Identification of the Mode of Evolution in Incomplete Carbonate Successions

Niklas Hohmann[*,1,2] ORCID: 0000-0003-1559-1838

Joël R. Koelewijn[1] ORCID: 0000-0002-4668-3797

Peter Burgess[3] ORCID: 0000-0002-3812-4231

Emilia Jarochowska[1] ORCID: 0000-0001-8937-9405

[*]Corresponding author, email: N.H.Hohmann@uu.nl

[1]Utrecht University, Faculty of Geosciences, Department of Earth Sciences, Vening Meinesz building A, Princetonlaan 8A, 3584 CB Utrecht, The Netherlands.

[2]University of Warsaw, Faculty of Biology, Biological and Chemical Research Centre, Institute of Evolutionary Biology, ul. Żwirki i Wigury 101, 02-089 Warsaw, Poland.

[3]University of Liverpool, Department of Geology, Chatham St, Liverpool L69 7ZT, United Kingdom.

Abstract

Background: The fossil record provides the unique opportunity to observe evolution over millions of years, but is known to be incomplete. While incompleteness varies spatially and is hard to estimate for empirical sections, computer simulations of geological processes can be used to examine the effects of the incompleteness *in silico*.

We combine simulations of different modes of evolution (stasis, (un)biased random walks) with deposition of carbonate platforms strata to examine how well the mode of evolution can be recovered from fossil time series, and how test results vary between different positions in the carbonate platform and multiple stratigraphic architectures generated by different sea level curves.

Results:

Stratigraphic architecture and position along an onshore-offshore gradient has only a small influence on the mode of evolution recovered by statistical tests. Tests fail to identify the correct mode of evolution in the absence of stratigraphic effects, and support for the correct mode decreases with time series length.

Visual examination of trait evolution in lineages shows that rather than stratigraphic incompleteness, maximum hiatus duration determines how much fossil time series differ from the original evolutionary process. Directional evolution is more susceptible to stratigraphic effects, turning it into apparent punctuated equilibrium. In contrast, stasis remains unaffected.

Conclusions:

- Tests for the mode of evolution should be reviewed critically, as they do not find good support for the correct (simulated) mode of evolution, even for adequate models that

generated the data, in the absence of stratigraphic effects, and for exceptionally long time series.

- Fossil time series favor the recognition of both stasis and complex, punctuated modes of evolution.
- Not stratigraphic incompleteness, but the presence of rare, prolonged gaps has the largest effect on trait evolution. This suggests that incomplete sections with regular hiatus frequency and durations can potentially preserve evolutionary history without major biases. Understanding external controls on stratigraphic architectures such as sea level fluctuations is crucial for distinguishing between stratigraphic effects and genuine evolutionary process.

Keywords:

Paleontology, Stratigraphy, Trait Evolution, Paleobiology, Carbonate Platform, Mode of Evolution, Time Series, Fossil Record.

Introduction

The fossil record as source of information

Fossils provide a unique record of evolution on temporal and spatial scales not accessible to experimentation or direct human observation (Gingerich 1983; 2001). Geological records have delivered fossil time series crucial in formulating and testing hypotheses on evolutionary dynamics and mechanisms of speciation spanning micro- to macroevolutionary scales (e.g., Dzik 2005; Strömberg 2006; Aze et al. 2011; Voje 2020). Nevertheless, fossils

remain underused in evolutionary biology. Their main application is still calibration of molecular clocks, which commonly relies on single occurrences and is subject to biases resulting from this small sample size (Springer 1995). The unique type of information contained in a fossil succession sampled over a long time interval is rarely exploited, likely due to the following barriers:

1. The fossil record, being a part of the stratigraphic record, is patchy and distorted. At the time when Darwin (1859) discussed this as a major limitation for the testing and development of the theory of evolution, little geological knowledge was present to elucidate the rules governing this incompleteness. Darwin's concern widely persists (e.g., Patterson (1981)), albeit mostly implicitly: most phylogenetic analyses published today do not use fossils which would have been relevant or use only a small fraction of them. Stratigraphy and sedimentology, which can provide relevant data on fossils and the constraints on their occurrence and sampling (Kidwell and Holland 2002; Hunt 2010), are jargon-laden, highly atomized disciplines whose utility for evolutionary biology is not obvious to biologists. Biostratigraphy, which uses fossil to establish the relative age of rocks and has amassed datasets that would be of high utility for evolutionary studies, employs taxonomic concepts that are often impractical for or incompatible with evolutionary questions (Dzik 1985; 1995; Pearson 1992; Haug and Haug 2017). As a results, scientific communities studying evolution and the fossil record function in parallel, with limited exchange (Grantham 2004).
2. There is a lack of methodological frameworks to incorporate fossils in evolutionary studies. Historically, phylogenetic methods rarely incorporated geological information such as the relative order of appearance of taxa or specimens in the fossil record, which is the main subject of biostratigraphy (Wills 1999). This has led to radical discrepancies between the outcomes of phylogenetic and stratigraphic, or

stratophenetic, approaches (Gingerich and Schoeninger 1977; Donoghue 2001; Dzik 2005). This barrier is gradually overcome by methodological advances, such as the Fossilized Birth-Death Model (Stadler et al. 2018), which allows incorporation of parameters specific to the fossil record, such as fossilization rate, sampling probability and age uncertainties of fossil occurrences (Barido-Sottani et al. 2020; Warnock, Heath, and Stadler 2020; Wright et al. 2022; Barido-Sottani et al. 2023).

Recently, there is renewed appreciation for the importance of fossils in phylogenetic reconstructions (Quental and Marshall 2010; Mitchell, Etienne, and Rabosky 2019; Mongiardino Koch, Garwood, and Parry 2021; Guenser et al. 2021). These studies focus on the role of the morphological information provided by extinct taxa, but less on what a modern understanding of the physical structure of the geological record contributes to reconstructing evolutionary processes from fossil-bearing stratigraphic successions.

Stratigraphic incompleteness and age-depth models

The incompleteness of the fossil record serves as an umbrella term for different effects that diminish the information content of the rock record, ranging from taphonomic effects and sampling biases to the role of gaps and erosion (Kidwell and Holland 2002). Here we focus on the role of gaps (hiatuses) in the rock record. Such gaps can arise due to sedimentation (including fossils) and subsequent erosion or lack of creation of rocks in the first place, e.g. when an environment remains barren of sediment formation or supply for a long time. Both processes result in gaps in the rock record and, as a result, in the fossil record. This type of incompleteness is termed stratigraphic (in)completeness, defined as the time (not) recorded in a section, divided by the total duration of the section (Dingus and Sadler 1982; Tipper 1987). Stratigraphic completeness provides an upper limit on the proportion of evolutionary history that can be recovered from a specified section, even with unlimited resources and perfect

preservation of fossils. Stratigraphic completeness is difficult to quantify in geological sections, and estimates range between 3 and 30 % (Wilkinson, Opdyke, and Algeo 1991), suggesting that more than 70 % of evolutionary history is either not recorded in the first place or destroyed at a later time.

Fossils older than a 1.5 million years cannot be dated directly, and their age has to be inferred from circumstantial evidence on the age of the strata in which they were found (Wehmiller et al. 1988; Kidwell and Flessa 1996). This inference is formalized by age-depth models (ADMs), which serve as coordinate transformations between the stratigraphic domain, where the fossils were found (length dimension L, SI unit meter), and the time domain (time dimension T, SI unit seconds - we use the derived units years, kyrs, or Myrs) (Hohmann 2021). Age-depth models are always explicitly or implicitly used when fossil data is used for evolutionary inferences. Because they convey how positions of fossils relate to their age, ADMs are the basis for calculating evolutionary rates. As a result, revising ADMs commonly leads to a revision of evolutionary rates. For example, Malmgren, Berggren, and Lohmann (1983) observed increased rates of morphological evolution in lineages of fossil foraminifera over a geologically short time interval of 0.6 Myr and proposed that this “punctuated gradualism” may be a “common norm for evolution”. MacLeod (1991) revised the age-depth model and showed that the interval with increased rates of evolution coincides with a stratigraphically condensed interval, i.e. more change is recorded in a thinner rock unit. Re-evaluating the evolutionary history based on the revised age-depth model removed the apparent punctuation and showed that morphological evolution in that case had been gradual rather than punctuated.

Age-depth models contain information on both variations in sediment accumulation rate and gaps in the stratigraphic and - as a result – the fossil record. For example, stratigraphic completeness corresponds to the fraction of the time domain to which an age-depth model

assigns a stratigraphic position. In the absence of an age-depth model, we can only make statements on the ordering of evolutionary events, but not on the temporal rates involved.

Forward models of stratigraphic architectures

Forward computer simulations of sedimentary strata provide a useful tool to study the effects of incompleteness and heterogeneous stratigraphic architectures. They have demonstrated that locations and frequency of gaps in the stratigraphic record are not random, but a predictable result of external controls, such as fluctuations in eustatic sea level (Warrlich 2000; Hutton and Syvitski 2008; Burgess 2013; Masiero et al. 2020).

Combined with biological models, forward models provide a powerful tool to test hypotheses on the effects of stratigraphic architectures on our interpretations of evolutionary history. For example, Hannisdal (2006) combined simulations of a siliciclastic basin with models of taphonomy and phenotypic evolution. The results showed that when sample sizes are small, morphological evolution will appear as stasis regardless of the underlying mode. This might explain why stasis is the most common evolutionary pattern recovered from the fossil record (Hunt, Hopkins, and Lidgard 2015).

Stratigraphic incompleteness and variations in sediment accumulation rates introduce multiple methodological challenges. Constructing complex ADMs requires sedimentological and stratigraphic expert knowledge, and they will potentially be associated with large uncertainties. Even in the “perfect knowledge” scenario where the age-depth model is fully known, evolutionary history in the time domain will inevitably be sampled irregularly: If two samples are separated by a hiatus, their age difference must be at least the duration of the hiatus, which might be millions of years. On the other hand, if sediment accumulation is rapid and no hiatuses are present, the age difference between samples might be only a few days.

Most studies "translate" fossil successions into time series using age-depth models based on simplified assumptions on the regularity of the stratigraphic record. These ADMs ignore stratigraphic incompleteness and often assume uninterrupted constant sediment accumulation (UCSA). This assumption implies that stratigraphic completeness is 100 %, rock thickness is proportional to time passed, and linear interpolation between tie points of known age can be used to infer fossil ages from their positions. Such ADMs are usually used implicitly, without discussing their limitations. While the assumption of UCSA is sedimentologically and stratigraphically unrealistic, it brings strong methodological simplifications. For example, if distance between samples collected in a rock section is kept constant, UCSA implies that the underlying evolutionary history in the time domain is sampled at a constant frequency, the generated fossil time series are equidistant in time and can therefore be analyzed by standard methods of time series analysis (Hunt 2006; Beran 2017).

Objectives and hypotheses

We examine how commonly made simplified assumptions on stratigraphic architectures influence how the mode of evolution is recovered from fossil time series. We use tropical carbonate platforms as a case study, because they host large parts of the fossil record and are evolutionary hotspots (Jablonski, Roy, and Valentine 2006).

We test the following hypotheses:

1. The mode of evolution identified in a fossil time series obtained under the assumption of uninterrupted constant sediment accumulation (UCSA) is the same as the mode of the original time series.
2. Lower stratigraphic completeness reduces the chance of identifying the correct mode of evolution from fossil time series constructed based on the assumption of UCSA.

The implication of this hypothesis is that different depositional environments have different chances of preserving the mode of evolution because of systematic differences in their completeness.

Methods

We simulate trait evolution in the time domain, pass it through a stratigraphic filter produced by the CarboCAT Lite model of carbonate platform formation (Burgess 2013; Burgess 2023), and compare how well the mode of evolution can be recovered from the fossil time series sampled in the stratigraphic domain and the time series reflecting the "true" evolutionary history in the time domain. A visual summary of the workflow is shown in Figure 1. Data is available in Hohmann, Koelewijn, and Jarochowska (2023a), code is available Hohmann, Koelewijn, and Jarochowska (2023b). See the README and REPRODUCEME files therein for details on computational reproducibility.

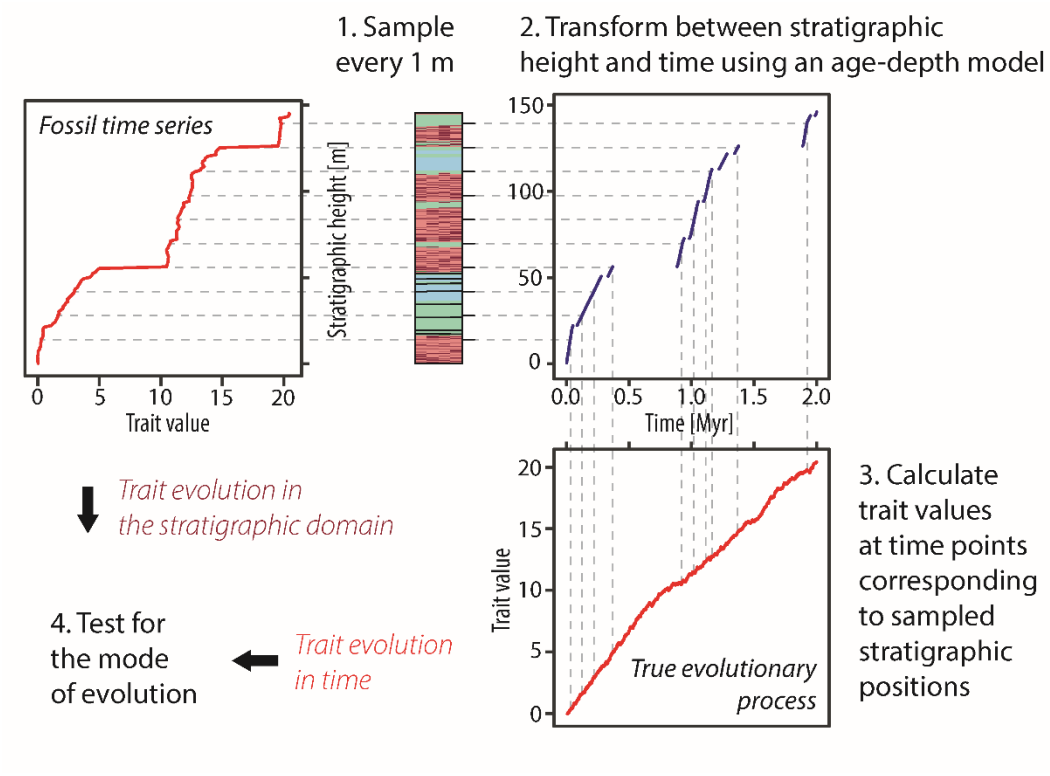

Figure 1: Study design for testing the mode of evolution in the stratigraphic domain. Computationally, first sampling positions are determined, then the age-depth model is used to determine the times that correspond to these positions. Last, the trait evolution at said times are simulated. The simulated mean trait values are the values observable at the sampled stratigraphic positions.

