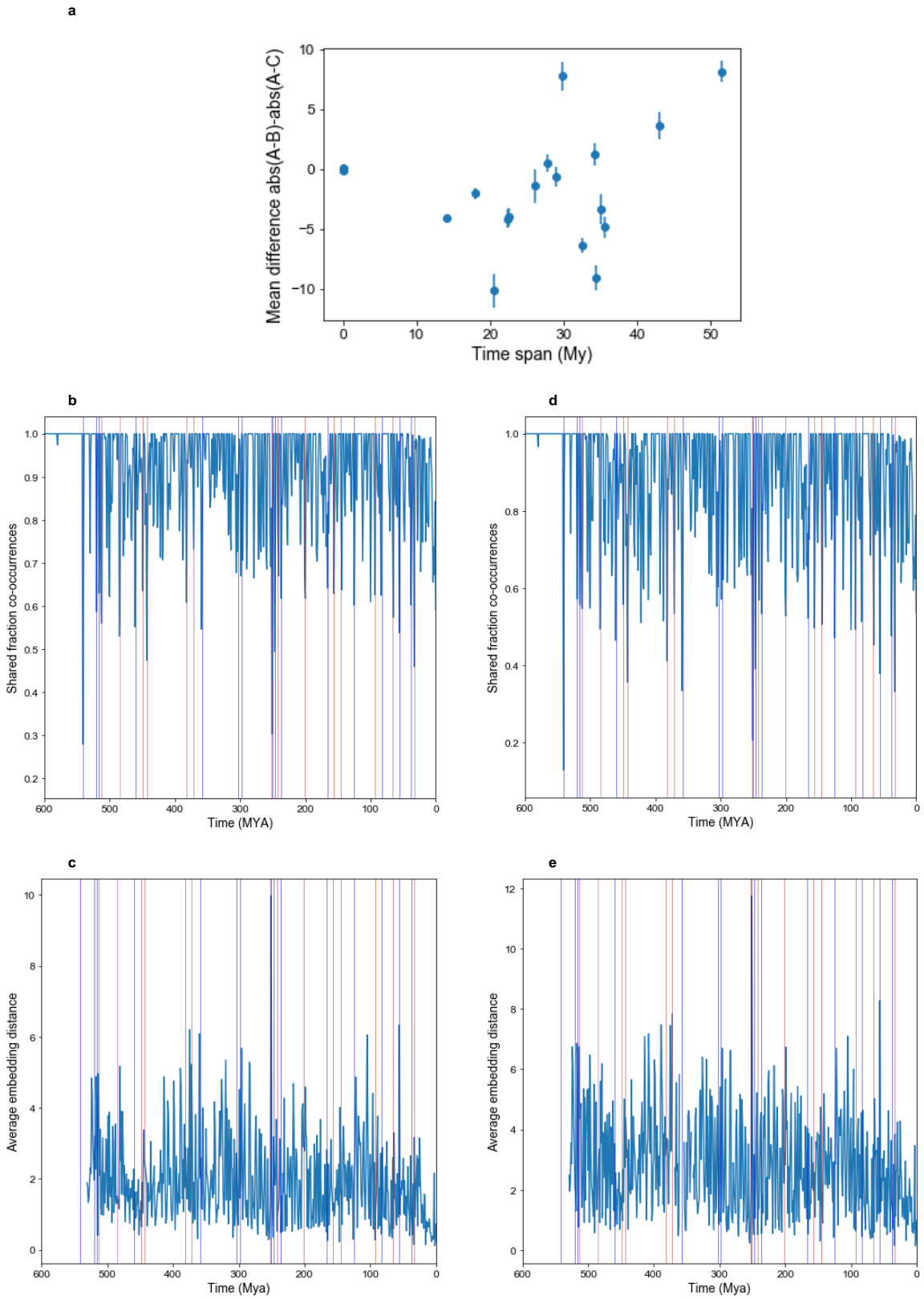
Author contributions J.F.H.C., N.G and G.E.B. designed research, N.G. and J.F.H.C. wrote computer code and performed analyses and J.F.H.C. wrote the paper with input from all authors.

Competing interests The authors declare no competing interests.

