Revisiting a Landmark Study System: No Evidence for a Punctuated Mode of Evolution in *Metrarabdotos*

Kjetil Lysne Voje,[1,*] Emanuela Di Martino,[2] and Arthur Porto[1]

1. Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, Oslo, Norway; 2. Natural History Museum, University of Oslo, Blindern, PO Box 1172, Oslo 0318, Norway

Submitted March 18, 2019; Accepted November 14, 2019; Electronically published March 17, 2020

Online enhancements: supplemental PDF, 3D figure rotations. *Dryad data:* https://doi.org/10.5061/dryad.t4b8gthxm.

abstract: Is speciation generally a "special time" in morphological evolution, or are lineage-splitting events just "more of the same" where the end product happens to be two separate lineages? Data on evolutionary dynamics during anagenetic and cladogenetic events among closely related lineages within a clade are rare, but the fossil record of the bryozoan genus *Metrarabdotos* is considered a textbook example of a clade where speciation causes rapid evolutionary change against a backdrop of morphological stasis within lineages. Here, we point to some methodological and measurement theoretical issues in the original work on *Metrarabdotos*. We then reanalyze a subset of the original data that can be meaningfully investigated using quantitative statistical approaches similar to those used in the original studies. We consistently fail to find variation in the evolutionary process during within-lineage evolution compared with cladogenetic events: the rates of evolution, the strength of selection, and the directions traveled in multivariate morphospace are not different when comparing evolution within lineages and at speciation events in *Metrarabdotos*, and genetic drift cannot be excluded as a sufficient explanation for the morphological differentiation within lineages and during speciation. Although widely considered the best example of a punctuated mode of evolution, morphological divergence and speciation are not linked in *Metrarabdotos*.

Keywords: speciation, anagenesis, cladogenesis, macroevolution, fossils, measurement theory.

Introduction

Eldredge and Gould's rereading of the fossil record (Eldredge and Gould 1972; Gould and Eldredge 1977), suggesting that stasis dominates lineage evolution and that phenotypic change mainly happens during rapid speciation events, initiated some of the fiercest debates in the history of evolutionary biology (e.g., Gould 1980, 2002; Charlesworth et al. 1982; Gingerich 1984, 2009). Although some of the most heated discussions regarding the punctuated equilibrium hypothesis have subsided, the debate regarding the relationship between speciation and morphological evolution remains to be settled (Lieberman and Eldredge 2014; Pennell et al. 2014*a*, 2014*b*; Venditti and Pagel 2014). The fossil record is our only direct source of information on how lineages evolve on timescales beyond a few centuries. How we analyze and interpret fossil data is therefore key in studying the pace of evolution during anagenesis and cladogenetic events. Several studies have investigated the tempo and mode of phyletic evolution in the fossil record (e.g., Malmgren and Kennett 1981; Bell et al. 1985, 2006; Chiba 1996; Theriot et al. 2006; Hunt 2007; Jones 2009; Hopkins and Lidgard 2012; Hunt et al. 2015; Voje 2016; Spanbauer et al. 2018; Voje et al. 2018), but the investigated fossil time series did not usually cover lineage-splitting events (but see Gingerich 1976; Kellogg 1983; Lazarus 1986; Pearson and Ezard 2014), making it challenging to address predictions from the punctuated equilibrium hypothesis directly. Empirical data on evolutionary dynamics during anagenetic and cladogenetic evolution among lineages belonging to a monophyletic group are exceedingly rare, but one important exception is the Neogene fossil record of the bryozoan *Metrarabdotos*, comprising "the most brilliant persuasive, and most meticulously documented, example ever presented for predominant (in this case, exclusive) punctuated equilibrium in a full lineage" (Gould 2002, p. 827).

Metrarabdotos Canu, 1914, is among the most speciose and abundant bryozoan genera in the Neogene fossil record of tropical America (fig. 1). Eight species of *Metrarabdotos* are known living at the present day worldwide, all with a tropical and subtropical distribution. Twenty-two *Metrarabdotos* species have been described from the

* Corresponding author; email: k.l.voje@ibv.uio.no.

ORCIDs: Voje, https://orcid.org/0000-0003-2556-3080; Di Martino, https://orcid.org/0000-0002-3892-4036; Porto, https://orcid.org/0000-0002-9210-8750.