The stratigraphic record: Forward models of carbonate platform architecture

We simulated two attached carbonate platforms using the CarboCAT Lite software (*Figure 2* and *Figure 3*) (Burgess 2023). CarboCAT Lite is implemented in MATLAB and simulates carbonate production by carbonate factories characterized by different production curves, which are functions of water depth (Bosscher and Schlager 1992; Burgess 2013; Masiero et al. 2020). It includes sediment transport that is a function of platform topography, i.e.,

sediment is transported downslope, but not a function of external parameters such as waves or currents. Simulations were run using time steps of 1 kyr (1000 years) and a grid of 1 km width (strike) and 15 km length (dip), subdivided into quadratic grid cells of 100 m length. We used three carbonate factories with production curves following the parametrization of Bosscher and Schlager (1992) (*Supplementary Figure 1*). Factory 1 is phototrophic with a maximum carbonate production rate of 500 m × Myr^{-1}, which it maintains up to 30 m of water depth. Factories 2 and 3 have maximum production rates of 160 and 150 m × Myr^{-1} respectively and maintain maximum productivity until 40 m water depth, but differ in how fast productivity decreases with depth (*Supplementary Figure 1*). We assumed constant subsidence of 70 m per Myr across the grid. This is approximately seven times higher than estimated subsidence rates in the last 125 kyr in the Bahamas (McNeill 2005), allowing for higher completeness than would be generated on a typical passive continental margin. As initial topography, a slope with a gradient of 5.33 m × km^{-1} was used, resulting in a total initial difference in water depth of 80 m between the shore and the offshore end of the simulated grid. We followed the default setting of CarboCAT Lite, described by Burgess (2013) with three lithofacies (sediment types) corresponding to the three carbonate factories, plus three secondary lithofacies representing sediment transported downslope after the deposition of the primary lithofacies (dark colors in *Figures 2* and *3*). As initial facies distribution, the default setting of CarboCAT Lite was used, which assigns each grid cell one of the three carbonate factories or no factory according to a uniform distribution.

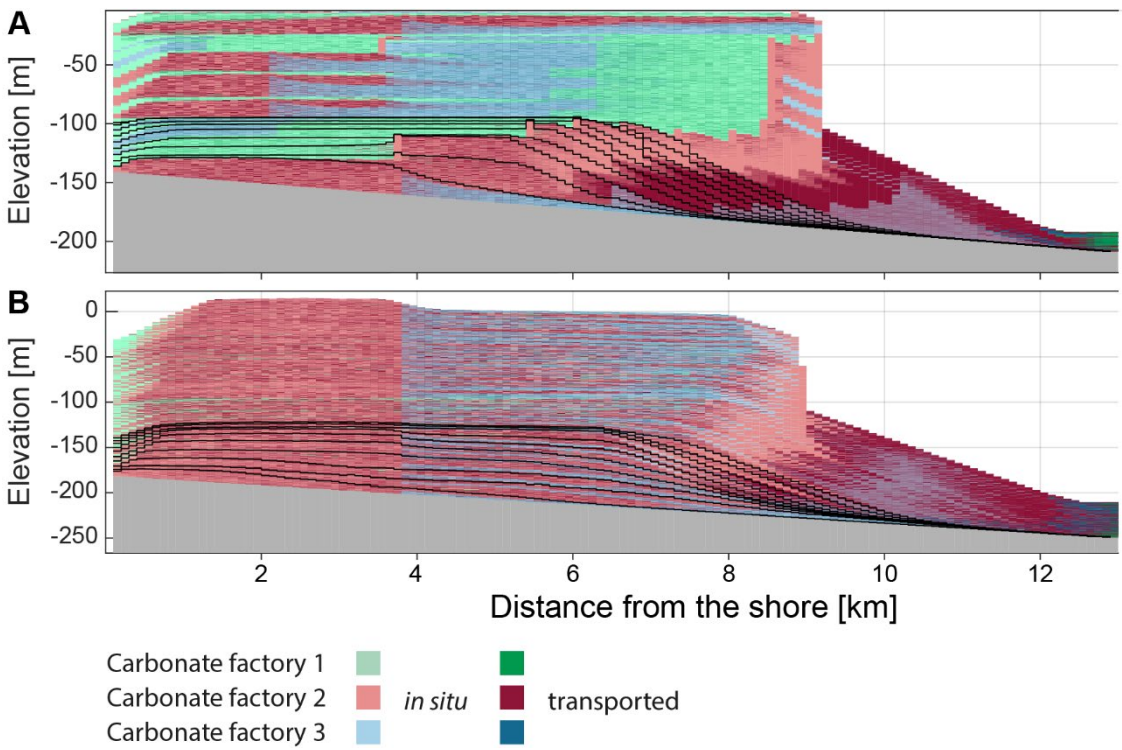

Figure 2: The outcome of simulating carbonate platforms in the stratigraphic domain. (A) Scenario A: deposition based on a fictional sea-level curve. (B) Scenario B: deposition based on the sea-level curve from Miller et al. (2020) for the last 2.58 Myr. Graphs represent the position in the middle of the simulated grid along the strike.

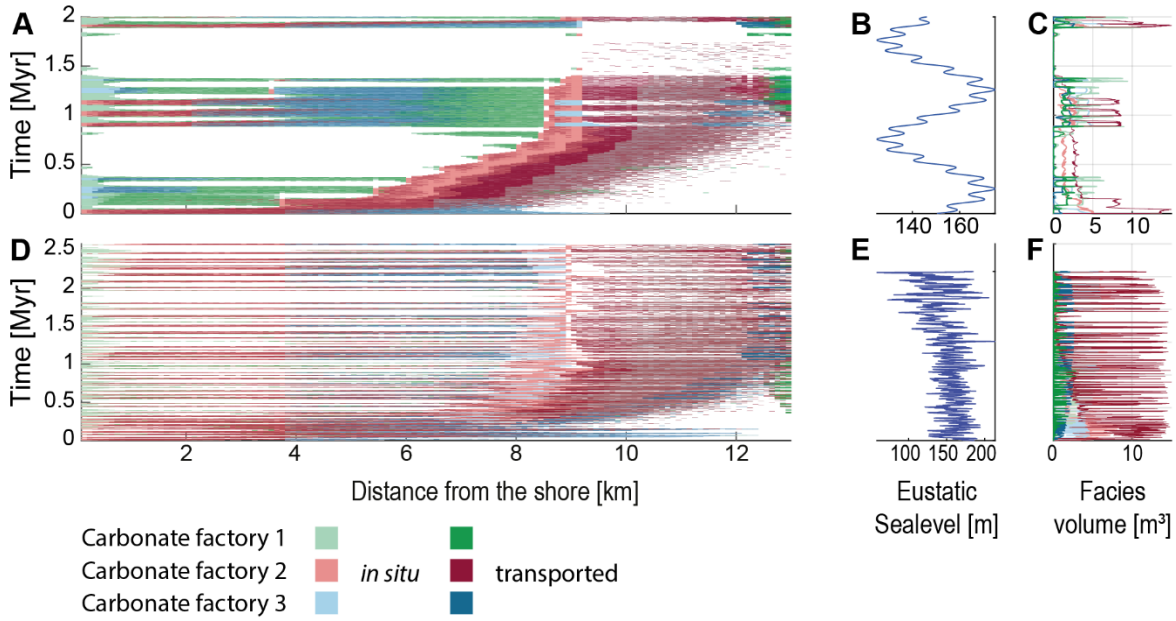

Figure 3: Simulated carbonate platforms in the time domain. (A-C) Scenario A. (D-F) Scenario B. (A, D) Chronostratigraphic (Wheeler) diagrams. (B, E) Sea level curves used as input for the simulation. (C, F) Facies volumes. Graphs represent the position in the middle of the simulated grid along the strike.

Two scenarios were simulated (*Figures 2* and *3*). In scenario A, the simulation was run for 2 Myr with changes in eustatic sea level given by a combination of sinusoids expressed as third-order sea level changes with a period of 1 Myr and an amplitude of 20 m and fourth-order changes with a period of 0.112 Myr and an amplitude of 2 m (*Figure 3* B). Scenario B was simulated for the period from 2.58 Mya (beginning of the Pleistocene) until the present using the global eustatic sea level curve estimated from global $\delta^{18}O_{benthic}$ records by Miller et al. (2020). This curve has a temporal resolution of approximately 2 kyr over the Pleistocene (mean: 2.13 kyr, median: 2.07 kyr), and was linearly interpolated to match the time increments of the simulation and normalized to have a mean sea level of 0 (*Figure 3* E).

For each scenario, CarboCAT Lite produces a grid containing the facies, including no deposition, at each time step and position on the grid. The outputs were extracted from

MATLAB and all further analysis was carried out in the R Software (R Core Team 2023). For both scenarios, age-depth models at every grid node were extracted from the simulation outputs. Age-depth models taken from the same distance from shore display only little variation as position along the shore (strike) is varied, indicating that lateral position in the grid is negligible. Accordingly, we focused on age-depth models taken from the middle transect along the dip. In each scenario, synthetic sections and the corresponding age-depth models at five locations were selected (*Supplementary Figure 2*). The selected locations are 2 km, 6 km, 8 km, 10 km, and 12 km from the shoreline. These distances corresponded to lagoonal environments, backreef, forereef, proximal slope, and distal slope. The simulated carbonate platforms accreted vertically with only minor shift of environments along the onshore-offshore axis, therefore each location can be interpreted as representing its respective environment across the entire simulation (platform lifetime). Only simulation outputs between the shore and 13 km offshore are shown in the figures, as the remainder (13-15 km) consisted entirely of redeposited carbonates and was not used further.

The simulated carbonate platforms were characterized using the following parameters calculated at any given location in the platform:

1. Stratigraphic completeness, calculated as the proportion of time steps with sediment accumulation relative to the total number of time steps in the simulation (function `get_completeness` in Hohmann, Koelewijn, and Jarochowska (2023b)), corresponding to completeness on the time scale of 1 kyr (Tipper 1987; Anders, Krueger, and Sadler 1987; Kemp 2012)
2. Distribution of hiatus durations in Myr (function `get_hiatus_distribution` in in Hohmann, Koelewijn, and Jarochowska (2023b))

Simulation of phenotypic trait evolution in the time domain

We simulated three commonly discussed modes of evolution in the time domain: stasis, unbiased random walk, and biased random walk (Hunt 2006; Jones 2009; Hopkins and Lidgard 2012; Hunt, Hopkins, and Lidgard 2015). These simulations serve as "true" evolutionary history against which we compare fossil time series derived under simplified assumptions on age-depth models.

In unbiased random walk models, the change in traits over a fixed time step is independent of the previous trait values, and drawn from a distribution with mean zero (Bookstein 1987; Sheets and Mitchell 2001). In biased random walk models, the mean can deviate from zero. The trait evolution along a lineage is then generated by summing up the incremental changes in traits. Commonly used models of random walks are based on equidistant time steps, meaning the time passed between two observations of the random walk is identical for all observations. We could not use these models, because depending on where in the stratigraphic column samples are collected, the time elapsed between two successive sampling positions can vary by more than three orders of magnitude. It will be less than a thousand years when sediment accumulation rate is high, and more than half a million years when they are separated by a hiatus (*Figure 4*). If samples have identical distances in the stratigraphic column, the underlying evolutionary history will be sampled at irregular times. We use continuous-time extensions of random walk and stasis models so we can sample lineages at arbitrary time points without relying on interpolation between discrete time steps. Interpolation would potentially introduce dependencies between successive samples, contradicting the assumption of the random walk model that change in traits is independent

and identically distributed. The continuous-time expansion provides exact trait values at the sampled times, and reduces to the standard discrete time step models when applied to equidistant time steps. In addition, discrete-time models incorporate time elapsed between observations into model parameters (e.g., the variance parameter of the random walk model), making their scaling behavior non-obvious when time scales are varied. Using continuous-time implementations makes it possible to compare observations across time scales and vary the granularity of observation while keeping the underlying model of evolution identical. We use this property to examine how increased sampling of the same time interval influences model selection performance. Another commonly considered mode is the Ornstein-Uhlenbeck (OU) process, which corresponds to noisy convergence towards a trait optimum (Lande 1976). The stochastic differential equations describing Ornstein-Uhlenbeck are usually solved using the Euler-Maruyama method with equidistant time steps (Platen and Bruti-Liberati 2010). We did not simulations of OU processes due to the difficulty of generating heterodistant samples (samples not equally spaced in time) from them.