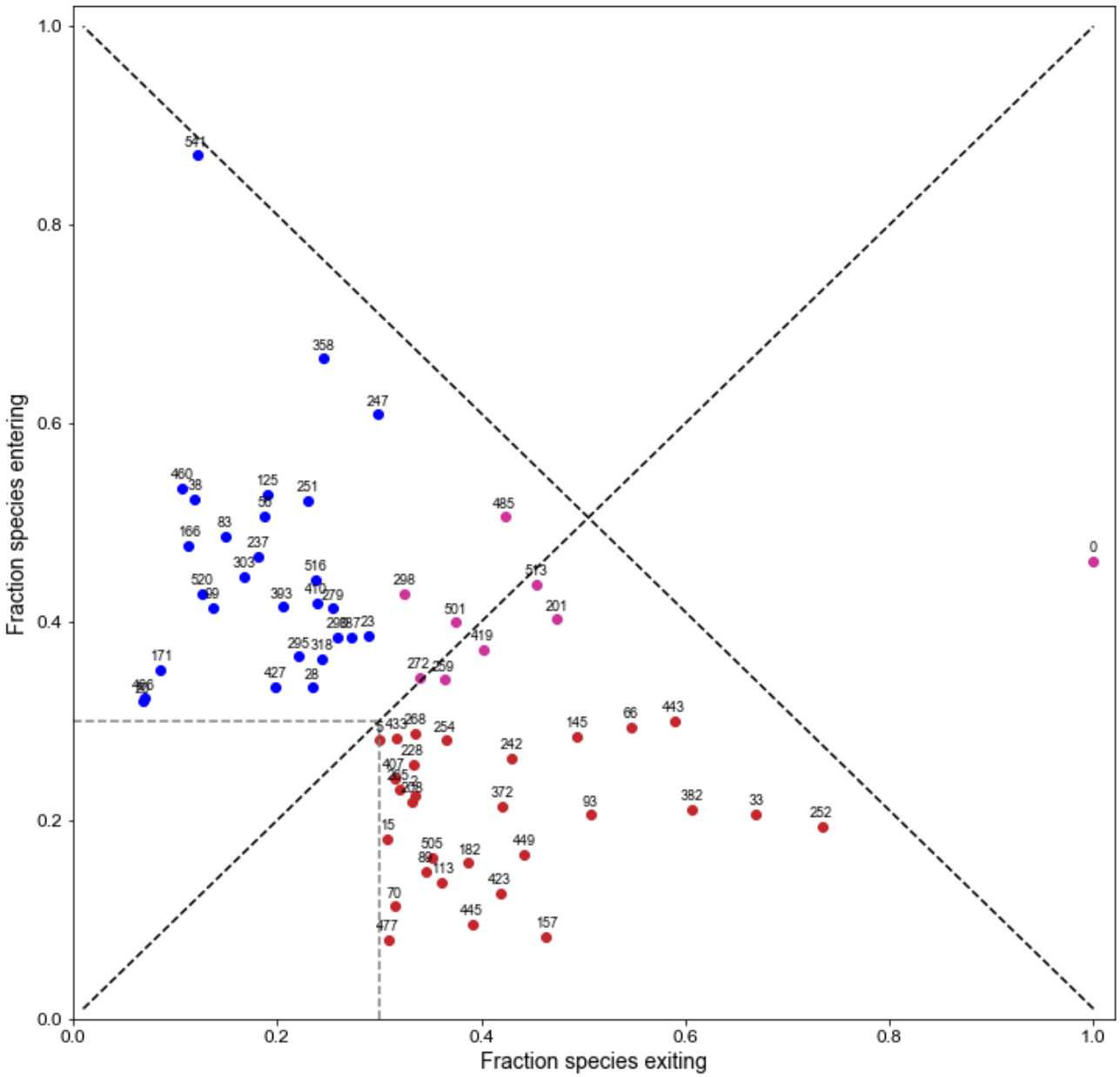
Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-3003-4.
Correspondence and requests for materials should be addressed to J.F.H.C.
Peer review information *Nature* thanks Emily Mitchell and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
Reprints and permissions information is available at http://www.nature.com/reprints.