Am. Nat. 2020. Vol. 195, pp. 899–917. DOI: 10.1086/707664

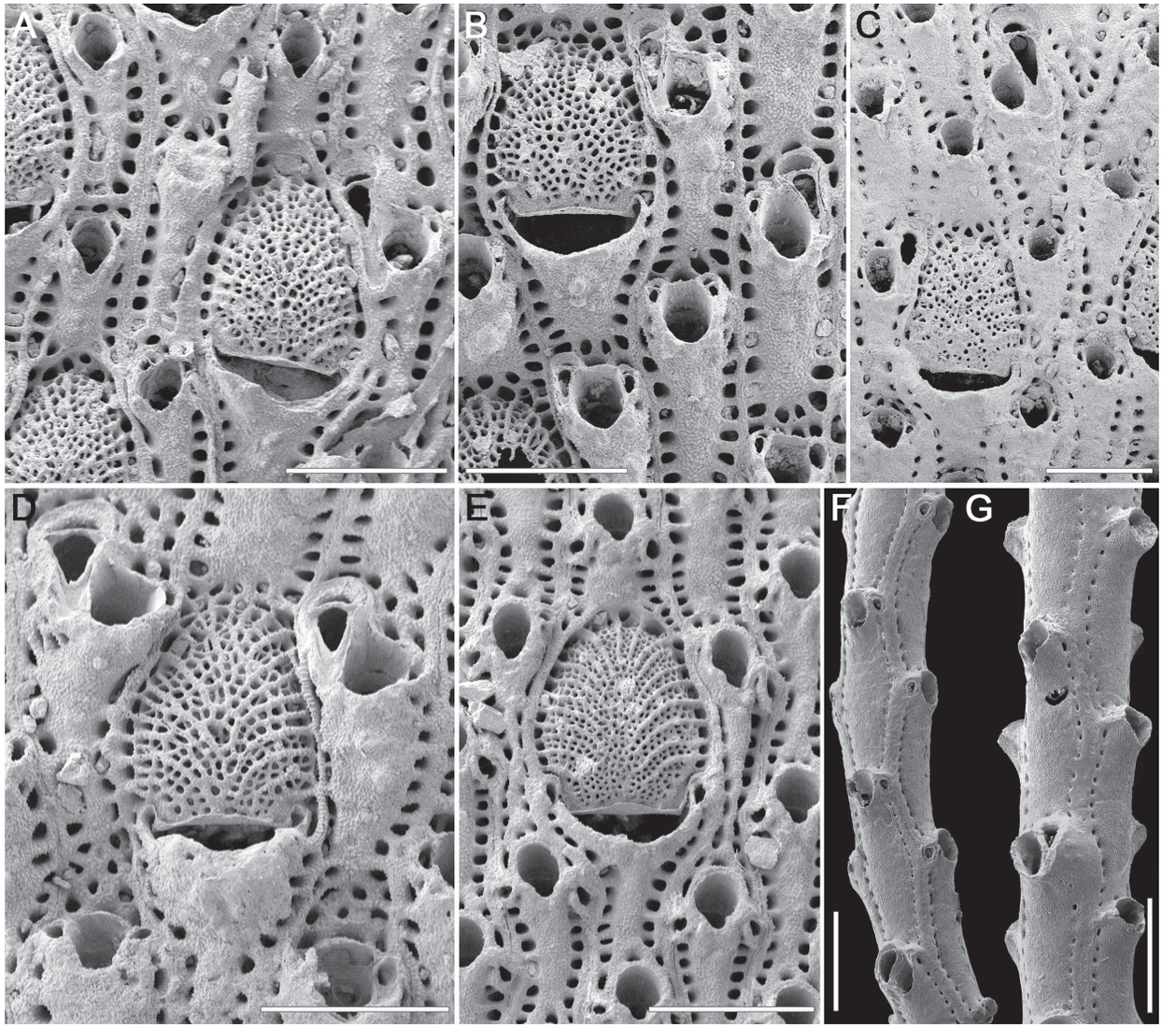

Figure 1: Scanning electron micrographs of the seven species of *Metrarabdotos* analyzed in this study. All specimens are from the upper Miocene (about 7 Ma) of the Dominican Republic. *A*, *Metrarabdotos colligatum* Canu & Bassler, 1919 (USNM 509422). *B*, *Metrarabdotos auriculatum* Canu & Bassler, 1923 (USNM 529754). *C*, *Metrarabdotos boldi* Cheetham et al., 2007 (USNM 529530). *D*, *Metrarabdotos coatesi* Cheetham et al., 2007 (USNM 529559). *E*, *Metrarabdotos lacrymosum* Canu & Bassler, 1919 (USNM 529816). *F*, *Metrarabdotos tainorum* Cheetham et al., 2007 (USNM 529855). *G*, *Metrarabdotos jungi* Cheetham et al., 2007 (USNM 529870). Scale bars: 500 μm.