We extracted age-depth models from CarboCAT Lite outputs. The age-depth models serve as functions $H : L \rightarrow T$ that connect the stratigraphic domain L with the time domain T. Given a model of trait evolution M_t in the time domain, we then define its stratigraphic expression as $M \circ H^{-1}$, which is the trait evolution observable in the stratigraphic column. This allows us to form triplets (h, t, m), where h is a stratigraphic position, t is the time when that position was formed, and m is the trait value that describes the true evolutionary history at time t and can be observed at height h (*Figure 1*). The R package DAIME (Hohmann 2021) was used for age-depth transformations to sample time and depth domain at arbitrary points. The theoretical underpinning for the need of these transformations can be found in Hohmann (2021). For the (un)biased random walk models, we use Brownian drift as a continuous-time expansion. It is given by

$$X_t = \mu \cdot t + \sigma B_t$$

where μ and σ are model parameters, t is time, and B_t is a Brownian motion. The parameter μ specifies how biased or directional the Brownian drift is and σ specifies how volatile the process is. After t time units have passed, the distribution of a trait following a Brownian drift model is normally distributed with standard deviation σ and mean $\mu \times t + x_0$, where x_0 is the trait value at the beginning of the observation. When sampled at equidistant points in time, this reduces to a standard random walk model where a normal distribution determines the step sizes.

Stasis is expressed as the lack of net change in traits over the period of observation. We model stasis as a series of independent, identically distributed random variables with mean m and standard deviation s (Sheets and Mitchell 2001). As a result, stasis is not affected by the heterodistant sampling in the time domain.

We examined the preservation of four evolutionary scenarios (*Supplementary Figure 3*):

(1) Stasis with mean $m = 0$ and standard deviation $s = 1$;
(2) Brownian motion as a special case of Brownian drift with parameters $\mu = 0$ and $\sigma = 1$;
(3) weak Brownian drift with parameters $\mu = 5$ and $\sigma = 1$, corresponding to a weakly directional random walk;
(4) strong Brownian drift with parameters $\mu = 10$ and $\sigma = 1$, corresponding to a strongly directional random walk.

After one million years of evolution, the Brownian motion has an expected change in traits of zero, while the weak and strong Brownian drift have an expected change of traits of 5 and 10, respectively. For comparability, the σ parameter on the Brownian motion and drift models was kept constant, resulting in a standard deviation of traits around their mean value

equal to one after one million years. The four evolutionary scenarios differ in their directionality, defined as the difference in traits accumulated over time. While weak and strong Brownian drift are directional, stasis and Brownian motion on average are not. Individual lineages following a Brownian motion can deviate drastically from their initial trait value – only when multiple lineages are observed, the mean trait values observed in them average to zero (*Supplementary Figure 3*).

Sampling procedure in the stratigraphic domain

For clarity, we distinguish between fossil time series and time series. Fossil time series are taken in the stratigraphic domain and record trait values observed at specific stratigraphic positions, whereas time series are located in the time domain and record trait values observed at specific points in time.

To isolate the effects of stratigraphic incompleteness, we assume all lineages have an equal chance of being sampled across all environments. We assume identical sampling procedures for fossil time series in both scenarios: a sample is taken every meter and consists of 100 specimens. The distribution of traits among the specimens in a sample is normally distributed with variance of 0.1 and a mean according to the simulated trait values. This choice is made so the fossil time series match the format required by the paleoTS package (Hunt 2006; 2022), which we use to identify the mode of evolution. The variance of 0.1 was chosen to ensure that variability within a sample is small compared to the expected mean evolutionary change that accumulates over the course of the simulations, thus reducing the chance to mistake variability within populations for evolutionary trends (Hannisdal 2006). This procedure generates equidistant fossil time series with, depending on the location in the platform, 20 to 150 sampling positions in scenario A and 50 to 220 sampling positions in

scenario B. This is more than most studies with stratophenetic datasets (Clyde and Gingerich 1994; Dzik 1999; Hunt and Roy 2006). For example, fossil time series compiled by Hunt, Hopkins, and Lidgard (2015) had a median of 14 sampling positions and the longest fossil time series consisted of 114 sampling positions.

Identification of the mode of evolution from simulated fossil time series

We applied the tests for the mode of evolution implemented in the `compareModels` function of the paleoTS package version 0.5.3 (Hunt 2022) to the simulated stratigraphic series. The tests use corrected Akaike's Information Criterion (AICc) to determine which of the following models fits best to the fossil time series: the stasis model, an unbiased random walk (URW), a biased random walk referred to as general random walk (GRW), and an Ornstein-Uhlenbeck process (OU). While we did not simulate Onstein-Uhlenbeck processes, we included it in the set of tested models as it corresponds to a distinct mode of evolution (convergence to a phenotypic optimum). The models identified by the `compareModels` function correspond to the simulated processes as follows: URW to the Brownian motion, GRW to the weak and strong Brownian drift, and stasis to stasis. The tests take the between-samples variance in trait values and the number of specimens found at a sampling position into account. At each location in the carbonate platform, 100 fossil time series per evolutionary scenario (stasis, Brownian motion, weak and strong Brownian drift) were simulated and tested for their fit to the stasis, URG, GRW, and OU model. Because raw AICc values carry no meaning, we use the derived AICc weights instead. The (uncorrected) AIC weight of a model can be interpreted as the probability that it is the best approximating model, given the data and the set of candidate models (Wagenmakers and Farrell 2004). Thus higher AIC weights (and, as a result, AICc weights) reflect better support for the model. We

considered a model as identified by the paleoTS package if it is in the 90 % confidence set of models (Symonds and Moussalli 2011), meaning its AICc weight is higher than 0.9, making it a single best model according to Portet (2020).

As a baseline for the test performance of the paleoTS package, the tests were also performed in the time domain, i.e. without any losses or distortions introduced by the stratigraphic record. To this aim, the time interval of observation (2 Myr and 2.56 Myr for scenario A and B, respectively) was subdivided into 5, 10, 15, 20, 25, 35, 50, 100 and 200 equally spaced sampling points. Lineages evolving according to the specified modes of evolution were sampled at these time points, and the test for the modes of evolution was performed on the resulting time series. Our hypothesis for this baseline is that in the absence of stratigraphic effects, increased sampling effort (i.e., a higher number of subdivisions of the time interval of observation, resulting in longer time series) increases support for the correct mode of evolution.

Results

Stratigraphic architectures

Carbonate platform A

Platform A (*Figure 2* A; *Figure 3* A-C) has a steep topography with high build-up up to 8 km into the basin and thin, condensed off-platform deposits between 8 and 13 km, consisting mostly of transported sediment. The majority of transported sediment is derived from the phototrophic factory, with a sharp drop in production rate at around 30 m water depth

(*Supplementary Figure 1*). The thinnest interval at ca. 13 km into the basin contains also sediment transported from the two other factories.

The platform consists of two "depositional sequences" with different topographies. These "sequences" correspond to two sea-level highs, with the first one resulting in low topography and rapid progradation that led to basinward, rather than upward, growth of the platform. The second "sequence" is aggradational, with distinct "parasequences" expressed in facies, corresponding to the 125 kyr period in the sea-level curve used as simulation input. This "sequence" is responsible for the steep topography of the platform. The top of the platform is abruptly truncated as a result of the long-term cycle sea-level fall and covered with a thin deposit of the ensuing initial transgression.

The two long-term cycle sea-level falls result in two major gaps (*Figure 3* A). The gap between the first and the second "depositional sequence" does not extend uniformly across the entire platform. During the time of the gap formation, the euphotic factory and sediment derived from transport of this factory's products accumulated at the platform edge, prograding between 5 and 12 km basinward. The last part of this interval offlaps the platform, forming an architecture resembling the Falling Stage Systems Tract (Plint and Nummedal 2000). In contrast, the second major gap resulted in almost no deposition, reflecting almost no sediment transport in the second stage of platform formation. As a result, this platform is characterized by: two long gaps in deposition and several shorter gaps with limited spatial extent; large differences in thickness along the onshore-offshore gradient and a substantial contribution of sediment transport and offshore deposits formed entirely from sediment exported from the platform in the first half of the platform formation.

Carbonate platform B

For scenario B, we used the Pleistocene-Holocene global mean sea-level estimate of Miller et al. (2020). In the original dataset, the sea level varied between -121 m to 31.9 m and was here normalized to a mean of 0 m (*Figure 3* E). The simulated time interval corresponds to gradually lowering global sea level dominated by 41-ka tilt forcing leading to amplitudes reaching 50 m. A gradual increase in sea-level amplitude throughout this interval is attributed to the onset of the glaciation of the Northern Hemisphere. In the middle Pleistocene, ca. 800 ka, a shift to "quasi-100-ka" periodicity is associated with higher sea level amplitudes reaching 140 m (Miller et al. 2020). This Milankovitch-paced sea-level changes resulted in a simulated carbonate platform with short gap durations (*Figure 3* D, *Figure 4* B). Where present, the gaps are widespread across the platform, i.e. sea-level drops resulted in gaps uniformly distributed across all environments. The continuity of gaps is even more pronounced in the youngest part of the platform corresponding to the last 800 kyr, reflecting the higher sea-level amplitude. Most gaps are concentrated in the central part of the platform, whereas the most shore- and basinward parts of the platform may be filled with sediment at the time when no other strata is formed in the grid, leading to the highest completeness of these two opposite ends. The dominant facies is sediment transported from the phototrophic factory (dark red in *Figure 2*), which is present across the entire platform, rather than exported off the platform as in scenario A. Hence, this platform is dominated by short-distance sediment transport. In the older part of the platform, the offshore part (at a distance of 11-13 km from the shore) has a higher stratigraphic completeness (more time represented by sediment) than the onshore end, but in the younger part of the platform, the difference disappears. The lack of large sea-level falls results in a lack of major facies shifts, low topography and a gradual, nearly aggradational buildup (*Figure 2* B).

Completeness and distribution of hiatuses

Changes in completeness along the onshore-offshore gradient were very similar for both scenarios: completeness increased monotonously from the shore and had a double peak around 9 and 11 km offshore, where it reached values of 75 and 95 percent (onshore and offshore peak) in scenario A and around 80 percent (both peaks) in scenario B (*Figure 4*). The onshore peak in completeness coincided with the forereef environment, where organisms continuously grow, while the second peak coincided with the slope onset, where transported sediment is continuously provided by the forereef. Average completeness across the platform was 53.7 % in scenario A and 46.3 % in scenario B, completeness was lowest 13 km from shore in scenario A and 1.9 km from shore in scenario B.

Qualitatively, the distribution of hiatus durations was similar for both scenarios. Maximum hiatus duration was constant over the entire platform, dropped dramatically on the slope, and then increased again off-platform. The first quartiles, medians and third quartiles of hiatus durations differed substantially from maximum hiatus durations. The exception to this is the backreef in scenario A (approx. 4–6 km from shore), where the 3rd quartile of hiatus durations reaches values of 0.25 to 0.4 Myrs. Averaged over the platform, maximum hiatus duration is one to two orders of magnitude larger than median hiatus duration (135 times longer in scenario A and 13 times longer in scenario B). This shows that the distribution of hiatus durations on the platform is heavy-tailed: most hiatuses are short, but a few exceptionally long hiatuses that are associated with long-term drops in sea level are present.

Quantitatively, hiatus durations in scenario A are longer (average median hiatus duration across the platform: 26 kyr in scenario A, 10 kyr in scenario B). The maximum hiatus duration is on average almost four times higher in scenario A than in scenario B.

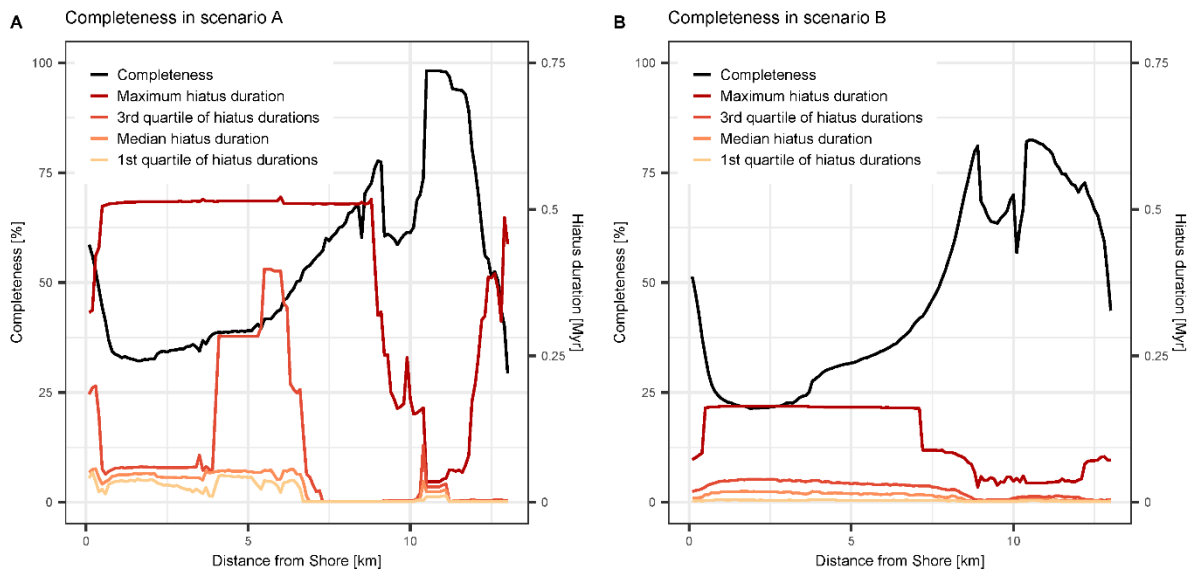

Figure 4: Stratigraphic completeness and distribution of hiatus durations along the onshore-offshore gradient in scenario A (left) and B (right). Maximum hiatus duration in scenario B is four times lower than in scenario B, while completeness is comparable.