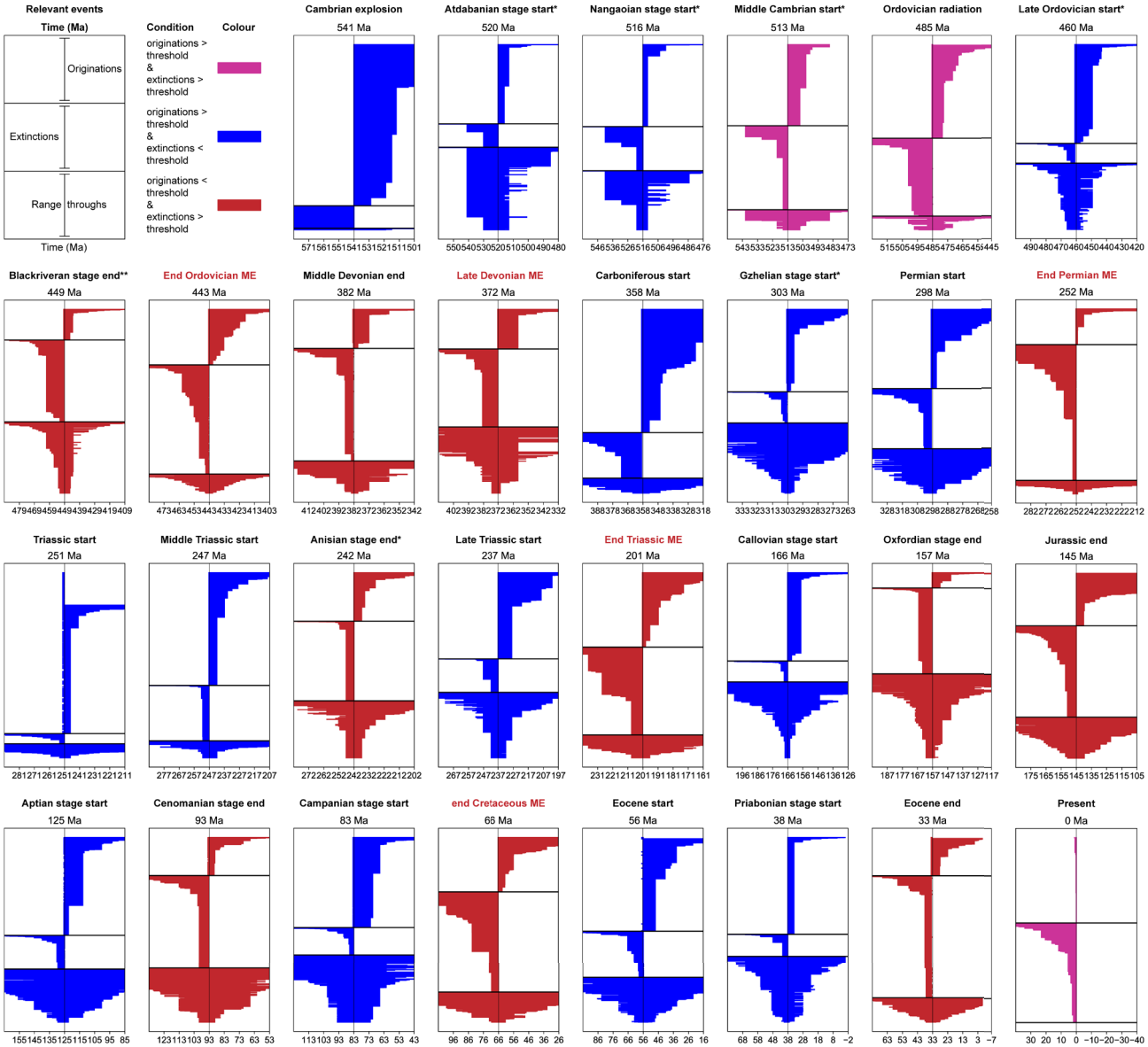
Extended Data Fig. 1 | Methods graphical summary and effects of computer simulated diversity increases. a, Graphical summary of the ML method. **b**–**g**, Computer simulations of secular variation in diversity, testing effects on measures of co-occurrence structure used in this study. **b**–**g**, Linear (**b**–**d**) and exponential (**e**–**g**) diversity increases (Supplementary Computer Code 3). **b**, **e**, Heatmaps visualizing the machine learnt spatial embedding distance between mean species locations at different times: yellow, closest; purple, farthest. **c**, **f**, Plot of embedding distances between successive times. **d**, **g**, Plot of first two principal component axes from the 16-dimensional spatial embedding. ML training times were 3,000 training epochs.