fossil record from tropical America alone, making the genus one of the most diverse genera in Neogene deposits from this region (Cheetham et al. 2007). Twelve of the extinct species are found in the stratigraphically well-constrained Miocene-Pliocene deposits of the Dominican Republic, where the average time resolution of the fossil record of the most abundant species is 0.2 million years (Cheetham et al. 2007).

A series of seminal articles by Cheetham and colleagues consolidated *Metrarabdotos* as a textbook example of a punctuated mode of evolution during the 1980s and 1990s (Cheetham 1986, 1987; Cheetham et al. 1994; Jackson and Cheetham 1994). A phylogenetic hypothesis of the genus was created on the basis of large-scale analyses of morphology and stratigraphic positions of fossils (Cheetham 1986; Cheetham and Hayek 1988; Jackson and Cheetham 1994). Using this phylogeny as a backbone, several articles with detailed analyses of morphological evolution within lineages and at speciation events found mostly strong support for minimal within-species evolution and large-scale changes during speciation (Cheetham 1986; Cheetham et al. 1993, 1994; Jackson and Cheetham 1994). To this day,

Metrarabdotos is generally considered the best example of a punctuated mode of evolution (Jackson and Cheetham 1999; Benton and Pearson 2001; Gould 2002; Hunt and Slater 2016; Gingerich 2019).

Cheetham and colleagues' groundbreaking work on *Metrarabdotos* is a powerful example of how paleontology and evolutionary biology can be merged to generate insights not possible within each discipline alone. However, the studies finding evidence in favor of a punctuated mode of evolution in *Metrarabdotos* are not without issues that potentially affect the original conclusions. Some authors have pointed to the general challenge regarding the biologic validity of morphospecies in the fossil record (e.g., Levinton 1988; but see Jackson and Cheetham 1990). Others have pointed to issues in sampling quality that might affect the hypothesized phylogenetic relationships within *Metrarabdotos* (Benton and Pearson 2001). Our concern here is issues related to measurement theory. While statistical models can aid in making inferences from measurements, measurement theory provides guiding principles to avoid disconnecting measurements from the reality we are interested in studying (Houle et al. 2011). The combination of statistics and measurement theory is accordingly helpful for making informed inferences about the biological question we intend to investigate.

We start by pointing to some methodological and measurement theoretical issues in the original work on *Metrarabdotos* and discuss how these issues can have affected the conclusions of a strongly punctuated mode of evolution within the genus. We then present a reanalysis of the original *Metrarabdotos* data using quantitative statistical approaches similar to those used in the original studies to reassess the evidence for a punctuated mode of evolution. We restrict the analyses to traits where the scale type allows for a meaningful interpretation of differences among population trait means, since this is an important assumption underlying the analyses in much of the original work on *Metrarabdotos*. Following Houle et al. (2011), the scale type defines what properties of a set of measurements that can be used to draw empirically meaningful conclusions. Furthermore, only traits that have actually been measured (in contrast to traits being assigned a mean value on the basis of which species the specimen belongs to; see below) are included in our reanalysis. We first compare rates of phenotypic evolution during anagenesis to rates at speciation events. Second, we conduct discriminant function analyses to investigate multivariate evolution within lineages. We then capitalize on the clonal nature of *Metrarabdotos* and estimate (broad-sense) genetic variance-covariance (**G**) matrices for lineages with a reasonable sample size of measured colonies. Using these **G** matrices, we investigate to what extent the observed morphological differentiation within lineages and at speciation events can be explained by genetic drift. We also test whether the estimated selection gradients required to produce the observed morphological changes within lineages and at speciation events are similar or different. We proceed by investigating whether the evolvability in directions traveled during anagenetic evolution was different compared with the evolvability in directions traveled during speciation. We end by discussing some potential shortcomings of the *Metrarabdotos* data in evaluating predictions from the punctuated equilibrium hypothesis.