Stratigraphic expression of evolution

Overall, we find that stratigraphic effects on the preservation of the mode of evolution are spatially heterogeneous within the carbonate platforms (*Figure 5*), and strongly depend on (1) the directionality of the examined mode of evolution (*Figure 6)* and (2) the presence of single, long hiatuses rather than the stratigraphic completeness at the location where the fossil time series is sampled (*Figure 7*). As a result, under high-frequency sea-level changes (scenario B), trait evolution is much more similar to true evolutionary change in the time domain than under slow-frequency sea-level changes (scenario A).

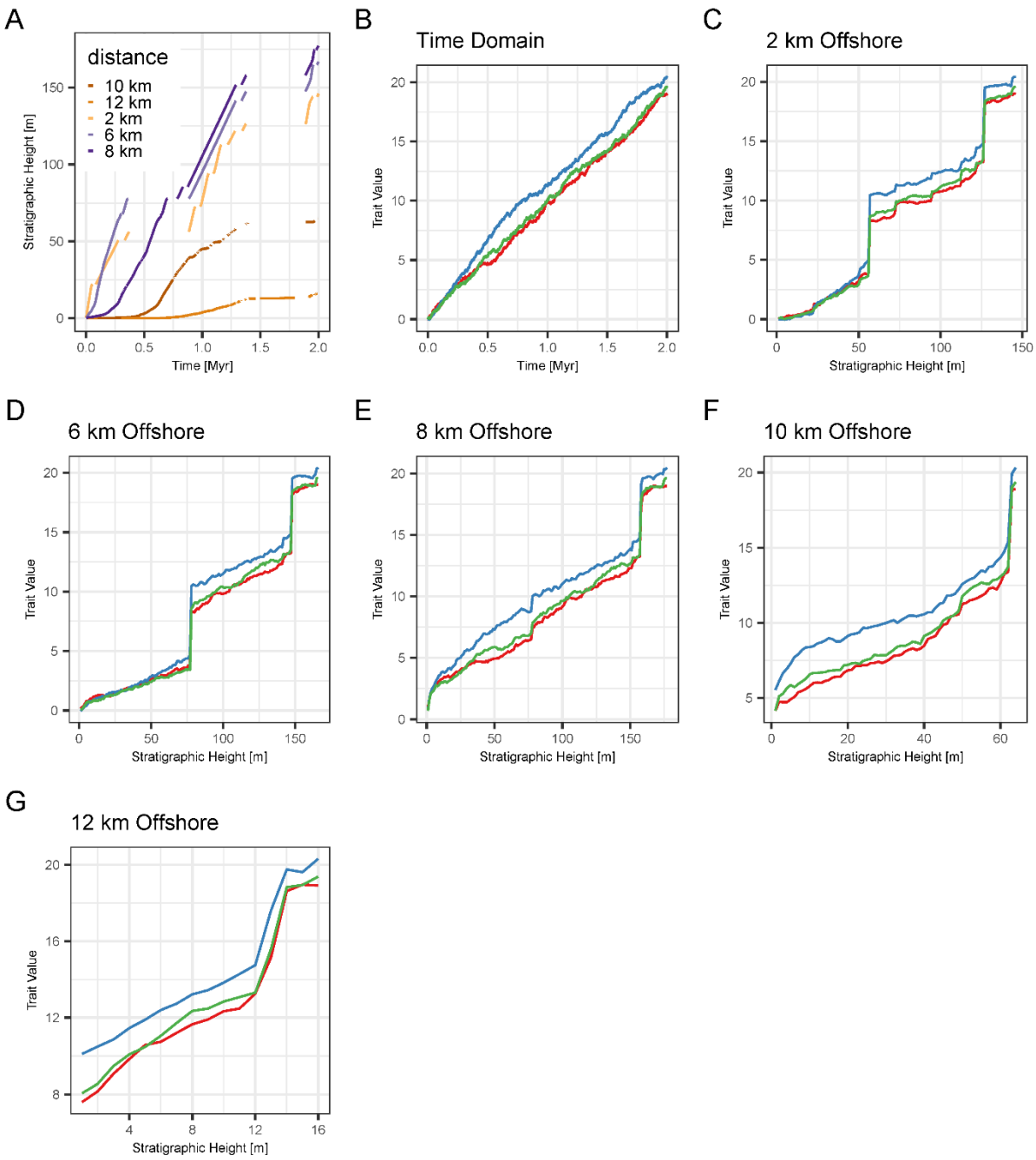

Figure 5: Spatial variability of the preservation of evolution in scenario A. (A) Age-depth models at varying distances from shore (B) three simulations of strong Brownian drift in the time domain (C), (D), (E), (F), (G) preservation of the lineages from (B) in the stratigraphic domain at 2 km, 6 km, 8 km, 10 km, and 12 km from shore in platform A. The same evolutionary history (B) is preserved differently dependent on where it is observed (C to G).

Differential effects of stratigraphy on modes of evolution

The extent to which stratigraphic series of trait values are affected by the architecture of the carbonate platform strongly depends on the mode of evolution (*Figure 6, Supplementary Figures 4–12*). For the same scenario and position in the platform, lineages evolving according to the stasis model are unaffected by stratigraphy. In contrast, hiatuses introduce jumps into the lineages evolving according to a weak and strong Brownian drift, which are not present in the time domain. For the Brownian motion, the presence of jumps depends on whether the trait series accumulates sufficient trait differences during the hiatus, making the presence or absence of jumps over gaps effectively random (*Figure 6*, B & F, red and green vs. blue lineage). In general, jumps in traits are more pronounced when (1) evolution is more directional and (2) gaps are longer.

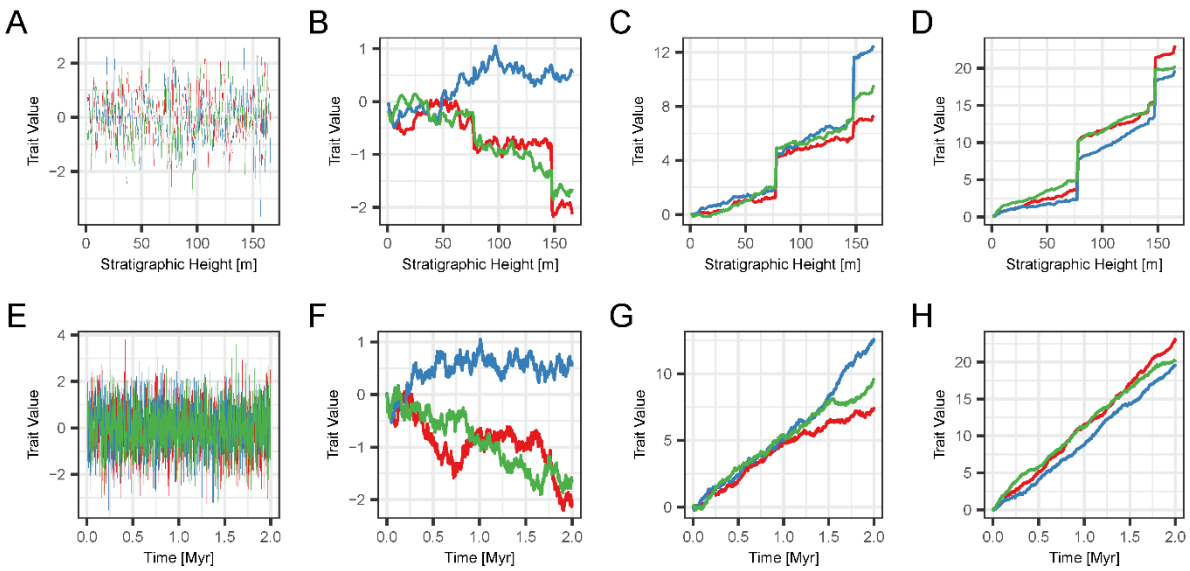

Figure 6: Differential preservation of different modes of evolution at the same location. First row: preservation of three lineages evolving according to the stasis (A), Brownian motion (B), weak Brownian drift (C), and strong Brownian drift (D) model 6 km offshore in scenario A. Second row: The corresponding true evolutionary history in the time domain. The change

in traits observable in fossil time series over a gap depends on the directionality of evolution and gap duration – Stasis is unaffected by the gaps, while the directional weak and strong Brownian drift displays jumps in phenotype over long gaps in the stratigraphic record.

Effects of completeness

Comparing identical modes of evolution at locations in the carbonate platform with similar stratigraphic completeness shows that stratigraphic completeness is not the major driver of stratigraphic control (*Figure 7*). If hiatuses are frequent and have similar duration, evolution in the stratigraphic domain is very similar to the time domain. If hiatuses are rare and long, traits can change significantly during the hiatus, leading to distinct differences in trait values from before to after the hiatus, generating dissimilarity between observations made in the stratigraphic domain and the "true" evolutionary history in the time domain.

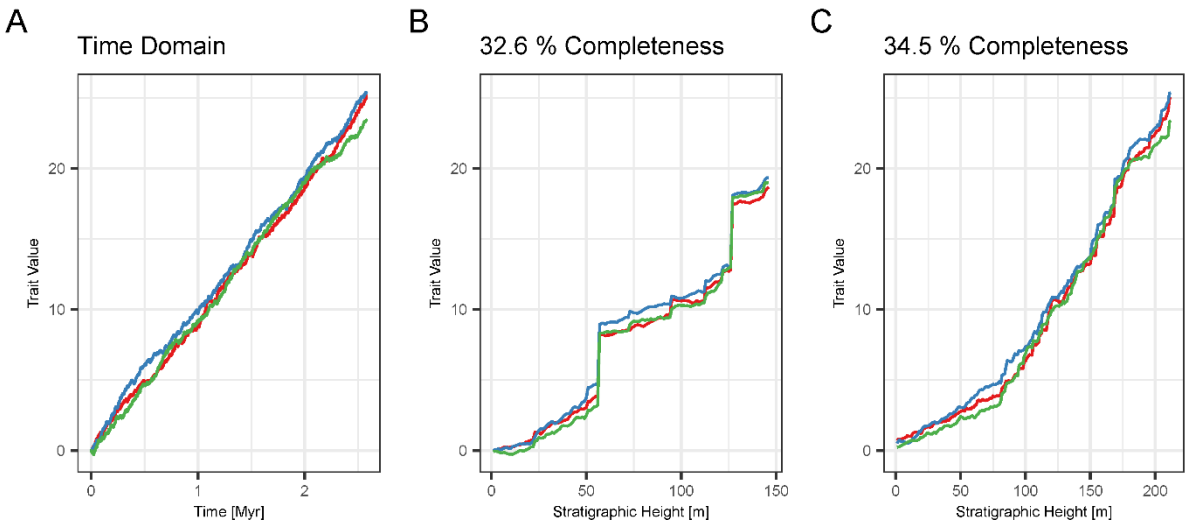

Figure 7 Effects of completeness vs. hiatus duration. Strong Brownian drift in the time domain (A), 2 km offshore in scenario A (B), and 6 km offshore in scenario B (C). In these sections, stratigraphic completeness differs by only 2 %, but preservation of the lineages

differs drastically due to the presence of few, but long hiatuses in scenario A generated by prolonged intervals of low sea level.

Spatial variation

The stratigraphic expression of trait evolution varies spatially within a carbonate platform (*Figure 5*). This is a direct result of the variability of stratigraphic completeness and hiatus frequency and duration along the onshore – offshore gradient (*Figure 4*). Because of the selective effects of stratigraphic architectures on the identifiability of directional evolution, spatial variability is most pronounced for weak and strong Brownian drift, and absent for stasis (*Supplementary Figures 14–19*).

Identification of modes of evolution

Surprisingly, neither the stratigraphic architecture driven by different sea-level histories nor the location along the onshore-offshore gradient has an influence on the best supported mode of evolution recovered by the paleoTS package (Hunt 2022). In the absence of stratigraphic effects, for adequate models, and under excellent sampling conditions (large number of specimens, low intrapopulation variability in traits, long, equidistant time series), tests fail to recover the correct (simulated) mode, and support for the correct mode of evolution decreases as time series length increases.

Test results in the stratigraphic domain

In the stratigraphic domain, the tests rarely yielded good support for the correct mode of evolution as defined by AICc weight > 0.9 (Portet 2020). This holds for both scenarios and all locations within the carbonate platform (Figures *8* and *9*, top rows).

Under simulated stasis (*Figure 8* A and *Figure 9* A), AICc weight values for the correct mode of evolution are low (between 0.1 and 0.25). The mode of evolution best supported is OU with AICc weight values larger than 0.75. This is independent of the location in the carbonate platform and scenario, except for scenario A, 12 km from shore. Here, AICc weights are around 0.5 for OU, 0.3 for stasis, and 0.1 for undirected random walk (URW).

Under simulated Brownian motion (*Figure 8* B and *Figure 9* B), AICc weights are moderate to low for all modes of evolution and have high dispersion. For the correct mode of evolution, URW, AICc weights are around 0.5. For stasis and OU, weights are around 0.25. This is independent of the location in the platform and of the scenario.

Results for simulations of weak and strong Brownian drift are very similar (*Figure 8* C, D and *Figure 9* C, D) . AICc weights for the correct mode of evolution (GRW) are around 0.75, whereas weights for OU are around 0.25. This holds across both scenarios and all locations in the platform, with the exception of scenario A, where minor support is found for undirected random walk (URW).