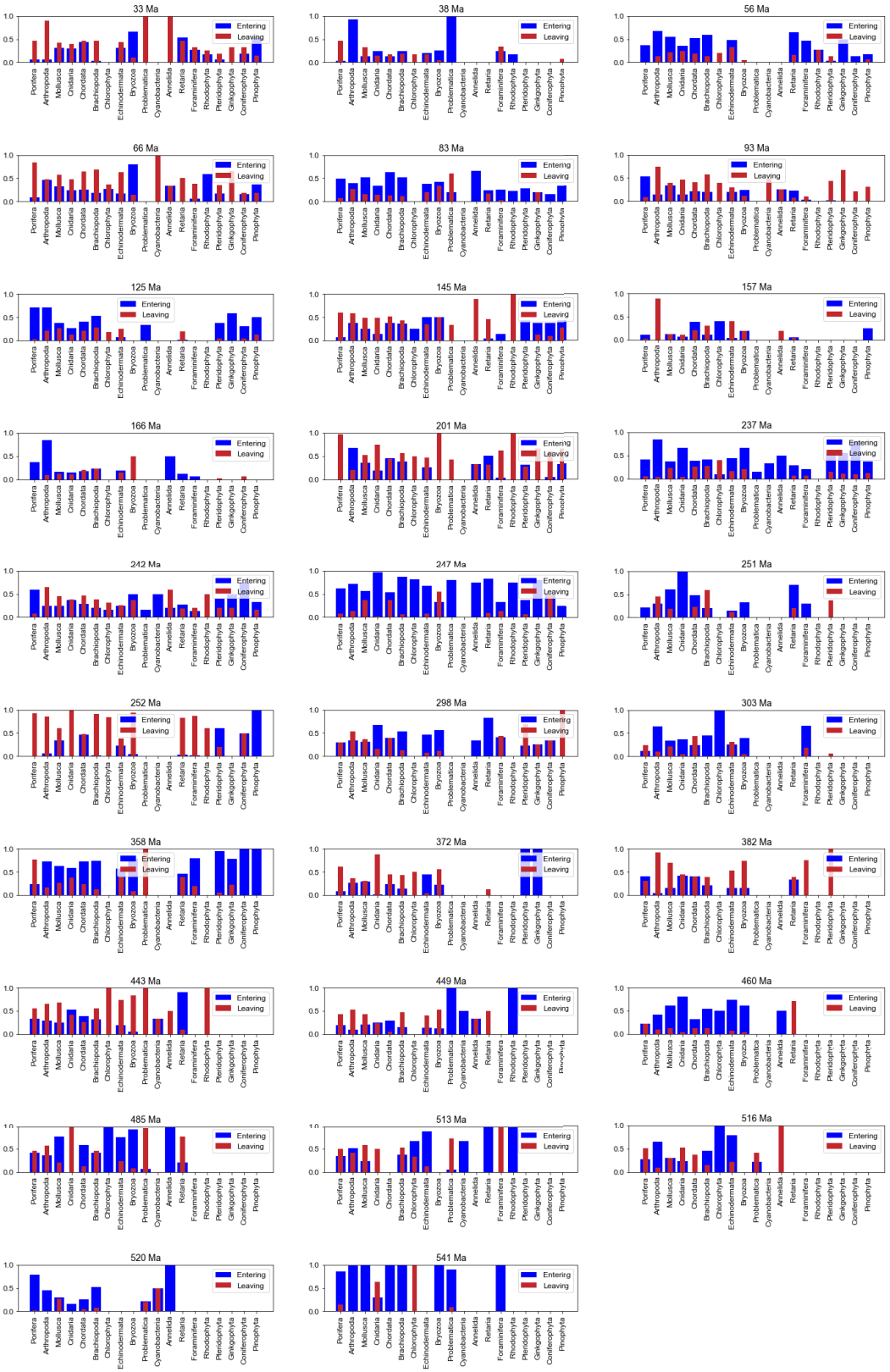
Extended Data Fig. 2 | Bootstrap data-resampling results and shared fraction of species between successive times versus mean embedding distance. a, Bootstrap data-resampling results: differences in embedding distances for 60 reference fossils, compared within 20 A, B, C triplets over 18 technical replicates of bootstrap data re-sampling and ML embedding training. Error bars, s.d. of the distance absolute(A–B) – absolute(A–C): mean 0.77. We expect the embedding distances to be comparatively stable within the time range over which co-occurrence probability is within the evolutionary decay range (observed to be mean 30 Myr for co-occurrence probability to reach 0.1 in the complete data set). **b**, **d**, Fraction of species shared between successive times, calculated exhaustively from raw species time ranges (histogram, Extended Data Fig. 6a). **c**, **e**, Distance in the ML spatial embedding between mean species locations at successive times. Compared times are at increments of 1 Myr. **b**, **c**, Complete fossil occurrence data set. **d**, **e**, Taxonomically screened data set. Vertical lines indicate the 5% most significant times of fractional species turnover (Fig. 3, Extended Data Fig. 4): mass extinctions (red), mass radiations (blue) and mixed mass extinction–radiations (magenta).