Methodological and Measurement Theoretical Issues in the Original Studies of Tempo and Mode in *Metrarabdotos*

Respecting Scale Type

Skepticism is warranted if the outcome of an analysis depends on the numbers or coding we choose to represent the traits we are interested in studying (i.e., how the traits were measured). A key principle within measurement theory is that the question we want to investigate (the theoretical context) should guide our choice of measurements (the scale types) and that the scale of these measurements put constraints on how they can be sensibly analyzed (Houle et al. 2011). One type of measurement might be favored given one theoretical context, while a different type of measurement might be better suited given a different theoretical context. For example, a taxonomist needs to be able to differentiate a range of characters that describe the positions or orientation of morphological structures in a species (left/center/right position, proximal/transversely/distal orientation of a structure, etc.). Such nominal traits are often given arbitrary values to differentiate between character states (e.g., 0, 1, and 2 represent, respectively, the left, center, and right positions of a morphological structure), and individuals are assigned the same number if they possess the same trait/character state. Order is meaningless on a nominal scale (e.g., the claim that left is less than right does not make sense), in contrast to an ordinal scale, where there is an order relation between the numbers (the categories small, medium, and large can be given the numbers 1, 2, and 3, respectively, where the order relation has meaning in the sense that $1 < 2 < 3$).

While nominal and ordinal traits can be meaningful, differences in magnitude between the numbers we choose to represent the character states in nominal and ordinal traits are not, since such differences assume that values reflect magnitude (Houle et al. 2011). The conclusions of a statistical analysis that calculates means and variances on the basis of nominal and ordinal traits are accordingly not meaningful, since the conclusion may depend on the arbitrary numbers we chose to represent the different character states.

The original *Metrarabdotos* data set analyzed in Cheetham (1986) consisted of 46 traits, 25 of which are categorical traits scored using arbitrary numbers (see table 1 in Cheetham et al. 2007). Intuitively, many of the categorical traits could be interpreted as being on an ordinal scale. For example, many of these traits describe the orientation of a structure as "directing inward" (scored as 1), "parallel to autozooidal axis" (scored as 2), or "directing outward" (scored as 3), which seems to indicate ordered characters states. However, if the categorical traits in the *Metrarabdotos* data set were truly ordered, this would predict that most within-species polymorphisms should be type 1,2 or 2,3 but rarely (if ever) 1,3. Yet character states 1 and 3 are actually the two most common states and are frequently polymorphic within populations in the absence of 2. In other words, it does not seem like a population needs to go through state 2 to get from state 1 to 3 for most of the discrete traits. Without knowledge about the developmental and genetic background of each discrete trait, we will therefore refer to these traits as nominal in the rest of this study. Note, however, that none of the conclusions we reach in this study depend on this distinction, as magnitudes between trait states within both nominal and ordinal traits are not meaningful.

Cheetham (1986) analyzed all 46 traits using a series of linear discriminant analyses in order to delimit species and to estimate morphological distances within and between lineages of *Metrarabdotos* (in units of discriminant scores). The results from these analyses were summarized in a now-canonical figure that portrays a rate of evolution within lineages that is zero or very close to zero, while morphological evolution at speciation events is mostly large and rapid. Cheetham et al. (2007) reanalyzed the original data from the article in 1986 together with new colonies of *Metrarabdotos* and concluded that the previously identified pattern of a punctuated mode of evolution within the genus still holds. However, the results of the linear discriminant analyses conducted in Cheetham (1986) and Cheetham et al. (2007) are difficult to interpret and rely on, since this statistical analysis assumes that magnitudes between numbers have meaning, an assumption that does not hold for about half of the traits analyzed. A large portion of skepticism is needed when interpreting trait dynamics and morphological distances based on analyses of nominal traits.

Furthermore, Cheetham et al. (1993, 1994, 1995) pioneered the application of quantitative genetics on data from the fossil record, capitalizing on the fact that colonies of bryozoans consist of multiple genetically identical units (zooids). To investigate whether the morphological differentiation during speciation within *Metrarabdotos* could be accounted for by random processes (genetic drift) or whether explanations involving directional selection were needed, Cheetham et al. (1994) performed sophisticated quantitative genetic analyses, which involved estimating broad-sense **G** matrices from the fossil record. The conclusion was that agents of evolution acting during speciation were different compared with the strong stabilizing selection needed to explain the extreme stasis within lineages, a result "consistent with the 'stronger' version of punctuated equilibria theory decoupling speciation from forces acting within species" (Cheetham et al. 1994, p. 373). However, seven out of 15 traits used to estimate variance-covariance matrices were on a nominal scale (Cheetham et al. 1994). The quantitative genetic analysis of the *Metrarabdotos* data therefore needs to be revisited to ensure the original conclusion remains valid.