Overall, we find that neither the stratigraphic scenario nor the location in the platform (and, as consequence, stratigraphic completeness) have a strong effect on the test results.

Difference between the stratigraphic and time domains

After accounting for time series lengths, we find that test results with and without stratigraphic effects are very similar (*Figure 8* and *Figure 9*, comparison between the top and bottom rows). Differences in test results between the time domain and the stratigraphic domain are (1) weak support for URW under simulations of weak and strong Brownian drift in scenario A in the stratigraphic domain, while tests in the time domain yield weak support

for OU, and (2) lower dispersion of AICc weights in scenario B under simulations of weak and strong Brownian drift in the stratigraphic domain compared to the time domain.

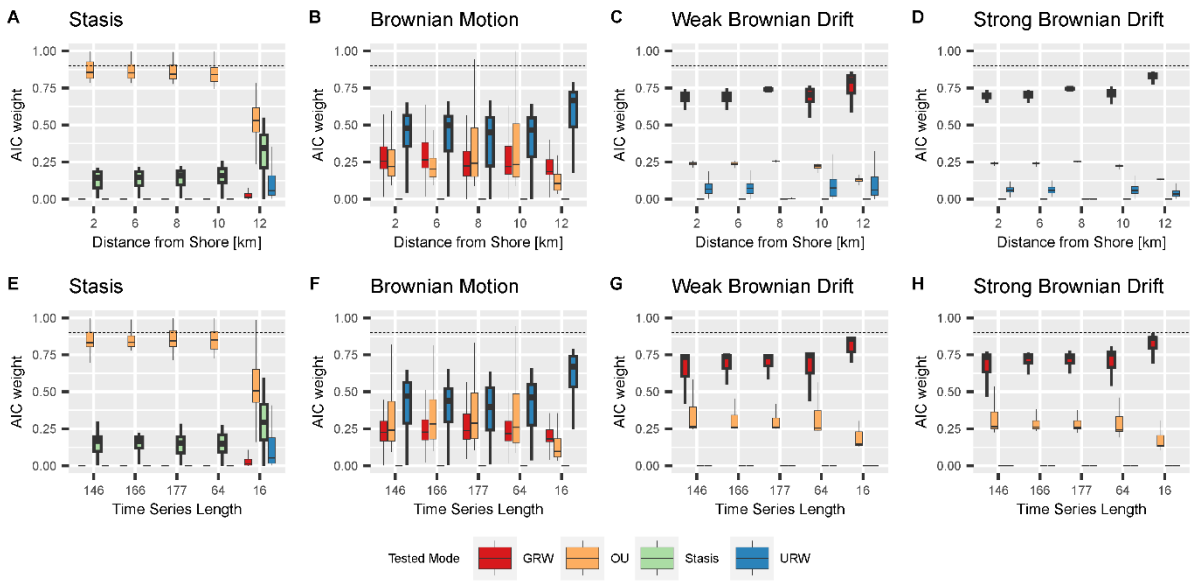

Figure 8 AICc weights of different modes of evolution in the stratigraphic domain in scenario A (first column, A-D) and for time series of equal length, but without stratigraphic biases (second column, E-H) under simulated stasis (first row), Brownian motion (second row), weak Brownian drift (third row), and strong Brownian drift (fourth row) at different positions in the platform. The dashed line indicates the threshold for good support (AICc weight > 0.9), the highlighted boxes are the correct test result for the simulated mode of evolution. Abbreviations for the tested modes are: GRW - general random walk, OU – Ornstein-Uhlenbeck process, Stasis – stasis, URW – undirected random walk.

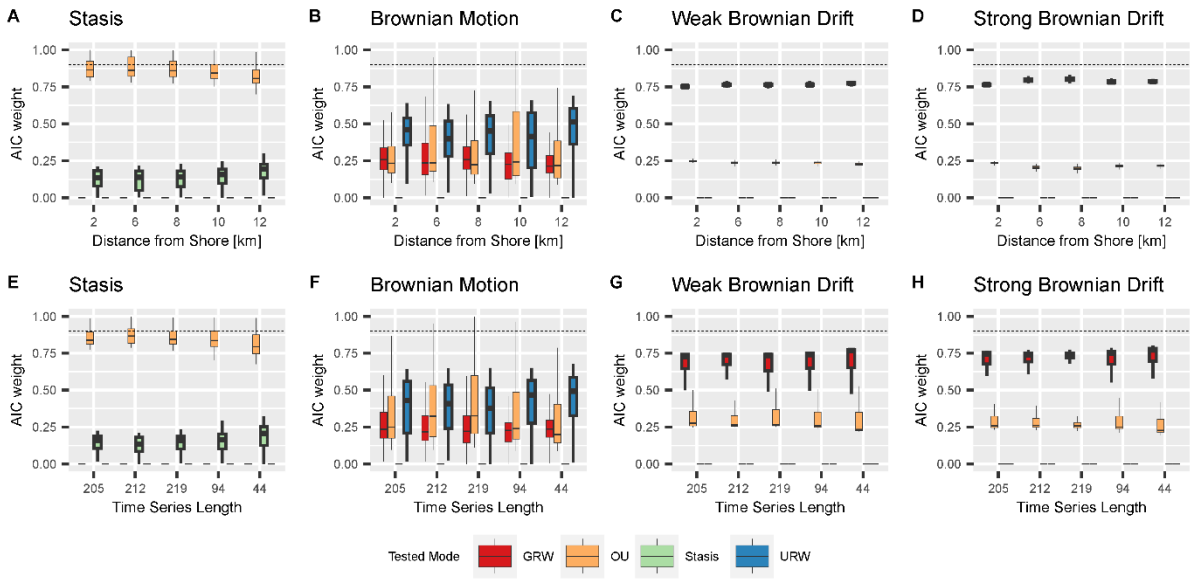

Figure 9: AICc weights of different modes of evolution in the stratigraphic domain in scenario B (first column, A-D) and for time series of equal length, but without stratigraphic biases (second column, E-H) under simulated stasis (first row), Brownian motion (second row), weak Brownian drift (third row), and strong Brownian drift (fourth row) at different positions in the platform. The dashed line indicates the threshold for good support (AICc weight > 0.9), the highlighted boxes are the correct test result for the simulated mode of evolution. Abbreviations for the tested modes are: GRW - general random walk, OU – Ornstein-Uhlenbeck process, Stasis – stasis, URW – undirected random walk

Test performance without stratigraphic biases

Testing for the mode of evolution without stratigraphic effects (i.e., in the time domain), we found that increasing sampling resolution (time series length) from 5 to 200 sampling points per simulation time reduced support for the correct mode of evolution as measured by AICc weights (*Figure 10*, *Supplementary Figure 20*). The results for scenario A (*Figure 10*) and scenario B (*Supplementary Figure 20*) were very similar, thus we focus on scenario A here.

Under simulated stasis (*Figure 10* A) AICc weights under low sampling effort (low number of sampling points) support URW, but this support quickly decreases and eventually vanishes for more than 20 sampling points. Support for the correct mode of evolution, stasis, peaks at 10 to 15 sampling points at a mean AICc weight of around 0.3, decreases monotonously as sampling effort is further increased and stabilizes around 0.15. Support for stasis never reaches the threshold value of 0.9. For time series with 15 or more samples, the AICc weights are the highest for OU and a portion of them exceeds the 0.9 threshold set as the criterion for good support. Support for GRW is low to none across all sampling resolutions.

Under simulated Brownian motion (*Figure 10* B), support for the correct mode of evolution (URW) is high for low sampling effort and mostly above 0.9 for series of five samples. It decreases monotonously to about 0.4 as sampling effort increases. Conversely, support for both OU and GRW increases monotonously until a mean of 0.25 as sampling effort increases. Support for stasis is low to null.

The results for weak and strong Brownian drift are similar (*Figure 10* C and D). Initial high (mostly >0.9 in the case of weak Brownian drift) support for URW vanishes with increasing number of sampling points. In all cases, support for OU increases with sampling effort. Support for the correct mode of evolution is high under low sampling effort, but decreases as sampling effort is increased, stabilizing at values around 0.7. Conversely, support for OU increases monotonously up to a mean of ca. 0.25 as sampling density is increased.

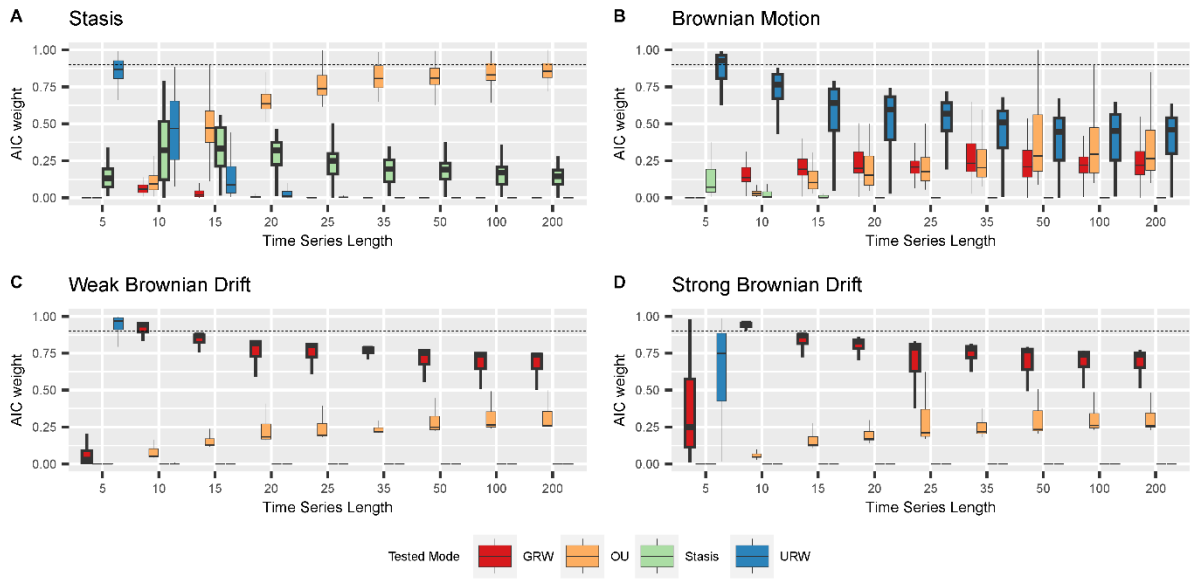

Figure 10: AICc weights of modes of evolution in the time domain under simulations of (A) stasis; (B) Brownian motion; (C) weak Brownian drift; (D) strong Brownian drift as a function of time series length. The sampled time interval is 2 Ma long (corresponding to the duration of scenario A), and is sampled with increasing frequency to reflect increasing sampling efforts. The dashed line indicates the threshold for good support (AICc weight > 0.9), the highlighted boxes are the correct test result for the simulated mode of evolution. Abbreviations for the tested modes are: GRW - general random walk, OU – Ornstein-Uhlenbeck process, Stasis – stasis, URW – undirected random walk

In summary, we find that in the absence of stratigraphic effects, the test correctly identified the mode of evolution in a portion of simulations when sampling intensity was very low (5 points) and only for Brownian motion and weak and strong Brownian drift. In the case of stasis, the correct mode has never been identified correctly if the threshold of 0.9 AICc weight is applied. Support for the correct mode of evolution decreases with time series length (Figure 10, Supplementary Figure 20). We thus reject the hypothesis that longer time series provide for better identification of the mode of evolution.

Discussion

Simulating the preservation of trait evolution in incomplete records formed in carbonate platforms, we found that – independently of the simulated mode of evolution, the presence or absence of stratigraphic effects, time series length, or location in the carbonate platform – tests for the mode of evolution failed to identify the correct mode of evolution according to the applied criterion of an AICc weight above 0.9 (Portet 2020), and support for the correct mode of evolution decreases with time series length. Visually, it is not stratigraphic completeness, but rather maximum hiatus duration and directionality of evolution that determine how much series of trait values derived from the stratigraphic record differ from the true evolutionary history.

Identifying the mode of evolution from fossil time series

We tested the hypothesis that the mode of evolution identified in a fossil time series obtained under the assumption of uninterrupted constant sediment accumulation (UCSA) is the same as the mode of the original time series in the time domain (the "true" evolutionary history). Although AICc weights of fossil time series derived under this assumption are identical to the AICc weights of the true evolutionary history, the tests fail to identify the correct mode of evolution based on the initially set criterion of an AICc weight larger than 0.9 for both stratigraphies studied here (*Figure 8* and *Figure 9*). Spatial variability in AICc weights in both carbonate platforms can be explained by differences in time series length resulting from differences in total accumulated thickness (*Figure 10*) rather than differences in stratigraphic architectures. Because of the inability of the test to identify the correct mode of evolution in the absence of any stratigraphic effects, our first hypothesis proved untestable with the

chosen approach. This test behavior is not due to our implementations of models of phenotypic evolution, as it persists when the internal simulation procedures provided by the paleoTS package are used (Hohmann and Hopkins 2024). This suggests that the test commonly yields an incorrect mode beyond our own study.