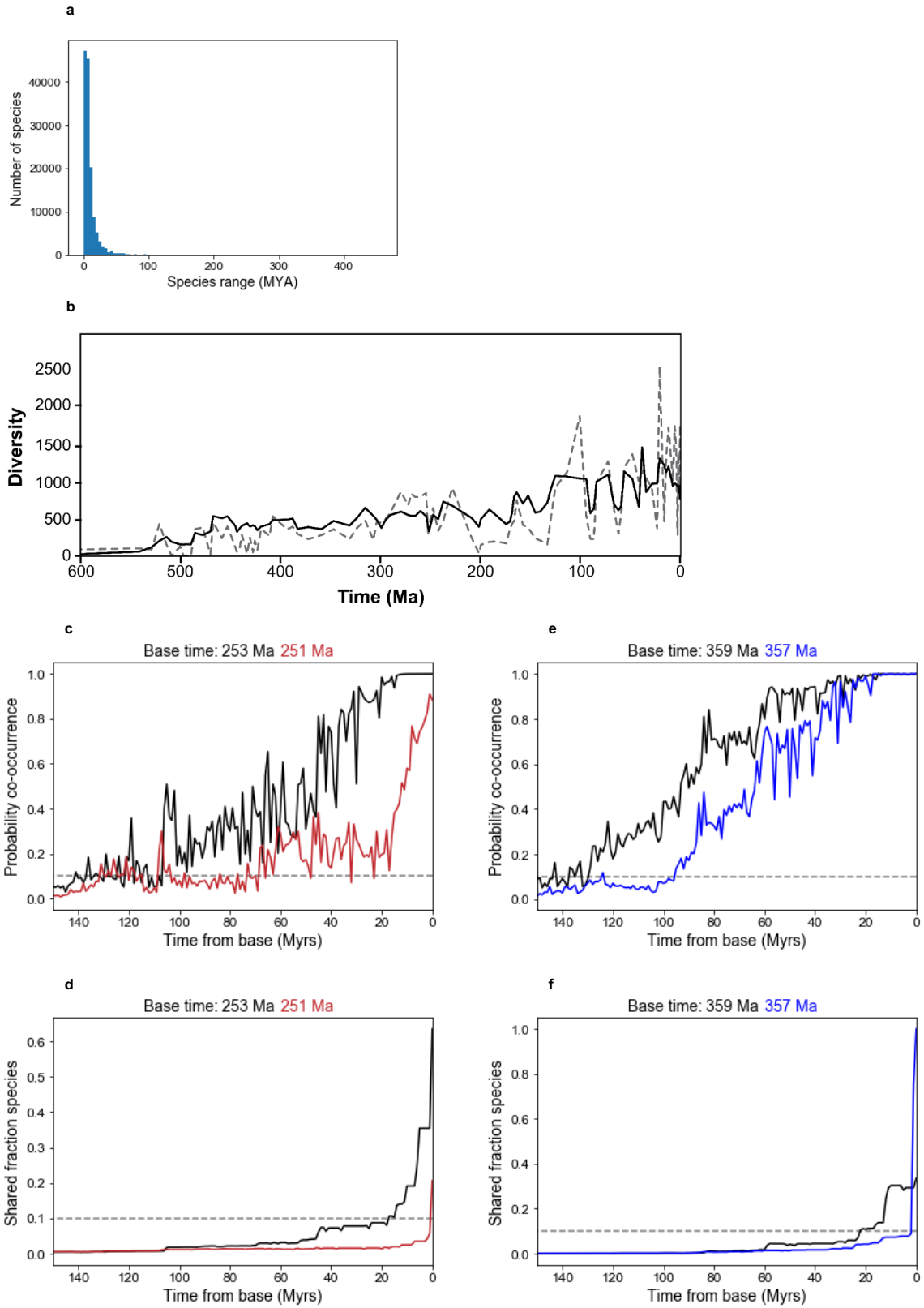
Extended Data Fig. 3 | Proportions of species originating versus going extinct. 1-Myr increments from 600 to 0 Ma with a threshold of 30% species entry/exit threshold, grey square. This threshold highlights the top 66 times of turnover from 222 total turnover times identified among 600 times included in the analysis. Red, mass extinctions; blue, mass radiations; magenta, mixed mass extinction–radiations.