Treatment of Missing Data

Another reason for reassessing the conclusion of a strongly punctuated pattern of evolution in *Metrarabdotos* is the way missing data were dealt with. Cheetham (1986) conduct an initial series of discriminant function analyses on a reduced set of colonies from the *Metrarabdotos* data set to define distinct groups/species (Cheetham's [1986] description of how he conducted these analyses does not contain enough detail to pinpoint exactly how they were performed, but Cheetham et al. [2007] contains a much more thorough description of a very similar analysis and how it deviates from the analyses performed in Cheetham [1986]). These initial discriminant analyses were conducted on 15 traits (seven on a nominal scale) from 166 colonies, where these 15 traits had been measured for one to five zooids per colony. A second set of discriminant analyses were then performed on the whole *Metrarabdotos* data set (all 46 traits on 240 colonies) to place the remaining colonies in a particular group/species. The colonies that were part of this second set of discriminant analyses had a lot of missing data for many of the 46 traits, since many of the 46 traits could be measured only on specific types of zooids that were often not present in the investigated colonies. When a colony had been placed in a group/species during the second round of discriminant analyses, missing trait data were replaced with mean trait values from the species it now belonged to.

The use of a discriminant function analysis to separate species based on a much smaller set of colonies that are then used to assign species status to other colonies can create a strong bias toward morphological homogeneity within lineages through time (i.e., low rate of evolution). We detail in the supplemental PDF (available online) how assigning species status based on proximity in this undersampled multivariate space might enforce a level of homogeneity within lineages that is not reflected in the "true" trait distribution in multivariate space (for an extended

discussion in the context of archeology, see also Kovarovic et al. 2011). Importantly, the amount of missing data in the *Metrarabdotos* data set is large (table 1). Cheetham et al. (2007) added 86 new colonies and 10 new traits (ratios of continuous traits already in the original data set) to the *Metrarabdotos* data set. Based on this extended data set and excluding the 15 traits with no or very few missing data and traits on a nominal scale, 19 traits remain. These 19 traits have on average not been measured in 42% (median, 50%) and 35% (median, 42%) of the colonies in the lineages *M. lacrymosum* and *M. colligatum*, while the situation is slightly better for the other species (table 1). However, the situation is worse across time: for these 19 traits, six of the seven lineages that constitute ancestor-descendant pairs according to Cheetham et al. (1994) are missing measurements from at least 40% of the time intervals they are present (median), and two lineages (*M. tainorum* and *M. jungi*) are on average missing measurements from 79% and 86% (median, 79% and 86%, respectively) of the time intervals they are present. The effect of replacing all of these missing data with species means bias trait dynamics within lineages toward strict stasis. For example, the species *M. lacrymosum* occurs in 16 time intervals covering a time span of 4.35 million years, but 12 of the 19 traits have been measured in a maximum of nine time intervals, which means the time intervals with missing data will have trait values identical to the species mean for these traits. This situation is not unique for *M. lacrymosum* and applies to all of the lineages to various extents. In *M. tainorum*, 11 of the 19 traits have been measured in only one out of eight time intervals, causing these traits to appear invariant during the 3.85 million years this species is present.

To summarize, given the status of *Metrarabdotos* as "the best documented, indeed already canonical, example of punctuated equilibrium as an invariant pattern for an entire clade across its full geographic range" (Gould 2002, p. 844), a reanalysis of these data seems justified in light of the methodological and measurement theoretical issues in some of the original work on the genus. Of particular interest is the extent to which the conclusion of a punctuated mode of evolution within the genus (Cheetham 1986, 1987; Cheetham et al. 1993, 1994, 2007; Jackson and Cheetham 1994) remains valid.