One possible explanation for the poor performance of the tests is that they are based on the Akaike information criterion (AICc), which measures relative fit of models to data. In the stratigraphic domain, none of the standard models of phenotypic evolution adequately describe the data. For example, the stratigraphic expression of the Brownian drift models resembles a Lévy process with rare, but large, jumps, rather than a Brownian drift (Landis, Schraiber, and Liang 2013; Landis and Schraiber 2017) (*Figure 5*). The issues arising from using AICc when models are not adequate are well known, and the adequacy of models of phyletic evolution has been discussed by Voje, Starrfelt and Liow (2018) and (Voje 2018), While they do not provide a general definition of what it means for a model to be adequate, they propose four tests for adequacy of models of phyletic evolution. In the time domain, we simulate lineages according to the same model we test for. Accordingly, the models must be adequate to describe the data as they generated the data in the first place, yet the test do not find good support for the correct mode of evolution. Another possible explanation for our inability to recover the correct simulated mode of evolution is that our definition of what it means to identify a model is too strict. Other criteria for good fit of a model are available (see e.g. Hopkins and Lidgard (2012)). However, this does not explain why, in the absence of stratigraphic biases, support for the correct mode as measured by AICc weights decrease as time series get longer (Figures *8* and *9*, lower row, *Figure 10*) — a qualitative behavior that is independent of any threshold for good model support.. For time series longer than 15 observations, support for the correct mode of evolution measured by AICc weights decreases (*Figure 10*, *Supplementary Figure 20*). Based on the probabilistic interpretation of the AICc

weights, this implies that the more samples are collected, the lower the probability that the data-generating model is the best model (Wagenmakers and Farrell 2004). One explanation for this apparent statistical paradox is that sampling a fixed time interval more frequently leads to diminishing returns on the information gain per sample. Intuitively, a few samples should be sufficient to separate directional evolution from stasis. However, this does not explain why AICc weights for the correct mode of evolution decrease instead of increase. Another potential explanation is that AIC is not consistent in the sense that if sample size grows, the probability that AIC identifies the true model does not approach one (Bozdogan 1987). It is unclear if the considered time series are long enough to display such asymptotic effects, and how they translate from AIC to AICc weights.

There is an ongoing debate about the usage of Ornstein-Uhlebeck models in phylogenetic comparative approaches (Ho and Ané 2014). Cooper et al. (2016) argue that OU processes are frequently incorrectly favored over simpler processes, while Grabowski et al. (2023) argue that this position is unsubstantiated. Using model selection, we found increasing support for OU as time series length increases. One potential explanation for this might be that OU processes contain many other processes as non-nested endmembers: With no selection, they are an unbiased random walk (Cooper et al. 2016), with a distant phenotypic optimum they resemble directional evolution with noise, and when starting at the phenotypic optimum they represent (autocorrelated) stasis.

Summarizing, we cannot resolve why tests do not identify the correct mode of evolution, longer time series yield lower AICc weights for the correct mode of evolution, and AICc weights are unaffected by stratigraphic biases.

Effects of stratigraphic incompleteness

The second hypothesis we examined was that the lower the stratigraphic completeness, the lower the chance to identify the correct mode of evolution from fossil time series that are constructed based on the assumption of uninterrupted constant sediment accumulation (UCSA). Because of the inability of the tests to find support for the correct mode of evolution in the absence of stratigraphic biases, we visually compared fossil time series derived under this assumption with time series of the true evolutionary history (*Figures 5, 6, 7, Supplementary Figures 4—19*).

Trait evolution reconstructed using simplified age-depth models varies spatially throughout the platform (*Figure 5*). Even if gaps are identified and their duration is known, evolutionary history coinciding with sea level drops is not preserved on the platform top. These time intervals can be recovered from the distal slope where sediment is continuously accreted during lowstand shedding (Figures 3 and 4)(Grammer and Ginsburg 1992). This highlights the importance to combine information from across the entire sedimentary basin to reconstruct past changes (Holland and Patzkowsky 2015; Zimmt et al. 2021), although the information gained from this approach might vary between carbonate platforms and siliciclastic systems.

We found that stratigraphic incompleteness (Dingus and Sadler 1982; Tipper 1987) is an imperfect measure to quantify the biasing effects of stratigraphic architectures on phenotypic evolution. For similar values of stratigraphic completeness, the stratigraphic expression of the same lineage can strongly differ. For example, fossil time series of Brownian drift reflect the underlying true evolutionary history, or display apparent jumps in phenotype for very close values of completeness (*Figure 7*). In addition, different modes of evolution are biased to a different degree when compared within the same section (and thus identical stratigraphic

723 effects and values of incompleteness). For example, in the same section Brownian drift
724 displays jumps in phenotype over gaps that will be misinterpreted as elevated rates of
725 evolution, whereas stasis remains unaffected (*Figure 6*). Multiple definitions of
726 incompleteness have been applied in stratigraphy (Anders, Krueger, and Sadler 1987), some
727 of which incorporate spatial variability (Straub and Foreman 2018). The definition of
728 stratigraphic completeness from (Tipper 1987) we used here reflects the intuition that gaps in
729 the geological record are the dominant factor that diminishes the quality of the fossil record.
730 While other definitions of incompleteness might be more suitable to quantifying stratigraphic
731 effects, our results show that the nature of the underlying evolutionary process cannot be
732 neglected when assessing the fidelity of the fossil record.
733 The change of phenotype over a gap is determined by gap duration and the change in
734 phenotype accumulated over this time interval. Stasis remains unaffected by gaps as it
735 accumulates no change in phenotype. On the other hand, for Brownian drift models, change
736 in phenotype is proportional to gap duration. Under the simplified assumptions on ADMs we
737 used, the change in traits over gaps will be taken at face value, giving trait evolution a distinct
738 punctuated look (Rita et al. 2019). While a precise modern formulation of the punctuated
739 equilibrium hypothesis (Gould and Eldredge 1972) is debated (see Pennell, Harmon, and
740 Uyeda (2014) vs. Lieberman and Eldredge (2014)), these jumps will be mistaken for intervals
741 with elevated evolutionary rates. Models of trait evolution that can incorporate punctuations
742 are Lévy processes, a class of stochastic process that combines gradual change with discrete
743 jumps (Landis, Schraiber, and Liang 2013; Landis and Schraiber 2017). In Lévy processes,
744 the jump components are random and have a Poisson structure. This provides a way to
745 separate between artefactual jumps introduced by simplistic ADMs and true elevated rates of
746 evolution: Artefactual jumps will coincide with gaps in the record or interval of reduced
747 sediment accumulation rates, and thus connect to external drivers of stratigraphic

architectures such as drops in sea level, rather than being random. Our study provides the stratigraphic null hypothesis that punctuated equilibrium should be more prevalent at times with low frequency sea-level changes. Combined, our results suggest that due to gaps in the stratigraphic record, fossil time series favor the recognition of both stasis and complex, punctuated models of evolution (Hunt, Hopkins, and Lidgard 2015).

Hoffman (1989) pointed out that most paleontological evidence for punctuated equilibrium comes from shallow water habitats (e.g. (Kelley 1983; Williamson 1981)), which they argue are more incomplete and thus favor the recognition of punctuations. This reflects the common idea that different environments have different incompleteness and thus different abilities to preserve evolution, as we have phrased in our hypothesis two. In contrast to this preconception, we found that many environments in a carbonate platform have very similar incompleteness and hiatus distributions (*Figure 4*). Due to the flat geometry of the platform, the separation is between platform top and slope, not the distinct environments within the platform such as forereef and lagoonal environments (Liu and Liu 2021). In addition, incompleteness alone is not sufficient to produce artefactual jumps in phenotype (*Figure 7*). Gaps in the stratigraphic record need to be sufficiently long compared to the rate of the evolutionary process, so that morphologies can change enough to display sufficient offset over the gap. This indicates that, rather than incompleteness, maximum hiatus duration is driving the discrepancy between fossil time series and the true evolutionary history.

Naturally, maximum hiatus duration is limited by stratigraphic completeness and the time covered by the section. But even under high incompleteness, as long as hiatus frequency is high and durations are short, the fossil record can still give a good insight into evolutionary history, and be adequate to test a wide range of evolutionary hypotheses (Paul 1992).

In our stratigraphic models, hiatus frequency and duration are determined by the frequency of sea-level fluctuations. Preservation of evolutionary history under high-frequency sea level

changes (scenario B) is good, while the large-scale fluctuations in scenario A result in prolonged gaps with a large impact on the recovery of the mode of evolution (Figure 7). This demonstrates that understanding controls on the spatial and temporal heterogeneity of the stratigraphic record is decisive in correctly interpreting the evolutionary history of a lineage in the realistic case of spatially constrained sampling.

Limitations of the simulated carbonate platforms

Carbonate platforms used for this research only approximate stratigraphic architectures that would form in nature. First, CarboCAT Lite does not include erosion other than sediment transport, which removes sediment locally immediately after its formation, but preserves the total volume of generated sediment in the platform. In real carbonate platforms, part of the carbonate sediment is dissolved and returned to the water column (e.g., Albright, Langdon, and Anthony (2013), exported into the ocean, and a part is chemically and mechanically eroded when it becomes emerged. The addition of erosion to the model would have the effect of enlarging the present gaps in the record and merging the shorter ones, resulting in fewer, longer gaps and lower stratigraphic completeness. Second, carbonate production curves used to inform the models reflect activities of carbonate factories under stable conditions. In reality, a regression would often lead to removal of carbonate producing organisms from the emerged area. The recolonization of this area by carbonate producers would lead to a lag in the factory resuming its activity. Carbonate production and preservation is also sensitive to diurnal, seasonal and long-term astronomical cycles, e.g. through temperature control over the proportion of precipitated aragonite (Balthasar and Cusack 2015) and the length of the vegetative season of calcifiers (e.g., Marshall and Clode 2004; Mancuso et al. 2019).
Our simulations did not include pelagic carbonate production, spatially heterogeneous subsidence or gravitational sediment transport, which may be crucial in the formation of

some empirical carbonate platforms (Masiero et al. 2020). We relied here on high benthic *in situ* production rates as the dominant driver of sediment accumulation in tropical, attached carbonate platforms. Thus, the stratigraphic architectures presented here are conceptual endpoints with exaggerated completeness.

The construction of our study follows the assumption of uninterrupted constant sediment accumulation (UCSA), which implies that the ages of fossils found at a given stratigraphic position correspond to the age of the stratum. This assumption is made in many simulation studies of trait evolution in the fossil record (e.g. (Hannisdal 2006)) and inherent to most methods that estimate age-depth models (e.g. (Parnell et al. 2008)). With this respect, our computer experiment is representative for such studies. However, time averaging and sedimentary condensation mean that typically more time is represented by fossils than by sedimentary strata (Kowalewski and Bambach 2008; Tomašových et al. 2022). On the other hand, fossils of different ages will be found at the same stratigraphic height, limiting the temporal resolution of evolutionary steps in the population's mean. Because of high *in situ* sediment production, timescales of time averaging identified in modern tropical carbonate environments are typically much shorter than rates of trait evolution, regardless of the way they are measured (Hunt 2012; Voje 2016; Kowalewski et al. 2017; Philip D. Gingerich 2019). This means that trait evolution reconstructed in tropical carbonate platforms may not be representative for other environments with lower rates of sediment accumulation or higher proportion of transported material.

Conclusions

We tested the hypothesis that the commonly employed approach to identifying the mode of evolution in fossil succession, i.e. linear projection of stratigraphic positions of occurrences

into the time domain without considering changes in sedimentation rate and gaps in the record, recovers the correct mode of evolution. In the course of the study, it appeared that the test is, in fact, not possible, because the commonly used test does not recover the correct mode of evolution even from complete data (i.e. if the time domain is sampled without gaps or distortions). In both considered situations: original mode of evolution in time and its distorted record in the stratigraphic domain, the tests did not yield clear support for the correct mode of evolution. Our findings differ from those of Hannisdal (2006) , who found (using a different approach but asking the same question) that incomplete sampling in the stratigraphic record may result in all other modes of evolution being identified as stasis. In our study, we never found strong support for stasis.

Our findings did not vary substantially between two stratigraphic architectures with varying gap distributions and degrees of stratigraphic completeness. The difficulty in identifying the correct mode of evolution holds across depositional environments. This is counterintuitive, as deeper environments are often assumed to be more complete and therefore more suitable for sampling fossil series for evolutionary studies. Increasing the number of observations (i.e. sampling intensity, length of the fossil series) did not improve the identification of the mode of evolution, but rather worsened it.

Our study was motivated by improving the recovery of evolutionary information from highly-resolve fossil successions, particularly at microevolutionary scales. We are convinced that such successions can aliment models and understanding that is not accessible to exclusively neontological methodologies, as illustrated by e.g. Hopkins and Lidgard (2012); Voje (2016); Petryshen et al. (2020). Our contribution is the use of stratigraphic forward modeling to ground-truth the methodologies serving this palaeobiological research program. Forward modeling allows rigorous testing of concerns that the fossil record is too distorted, or too incomplete, to answer (micro)evolutionary questions. The conclusion of our study is,

unexpectedly, that it may be the analytical methods that limit our use of fossil data, rather than the quality of the fossil record. The proliferation of ever better stratigraphic forward models (e.g. CarboCAT (P. M. Burgess 2013; Masiero et al. 2020), SedFlux (Hutton and Syvitski 2008), strataR (Holland 2022), CarboKitten.jl (Hidding et al., 2024)) opens the possibility to validate these methods and improve our understanding of the fossil record.

Acknowledgements

The authors would like to thank Melanie Hopkins, Katharine Loughney, Bjarte Hannisdal, Kenneth De Baets, and one anonymous reviewer for their feedback on the manuscript. The authors would also like to thank the Peer Community In (PCI) Paleontology.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The dataset supporting the conclusions of this article is available in the Zenodo repository, https://doi.org/10.17605/OSF.IO/ZBPWA (Hohmann, Koelewijn, and Jarochowska 2023a). All code used can be found in Hohmann, Koelewijn, and Jarochowska (2023b) and is accessible via https://doi.org/10.5281/zenodo.10390267. Supplementary figures can be found in the supplementary materials.