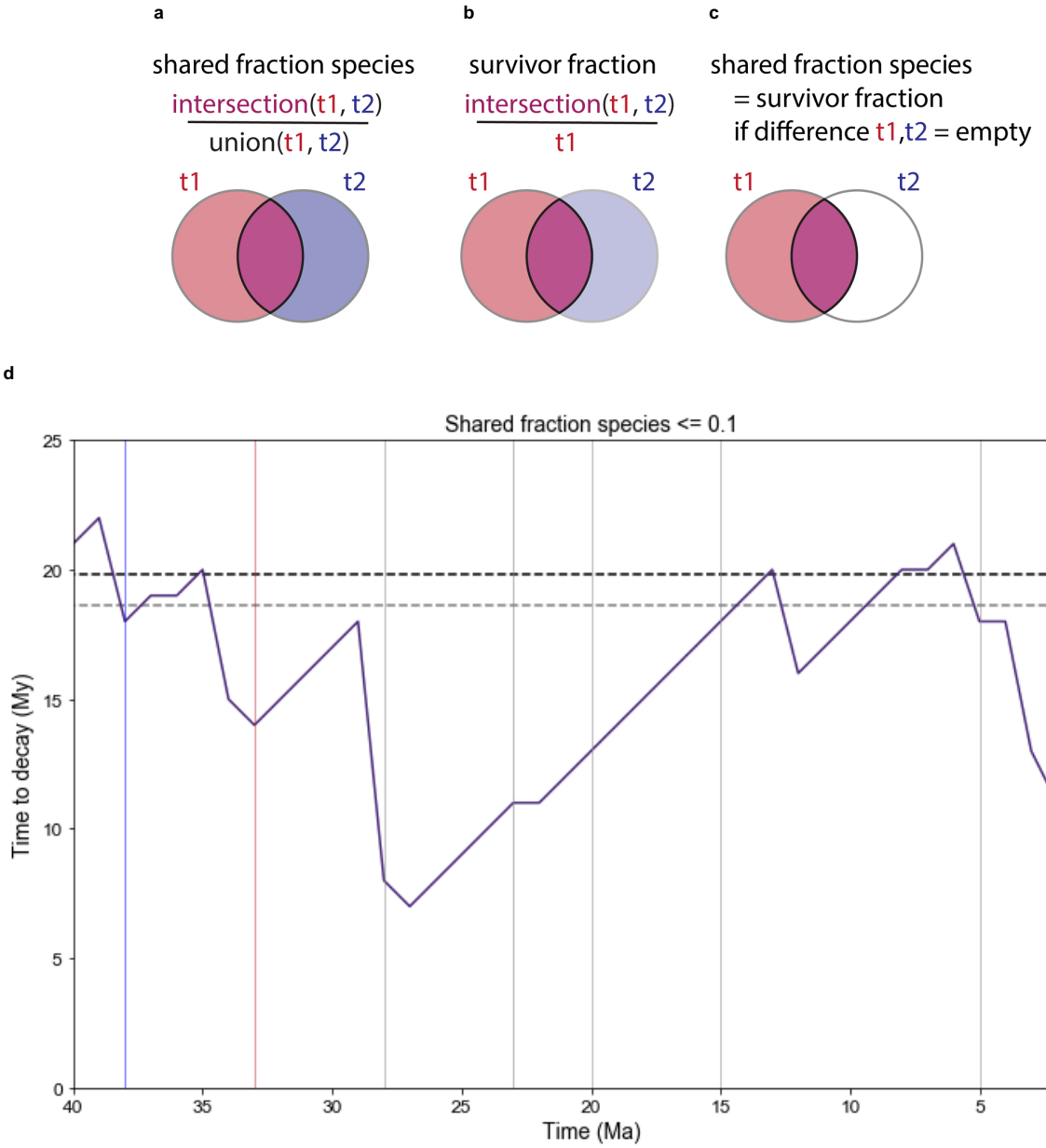
Extended Data Fig. 4 | Times of greatest fractional species turnover in the Phanerozoic fossil record. Top 5% most significant proportionate extinction or origination times (corresponding to the 30 labelled and coloured times in Fig. 3, > 42% species entry/exit threshold). Drill plots for focal times (key, top left) comparing stratigraphic ranges of all species occurring within 1 Myr of the focal time, vertically sorted into originations, extinctions and crossing ranges. Colours indicate over threshold mass extinctions (red), mass radiations (blue) and mixed mass extinction–radiations (magenta). Relevant stratigraphic unit names, dates and corresponding references are those used in the PBDB (*ref. [30]; **refs. [31,32]).



Extended Data Fig. 5 | Breakdown by phylum of species extinctions and originations at the top 5% of evolutionary disruption times. Associated with Fig. 3, Extended Data Fig. 4. Proportions of species entering (dark blue) or exiting (dark red) the fossil record are shown for the 19 most prevalent phyla in the data set (taxonomically screened data set).



Extended Data Fig. 6 | Raw species time ranges and diversity counts and examples of the decay in probability of temporal co-occurrence. a, Raw species time ranges: time ranges (maximum occurrence – minimum occurrence) for 137,779 fossil species (taxonomically screened data set). Taxonomically screened Phanerozoic data set (535–0 Ma): median = 6.5 Myr, mean = 9.95 Myr, s.d. = 12.86. Complete data set: median 7 Myr, mean 14.4 s.d. = 28.1 Myr. **b**, Raw diversity counts: sampled-in-bin taxonomic diversity of genera (grey dashed line) and families (black line) for the complete data set, output by the PBDB within the default time bin of geological ages (at maximum Ma). **c**–**f**, Examples of decays in co-occurrence probability (**c**, **e**) or in shared fraction of species (**d**, **f**), from base times 1 Myr before versus after major evolutionary disturbance events. Grey dashed lines indicate a value of 0.1. **c**, **d**, End-Permian mass extinction at 252 Ma. **e**, **f**, Carboniferous mass radiation at 358 Ma. Following a disturbance event, co-occurrence probabilities and shared fractions of species fall more rapidly to low levels because comparatively few living species co-occur with any species that were present in the past.