Competing interests

The authors declare that they have no competing interests

Funding

Funded by the European Union (ERC, MindTheGap, StG project no 101041077). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

Authors' contributions

According to the CRediT taxonomy

Niklas Hohmann: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – Original Draft, Writing – Review & Editing, Visualization.

Joël R. Koelewijn: Software, Validation, Formal analysis, Investigation, Visualization.

Peter Burgess: Writing – Review & Editing, Software.

Emilia Jarochowska: Conceptualization, Methodology, Validation, Visualization, Writing – Review & Editing, Supervision, Project administration, Funding acquisition.

References

Albright, R., C. Langdon, and K. R. N. Anthony. 2013. "Dynamics of Seawater Carbonate Chemistry, Production, and Calcification of a Coral Reef Flat, Central Great Barrier Reef." *Biogeosciences* 10 (10): 6747–58. https://doi.org/10.5194/bg-10-6747-2013.

Anders, Mark H., Scot W. Krueger, and Peter M. Sadler. 1987. "A New Look at Sedimentation Rates and the Completeness of the Stratigraphic Record." *The Journal of Geology* 95 (1): 1–14. https://doi.org/10.1086/629103.

Aze, Tracy, Thomas H. G. Ezard, Andy Purvis, Helen K. Coxall, Duncan R. M. Stewart, Bridget S. Wade, and Paul N. Pearson. 2011. "A Phylogeny of Cenozoic Macroperforate Planktonic Foraminifera from Fossil Data." *Biological Reviews* 86 (4): 900–927. https://doi.org/10.1111/j.1469-185X.2011.00178.x.

Balthasar, Uwe, and Maggie Cusack. 2015. "Aragonite-Calcite Seas—Quantifying the Gray Area." *Geology* 43 (2): 99–102. https://doi.org/10.1130/G36293.1.

Barido-Sottani, Joëlle, Alexander Pohle, Kenneth De Baets, Duncan Murdock, and Rachel C. M. Warnock. 2023. "Putting the F into FBD Analysis: Tree Constraints or Morphological Data?" *Palaeontology* 66 (6): e12679. https://doi.org/10.1111/pala.12679.

Barido-Sottani, Joëlle, Nina M. A. van Tiel, Melanie J. Hopkins, David F. Wright, Tanja Stadler, and Rachel C. M. Warnock. 2020. "Ignoring Fossil Age Uncertainty Leads to Inaccurate Topology and Divergence Time Estimates in Time Calibrated Tree Inference." *Frontiers in Ecology and Evolution* 8. https://www.frontiersin.org/articles/10.3389/fevo.2020.00183.

Beran, Jan. 2017. *Mathematical Foundations of Time Series Analysis: A Concise Introduction*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-74380-6.

Bookstein, Fred L. 1987. "Random Walk and the Existence of Evolutionary Rates." *Paleobiology* 13 (4): 446–64. https://doi.org/10.1017/S0094837300009039.

Bosscher, Hemmo, and Wolfgang Schlager. 1992. "Computer Simulation of Reef Growth." *Sedimentology* 39 (3): 503–12. https://doi.org/10.1111/j.1365-3091.1992.tb02130.x.

Bozdogan, Hamparsum. 1987. "Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytical Extensions." *Psychometrika* 52 (3): 345–70. https://doi.org/10.1007/BF02294361.

Burgess, Peter. (2023) 2023. "CarboCAT Lite." MATLAB. https://github.com/MindTheGap-ERC/CarboCATLite.

Burgess, Peter M. 2013. "CarboCAT: A Cellular Automata Model of Heterogeneous Carbonate Strata." *Computers & Geosciences*, Modeling for Environmental Change, 53 (April): 129–40. https://doi.org/10.1016/j.cageo.2011.08.026.

Clyde, William C., and Philip D. Gingerich. 1994. "Rates of Evolution in the Dentition of Early Eocene Cantius: Comparison of Size and Shape." *Paleobiology* 20 (4): 506–22. https://doi.org/10.1017/S0094837300012963.

Cooper, Natalie, Gavin H. Thomas, Chris Venditti, Andrew Meade, and Rob P. Freckleton. 2016. "A Cautionary Note on the Use of Ornstein Uhlenbeck Models in Macroevolutionary Studies." *Biological Journal of the Linnean Society* 118 (1): 64–77. https://doi.org/10.1111/bij.12701.

Darwin, Charles. 1859. *On the Origin of Species: A Facsimile of the First Edition*. Harvard University Press.

Dingus, Lowell, and Peter M. Sadler. 1982. "The Effects of Stratigraphic Completeness on Estimates of Evolutionary Rates." *Systematic Biology* 31 (4): 400–412. https://doi.org/10.1093/sysbio/31.4.400.

Donoghue, Philip C. J. 2001. "Conodonts Meet Cladistics: Recovering Relationships and Assessing the Completeness of the Conodont Fossil Record." *Palaeontology* 44 (1): 65–93. https://doi.org/10.1111/1475-4983.00170.

Dzik, Jerzy. 1985. "Typologic versus Population Concepts of Chronospecies : Implications for Ammonite Biostratigraphy." *Acta Palaeontologica Polonica* 30 (1–2): 71–92.

———. 1995. "Range-Based Biostratigraphy and Evolutionary Geochronology." *Paleopelagos Special Publication* 1 (94): 121–28.

———. 1999. "Relationship between Rates of Speciation and Phyletic Evolution: Stratophenetic Data on Pelagic Conodont Chordates and Benthic Ostracods." *Geobios* 32 (2): 205–21. https://doi.org/10.1016/S0016-6995(99)80033-3.

———. 2005. "The Chronophyletic Approach: Stratophenetics Facing an Incomplete Fossil Record." *Special Papers in Palaeontology* 73: 159–83.

Gingerich, P.D. 1983. “Rates of Evolution: Effects of Time and Temporal Scaling.” *Science* 222: 159–62.

———. 2001. “Rates of Evolution on the Time Scale of the Evolutionary Process.” In *Microevolution Rate, Pattern, Process*, edited by A. P. Hendry and M. T. Kinnison, 127–44. Contemporary Issues in Genetics and Evolution. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-010-0585-2_9.

Gingerich, Philip D. 2019. *Rates of Evolution: A Quantitative Synthesis*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316711644.

Gingerich, Philip D., and Margaret Schoeninger. 1977. “The Fossil Record and Primate Phylogeny.” *Journal of Human Evolution* 6 (5): 483–505. https://doi.org/10.1016/S0047-2484(77)80059-6.

Gould, Stephen Jay, and Niles Eldredge. 1972. “Punctuated Equilibria: An Alternative to Phyletic Gradualism.” *Models in Paleobiology* 1972: 82–115.

Grabowski, Mark, Jason Pienaar, Kjetil L Voje, Staffan Andersson, Jesualdo Fuentes-González, Bjørn T Kopperud, Daniel S Moen, Masahito Tsuboi, Josef Uyeda, and Thomas F Hansen. 2023. “A Cautionary Note on ‘A Cautionary Note on the Use of Ornstein Uhlenbeck Models in Macroevolutionary Studies.’” *Systematic Biology* 72 (4): 955–63. https://doi.org/10.1093/sysbio/syad012.

Grammer, G. Michael, and Robert N. Ginsburg. 1992. “Highstand versus Lowstand Deposition on Carbonate Platform Margins: Insight from Quaternary Foreslopes in the Bahamas.” *Marine Geology* 103 (1): 125–36. https://doi.org/10.1016/0025-3227(92)90012-7.

Grantham, Todd. 2004. “The Role of Fossils in Phylogeny Reconstruction: Why Is It so Difficult to Integrate Paleobiological and Neontological Evolutionary Biology?” *Biology and Philosophy* 19 (5): 687–720. https://doi.org/10.1007/s10539-005-0370-z.

Guenser, Pauline, Rachel C. M. Warnock, Walker Pett, Philip C. J. Donoghue, and Emilia Jarochowska. 2021. “Does Time Matter in Phylogeny? A Perspective from the Fossil Record.” bioRxiv. https://doi.org/10.1101/2021.06.11.445746.

Hannisdal, Bjarte. 2006. “Phenotypic Evolution in the Fossil Record: Numerical Experiments.” *The Journal of Geology* 114 (2): 133–53. https://doi.org/10.1086/499569.

Haug, Joachim Tobias, and Carolin Haug. 2017. “Species, Populations and Morphotypes through Time − Challenges and Possible Concepts.” Edited by Sylvain Charbonnier. *BSGF - Earth Sciences Bulletin* 188 (3): 20. https://doi.org/10.1051/bsgf/2017181.

Hidding, Johan, Emilia Jarochowska, Xianyi Liu, Peter Burgess, Niklas Hohmann, and Hanno Spreeuw. 2024. “CarboKitten.Jl.”

Ho, Lam Si Tung, and Cécile Ané. 2014. “Intrinsic Inference Difficulties for Trait Evolution with Ornstein-Uhlenbeck Models.” *Methods in Ecology and Evolution* 5 (11): 1133–46. https://doi.org/10.1111/2041-210X.12285.

Hoffman, Antoni. 1989. *Arguments on Evolution: A Paleontologist’s Perspective*. New York: Oxford University Press.

Hohmann, Niklas. 2021. “Incorporating Information on Varying Sedimentation Rates into Paleontological Analyses.” *PALAIOS* 36 (2): 53–67. https://doi.org/10.2110/palo.2020.038.

Hohmann, Niklas, and Melanie Hopkins. 2024. “paleoTS_test: Examine paleoTS Model Selection Performance with Time Series Length.” [object Object]. https://doi.org/10.5281/ZENODO.10843692.

Hohmann, Niklas, Jan R. Koelewijn, and Emilia Jarochowska. 2023a. “Identification of the Mode of Evolution in Incomplete Carbonate Successions - Supporting Data.” Open Science Framework. https://doi.org/10.17605/OSF.IO/ZBPWA.

———. 2023b. “Identification of the Mode of Evolution in Incomplete Carbonate Successions - Supporting Code.” Zernodo. https://doi.org/10.5281/zenodo.10390266.

Holland, Steven M. 2022. “The Structure of the Nonmarine Fossil Record: Predictions from a Coupled Stratigraphic–Paleoecological Model of a Coastal Basin.” *Paleobiology* 48 (3): 372–96. https://doi.org/10.1017/pab.2022.5.

Holland, Steven M., and Mark E. Patzkowsky. 2015. “The Stratigraphy of Mass Extinction.” *Palaeontology* 58 (5): 903–24. https://doi.org/10.1111/pala.12188.

Hopkins, Melanie J., and Scott Lidgard. 2012. “Evolutionary Mode Routinely Varies among Morphological Traits within Fossil Species Lineages.” *Proceedings of the National Academy of Sciences* 109 (50): 20520–25. https://doi.org/10.1073/pnas.1209901109.

Hunt, Gene. 2006. “Fitting and Comparing Models of Phyletic Evolution: Random Walks and Beyond.” *Paleobiology* 32 (4): 578–601. https://doi.org/10.1666/05070.1.

———. 2010. “Evolution in Fossil Lineages: Paleontology and The Origin of Species.” *The American Naturalist* 176 (S1): S61–76. https://doi.org/10.1086/657057.

———. 2012. “Measuring Rates of Phenotypic Evolution and the Inseparability of Tempo and Mode.” *Paleobiology* 38 (3): 351–73.

———. 2022. “paleoTS: Analyze Paleontological Time-Series.” https://CRAN.R-project.org/package=paleoTS.

Hunt, Gene, Melanie J. Hopkins, and Scott Lidgard. 2015. “Simple versus Complex Models of Trait Evolution and Stasis as a Response to Environmental Change.” *Proceedings of the National Academy of Sciences of the United States of America* 112 (16): 4885–90. https://doi.org/10.1073/pnas.1403662111.

Hunt, Gene, and Kaustuv Roy. 2006. “Climate Change, Body Size Evolution, and Cope’s Rule in Deep-Sea Ostracodes.” *Proceedings of the National Academy of Sciences* 103 (5): 1347–52. https://doi.org/10.1073/pnas.0510550103.

Hutton, Eric W. H., and James P. M. Syvitski. 2008. “Sedflux 2.0: An Advanced Process-Response Model That Generates Three-Dimensional Stratigraphy.” *Computers & Geosciences*, Predictive Modeling in Sediment Transport and Stratigraphy, 34 (10): 1319–37. https://doi.org/10.1016/j.cageo.2008.02.013.

Jablonski, David, Kaustuv Roy, and James W. Valentine. 2006. “Out of the Tropics: Evolutionary Dynamics of the Latitudinal Diversity Gradient.” *Science* 314 (5796): 102–6. https://doi.org/10.1126/science.1130880.

Jones, David. 2009. “Directional Evolution in the Conodont Pterospathodus.” *Paleobiology* 35 (3): 413–31.

Kelley, Patricia H. 1983. “Evolutionary Patterns of Eight Chesapeake Group Molluscs: Evidence for the Model of Punctuated Equilibria.” *Journal of Paleontology* 57 (3): 581–98.

Kemp, David B. 2012. “Stochastic and Deterministic Controls on Stratigraphic Completeness and Fidelity.” *International Journal of Earth Sciences* 101 (8): 2225–38. https://doi.org/10.1007/s00531-012-0784-1.

Kidwell, Susan M., and Karl W. Flessa. 1996. “The Quality of the Fossil Record: Populations, Species, and Communities.” *Annual Review of Earth and Planetary Sciences* 24 (1): 433–64. https://doi.org/10.1146/annurev.earth.24.1.433.