e

Measure	Sample size	Statistical test	Test statistic	Test statistic value	P value
Shared fraction species	532	Shapiro Wilk normality test	W	0.8117	1.94E-24
		Spearman nonparametric rank order correlation	r	0.054785	0.2071
Embedding distance	532	Shapiro Wilk normality test	W	0.9091	2.641E-17
		Spearman nonparametric rank order correlation	r	0.18554	1.6561E-05
Embedding distance 532-67 Ma	466	Spearman nonparametric rank order correlation	r	0.073272	0.1142
Fraction species going extinct	222	Shapiro Wilk normality test	W	0.8573	1.626E-13
		Spearman nonparametric rank order correlation	r	0.071236	0.29063
Fraction species originating	222	Shapiro Wilk normality test	W	0.8951	2.478E-11
		Spearman nonparametric rank order correlation	r	0.097941	0.1458

Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Conceptual diagram comparing measures of macroevolutionary decay, decay-clock detail focusing on the last 40 Myr and statistical relationships between measures of macroevolutionary disturbance and time. a, Set representation of the shared fraction of species between compared times (for example, times t_1 and t_2). This measure is used in this study and is closely conceptually related to the co-occurrence probability calculated using the ML spatial embedding (see Methods for further details). **b**, Fraction of surviving species, a core concept of standard methods of survivor analysis for example[4]. These measures (**a**, **b**) will be equal if no new species have originated by time t_2 (scenario in **c**). Where new species have instead originated by time t_2, their effect will be picked up by the measures used in this study (**a**), whereas the impact of new species would not be considered by measures only of the fraction of survivors from t_1 (**b**). **d**, Vertical lines indicate times of evolutionary disturbance (blue, mass radiations; red, mass extinctions, corresponding to Fig. 3; grey, turnover events below the mass-event threshold). **e**, (1), measures of disturbance to co-occurrence structure calculated between consecutive time windows are largely independent of Phanerozoic time (over which there have been secular trends in raw diversity[20]). The shared fraction of species shows no significant relationship with time (taxonomically screened data set). The embedding distance (complete data set) shows a weak relationship across the whole Phanerozoic that is removed when Cenozoic data are excluded (data excluded in order to isolate hypothesized effect after initial data analysis), consistent with a weak effect on Cenozoic embedding distance from fossils with ranges extending to 0 Ma (which are particularly abundant in the data set). (2), proportions of species exiting or entering the fossil record within 1 Myr of a given time show no significant relationship with time (taxonomically screened data set). All statistical tests are two-tailed.

nature research

Corresponding author(s): J. Hoyal Cuthill

Last updated by author(s): 20/09/2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection
Data analysis	Custom python code (using Python version 3.6.6) was used for data analysis and is provided as Supplementary Computer Code (sole documented version therefore unnumbered).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All raw data supporting the results are publicly available (Paleobiology Database), secondary data and statistics can be reproduced using the provided supplementary computer code.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Machine learning analysis of raw fossil occurrence data for 171,231 independent species downloaded from the public Palaeobiology Database
Research sample	1,273,254 fossil occurrence records of 71,231 independent species downloaded from the public Palaeobiology Database
Sampling strategy	Exhaustive sampling of the databased fossil occurrences from 1000 to 0 million years ago
Data collection	Data were downloaded on 2020-03-19 from http://paleobiodb.org/data1.2/occs/list.csv?datainfo&rowcount&taxon_reso=species&max_ma=1000&min_ma=0&show=class,acconly,coords,paleoloc,env,acconly". J.F.H.C downloaded data from a public repository. No instruments were used to personally record data for this study.
Timing and spatial scale	1000 to 0 million years ago, all global occurrences
Data exclusions	No data were excluded from the main analyses. One repeated statistical analysis excluded recent data to aid identification of an effect as documented in the captions to the relevant table of statistical results (Extended Data Table 1). These data were excluded in order to isolate hypothesised effect after initial data analysis.
Reproducibility	All data analyses were automatically scripted and can be reproduced by re-running the provided computer code. Statistical reproducibility of the machine learning analyses was formally and successfully tested using a bootstrap analysis as described in the paper.
Randomization	This is not relevant to this study which analysed all available data without splitting these into groups
Blinding	This is not relevant to this study which does not involve group allocation or statistical comparison

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☐	☒ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Palaeontology

Specimen provenance	This is not relevant to this study on aggregated fossil occurrence records
Specimen deposition	This is not relevant to this study as no fossil specimens were deposited
Dating methods	This is not relevant to this study as no new dates were obtained

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.