Kidwell, Susan M., and Steven M. Holland. 2002. “The Quality of the Fossil Record: Implications for Evolutionary Analyses.” *Annual Review of Ecology and Systematics* 33 (1): 561–88. https://doi.org/10.1146/annurev.ecolsys.33.030602.152151.

Kowalewski, Michał, and Richard K. Bambach. 2008. “The Limits of Paleontological Resolution.” In *High-Resolution Approaches in Stratigraphic Paleontology*, edited by P. J. Harries, 21:1–48. Topics in Geobiology. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-1-4020-9053-0_1.

Kowalewski, Michał, Sahale Casebolt, Quan Hua, Katherine E. Whitacre, Darrell S. Kaufman, and Matthew A. Kosnik. 2017. "One Fossil Record, Multiple Time Resolutions: Disparate Time-Averaging of Echinoids and Mollusks on a Holocene Carbonate Platform." *Geology* 46 (1): 51–54. https://doi.org/10.1130/G39789.1.

Lande, Russell. 1976. "Natural Selection and Random Genetic Drift in Phenotypic Evolution." *Evolution* 30 (2): 314–34. https://doi.org/10.2307/2407703.

Landis, Michael J., and Joshua G. Schraiber. 2017. "Pulsed Evolution Shaped Modern Vertebrate Body Sizes." *Proceedings of the National Academy of Sciences* 114 (50): 13224–29. https://doi.org/10.1073/pnas.1710920114.

Landis, Michael J., Joshua G. Schraiber, and Mason Liang. 2013. "Phylogenetic Analysis Using Lévy Processes: Finding Jumps in the Evolution of Continuous Traits." *Systematic Biology* 62 (2): 193–204. https://doi.org/10.1093/sysbio/sys086.

Lieberman, Bruce S., and Niles Eldredge. 2014. "What Is Punctuated Equilibrium? What Is Macroevolution? A Response to Pennell et Al." *Trends in Ecology & Evolution* 29 (4): 185–86. https://doi.org/10.1016/j.tree.2014.02.005.

Liu, Jianliang, and Keyu Liu. 2021. "Estimating Stratal Completeness of Carbonate Deposition via Process-Based Stratigraphic Forward Modeling." *Science China Earth Sciences* 64 (2): 253–59. https://doi.org/10.1007/s11430-020-9660-8.

MacLeod, Norman. 1991. "Punctuated Anagenesis and the Importance of Stratigraphy to Paleobiology." *Paleobiology* 17 (2): 167–88. https://doi.org/10.1017/S0094837300010472.

Malmgren, Björn A., W. A. Berggren, and G. P. Lohmann. 1983. "Evidence for Punctuated Gradualism in the Late Neogene Globorotalia Tumida Lineage of Planktonic Foraminifera." *Paleobiology* 9 (4): 377–89. https://doi.org/10.1017/S0094837300007843.

Mancuso, Arianna, Marco Stagioni, Fiorella Prada, Daniele Scarponi, Corrado Piccinetti, and Stefano Goffredo. 2019. "Environmental Influence on Calcification of the Bivalve Chamelea Gallina along a Latitudinal Gradient in the Adriatic Sea." *Scientific Reports* 9 (1): 11198. https://doi.org/10.1038/s41598-019-47538-1.

Marshall, A.T., and P. Clode. 2004. "Calcification Rate and the Effect of Temperature in a Zooxanthellate and an Azooxanthellate Scleractinian Reef Coral." *Coral Reefs* 23 (2): 218–24. https://doi.org/10.1007/s00338-004-0369-y.

Masiero, Isabella, Estanislao Kozlowski, Georgios Antonatos, Haiwei Xi, and Peter Burgess. 2020. "Numerical Stratigraphic Forward Models as Conceptual Knowledge Repositories and Experimental Tools: An Example Using a New Enhanced Version of CarboCAT." *Computers & Geosciences* 138 (May): 104453. https://doi.org/10.1016/j.cageo.2020.104453.

McNeill, Donald F. 2005. "Accumulation Rates from Well-Dated Late Neogene Carbonate Platforms and Margins." *Sedimentary Geology*, Sedimentology in the 21st Century - A Tribute to Wolfgang Schlager, 175 (1): 73–87. https://doi.org/10.1016/j.sedgeo.2004.12.032.

Miller, Kenneth G., James V. Browning, W. John Schmelz, Robert E. Kopp, Gregory S. Mountain, and James D. Wright. 2020. "Cenozoic Sea-Level and Cryospheric Evolution from Deep-Sea Geochemical and Continental Margin Records." *Science Advances* 6 (20): eaaz1346. https://doi.org/10.1126/sciadv.aaz1346.

Mitchell, Jonathan S, Rampal S Etienne, and Daniel L Rabosky. 2019. "Inferring Diversification Rate Variation from Phylogenies with Fossils." *Systematic Biology* 68 (1): 1–18. https://doi.org/10.1093/sysbio/syy035.

Mongiardino Koch, Nicolás, Russell J. Garwood, and Luke A. Parry. 2021. "Fossils Improve Phylogenetic Analyses of Morphological Characters." *Proceedings of the Royal*

Society B: Biological Sciences 288 (1950): 20210044. https://doi.org/10.1098/rspb.2021.0044.

Parnell, A. C., J. Haslett, J. R. M. Allen, C. E. Buck, and B. Huntley. 2008. "A Flexible Approach to Assessing Synchroneity of Past Events Using Bayesian Reconstructions of Sedimentation History." *Quaternary Science Reviews* 27 (19): 1872–85. https://doi.org/10.1016/j.quascirev.2008.07.009.

Patterson, Colin. 1981. "Significance of Fossils in Determining Evolutionary Relationships." *Annual Review of Ecology and Systematics* 12: 195–223.

Paul, Christopher R. C. 1992. "How Complete Does the Fossil Record Have to Be?" *Revista Española de Paleontología* 7 (2): 127–33.

Pearson, Paul N. 1992. "Survivorship Analysis of Fossil Taxa When Real-Time Extinction Rates Vary: The Paleogene Planktonic Foraminifera." *Paleobiology* 18 (2): 115–31. https://doi.org/10.1017/S0094837300013920.

Pennell, Matthew W., Luke J. Harmon, and Josef C. Uyeda. 2014. "Is There Room for Punctuated Equilibrium in Macroevolution?" *Trends in Ecology & Evolution* 29 (1): 23–32. https://doi.org/10.1016/j.tree.2013.07.004.

Petryshen, W., C. M. Henderson, K. De Baets, and E. Jarochowska. 2020. "Evidence of Parallel Evolution in the Dental Elements of Sweetognathus Conodonts." *Proceedings of the Royal Society B: Biological Sciences* 287 (1939): 20201922. https://doi.org/10.1098/rspb.2020.1922.

Platen, Eckhard, and Nicola Bruti-Liberati. 2010. *Numerical Solution of Stochastic Differential Equations with Jumps in Finance*. Vol. 64. Stochastic Modelling and Applied Probability. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-13694-8.

Plint, A. Guy, and Dag Nummedal. 2000. "The Falling Stage Systems Tract: Recognition and Importance in Sequence Stratigraphic Analysis." *Geological Society, London, Special Publications* 172 (1): 1–17. https://doi.org/10.1144/GSL.SP.2000.172.01.01.

Portet, Stéphanie. 2020. "A Primer on Model Selection Using the Akaike Information Criterion." *Infectious Disease Modelling* 5: 111–28. https://doi.org/10.1016/j.idm.2019.12.010.

Quental, Tiago B., and Charles R. Marshall. 2010. "Diversity Dynamics: Molecular Phylogenies Need the Fossil Record." *Trends in Ecology & Evolution* 25 (8): 434–41. https://doi.org/10.1016/j.tree.2010.05.002.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rita, Patrícia, Paulina Nätscher, Luís V. Duarte, Robert Weis, and Kenneth De Baets. 2019. "Mechanisms and Drivers of Belemnite Body-Size Dynamics across the Pliensbachian–Toarcian Crisis." *Royal Society Open Science* 6 (12): 190494. https://doi.org/10.1098/rsos.190494.

Sheets, H. David, and Charles E. Mitchell. 2001. "Uncorrelated Change Produces the Apparent Dependence of Evolutionary Rate on Interval." *Paleobiology* 27 (3): 429–45. https://doi.org/10.1666/0094-8373(2001)027<0429:UCPTAD>2.0.CO;2.

Springer, Mark S. 1995. "Molecular Clocks and the Incompleteness of the Fossil Record." *Journal of Molecular Evolution* 41 (5): 531–38. https://doi.org/10.1007/BF00175810.

Stadler, Tanja, Alexandra Gavryushkina, Rachel C. M. Warnock, Alexei J. Drummond, and Tracy A. Heath. 2018. "The Fossilized Birth-Death Model for the Analysis of Stratigraphic Range Data under Different Speciation Modes." *Journal of Theoretical Biology* 447 (June): 41–55. https://doi.org/10.1016/j.jtbi.2018.03.005.

Straub, Kyle M., and Brady Z. Foreman. 2018. "Geomorphic Stasis and Spatiotemporal Scales of Stratigraphic Completeness." *Geology* 46 (4): 311–14. https://doi.org/10.1130/G40045.1.

Strömberg, Caroline A. E. 2006. "Evolution of Hypsodonty in Equids: Testing a Hypothesis of Adaptation." *Paleobiology* 32 (2): 236–58. https://doi.org/10.1666/0094-8373(2006)32[236:EOHIET]2.0.CO;2.

Symonds, Matthew R. E., and Adnan Moussalli. 2011. "A Brief Guide to Model Selection, Multimodel Inference and Model Averaging in Behavioural Ecology Using Akaike's Information Criterion." *Behavioral Ecology and Sociobiology* 65 (1): 13–21. https://doi.org/10.1007/s00265-010-1037-6.

Tipper, John C. 1987. "Estimating Stratigraphic Completeness." *The Journal of Geology* 95 (5): 710–15.

Tomašových, Adam, Ivo Gallmetzer, Alexandra Haselmair, and Martin Zuschin. 2022. "Inferring Time Averaging and Hiatus Durations in the Stratigraphic Record of High-frequency Depositional Sequences." Edited by Christian Betzler. *Sedimentology* 69 (3): 1083–1118. https://doi.org/10.1111/sed.12936.

Voje, Kjetil Lysne. 2016. "Tempo Does Not Correlate with Mode in the Fossil Record." *Evolution* 70 (12): 2678–89. https://doi.org/10.1111/evo.13090.

———. 2018. "Assessing Adequacy of Models of Phyletic Evolution in the Fossil Record." *Methods in Ecology and Evolution* 9 (12): 2402–13. https://doi.org/10.1111/2041-210X.13083.

———. 2020. "Testing Eco-Evolutionary Predictions Using Fossil Data: Phyletic Evolution Following Ecological Opportunity*." *Evolution* 74 (1): 188–200. https://doi.org/10.1111/evo.13838.

Voje, Kjetil Lysne, Jostein Starrfelt, and Lee Hsiang Liow. 2018. "Model Adequacy and Microevolutionary Explanations for Stasis in the Fossil Record." *The American Naturalist* 191 (4): 509–23. https://doi.org/10.1086/696265.

Wagenmakers, Eric-Jan, and Simon Farrell. 2004. "AIC Model Selection Using Akaike Weights." *Psychonomic Bulletin & Review* 11 (1): 192–96. https://doi.org/10.3758/BF03206482.

Warnock, Rachel C. M., Tracy A. Heath, and Tanja Stadler. 2020. "Assessing the Impact of Incomplete Species Sampling on Estimates of Speciation and Extinction Rates." *Paleobiology* 46 (2): 137–57. https://doi.org/10.1017/pab.2020.12.

Warrlich, G. M. D. 2000. "3D Computer Forward Modelling of Carbonate Platform Evolution." PhD, London: Royal Holloway University of London.

Wehmiller, John F., Daniel F. Belknap, Brian S. Boutin, June E. Mirecki, Stephen D. Rahaim, and Linda L. York. 1988. "A Review of the Aminostratigraphy of Quaternary Mollusks from United States Atlantic Coastal Plain Sites." In *Dating Quaternary Sediments*, edited by Don J. Easterbrook, 227:0. Geological Society of America. https://doi.org/10.1130/SPE227-p69.

Wilkinson, Bruce H., Bradley N. Opdyke, and Thomas J. Algeo. 1991. "Time Partitioning in Cratonic Carbonate Rocks." *Geology* 19 (11): 1093–96. https://doi.org/10.1130/0091-7613(1991)019<1093:TPICCR>2.3.CO;2.

Williamson, P. G. 1981. "Palaeontological Documentation of Speciation in Cenozoic Molluscs from Turkana Basin." *Nature* 293 (5832): 437–43. https://doi.org/10.1038/293437a0.

Wills, Matthew A. 1999. "Congruence Between Phylogeny and Stratigraphy: Randomization Tests and the Gap Excess Ratio." Edited by C. Marshall. *Systematic Biology* 48 (3): 559–80. https://doi.org/10.1080/106351599260148.

Wright, April M., David W. Bapst, Joëlle Barido-Sottani, and Rachel C.M. Warnock. 2022. "Integrating Fossil Observations Into Phylogenetics Using the Fossilized Birth–Death Model." *Annual Review of Ecology, Evolution, and Systematics* 53 (1): 251–73. https://doi.org/10.1146/annurev-ecolsys-102220-030855.

Zimmt, Joshua B., Steven M. Holland, Seth Finnegan, and Charles R. Marshall. 2021. "Recognizing Pulses of Extinction from Clusters of Last Occurrences." *Palaeontology* 64 (1): 1–20. https://doi.org/10.1111/pala.12505.