Material and Methods

Trait Data

All lineages analyzed in this study are part of ancestor-descendant pairs as defined in the original work by Cheetham (Cheetham 1986; Cheetham et al. 1994). Table 2 lists all of the analyzed species and the hypothesized ancestor-descendant relationships. The data analyzed in this study were first published by Cheetham (1986) and analyzed in subsequent articles (Cheetham 1987; Cheetham et al. 1993, 1994, 2007; Jackson and Cheetham 1994). Cheetham et al. (2007) extended the original data set with 86 colonies

Table 1: Overview of missing data in the *Metrarabdotos* data set

		Mean (median) % of colonies with missing data		
Species	No. colonies	Reanalyzed traits	Excluded traits	All traits
M. tainorum	14	.1 (.0)	16.4 (16.7)	10.1 (16.7)
M. jungi	11	2.6 (.0)	12.2 (11.5)	8.4 (11.5)
M. lacrymosum	55	.9 (.0)	41.9 (50.0)	26.1 (18.0)
M. colligatum	95	.2 (.0)	35.2 (33.3)	21.6 (6.4)
M. auriculatum	78	.2 (.0)	37.1 (42.3)	22.8 (7.7)
M. boldi	29	.1 (.0)	19.1 (23.1)	11.8 (3.9)
M. coatesi	44	.1 (.0)	19.1 (18.0)	11.7 (10.3)
		Mean (median) % of time intervals with missing data		
	No. time intervals	Reanalyzed traits	Excluded traits	All traits
M. tainorum	8	1.0 (.0)	86.2 (87.5)	53.2 (87.5)
M. jungi	7	17.9 (.0)	79.0 (71.4)	55.3 (71.4)
M. lacrymosum	16	.5 (.0)	37.2 (43.8)	23.0 (18.8)
M. colligatum	16	.0 (.0)	36.5 (43.8)	22.4 (.0)
M. auriculatum	20	.0 (.0)	35.0 (40.0)	21.5 (5.0)
M. boldi	15	.6 (.0)	43.2 (53.3)	26.7 (6.7)
M. coatesi	14	.0 (.0)	22.9 (21.4)	14.1 (.0)

Note: The table shows the mean and median percentage of missing data on the colony level and the stratigraphic level within each lineage. Only traits on a ratio scale are part of the statistics. Reanalyzed traits refer to the 12 traits analyzed in this study (table 3), while the excluded traits are the 19 traits not analyzed in this study because of the extent of missing data.

Table 2: Differences in trait means between ancestor-descendant *Metrarabdotos* species pairs

	Ancestor → descendant				
Traits	*M. colligatum* → *M. auriculatum*	*M. auriculatum* → *M. boldi*	*M. boldi* → *M. coatesi*	*M. lacrymosum* → *M. tainorum*	*M. tainorum* → *M. jungi*
LZ	.013*	.021*	−.010	−.161***	−.003
WZ	−.031***	.051***	−.033***	−.119***	−.004
LO	−.003	.016**	−.019***	−.026*	.037
WO	.016***	.013*	.001	−.032***	.003
LD	.158***	.046**	−.046**	−.028	−.062*
LAVS	−.066***	.116***	−.035	.036	−.023
LAVL	−.105***	.164***	−.060***	.148***	−.032
NUMA	−.004	.039***	−.025**	−.129***	−.010

Note: Differences are illustrated on a proportional scale (i.e., standardized by the ancestor's trait means, calculated on the basis of all populations across all time intervals where the ancestor is present). Trait abbreviations are defined in table 3. Asterisks indicate statistically significant differences (*t*-test).

* $P < .05$.

** $P < .01$.

*** $P < .001$.

and 10 new traits, and it is this extended data set we analyze here. The unnamed species from the original work have been named according to Cheetham et al. (2007): n.sp.3 = *M. tainorum*, n.sp.4 = *M. jungi*, n.sp.9 = *M. boldi*, and n.sp.10 = *M. coatesi*.

As an initial step, we filtered the *Metrarabdotos* trait database and only selected traits that fulfilled the following criteria: traits selected (i) were on a ratio scale (comparisons of differences among trait means are meaningful on this scale, as is mean scaling; Houle et al. 2011) and (ii) had been measured (and not set to be equal to the species trait mean) on at least four zooids per colony. The four-zooid minimum is aimed at preventing an exaggerated effect of sampling error on mean estimates. Analyzing traits on a ratio scale allows us to reanalyze data using quantitative statistical approaches similar to those used in the original studies of *Metrarabdotos*. Only colonies with stratigraphic age information were used. We removed any trait that is defined as a ratio of two other traits already present in our data set, as those are a source of multicollinearity and lead to the estimation of rank-deficient matrices. Of the 46 traits in the original *Metrarabdotos* data set, 25 (54.4%) were filtered out for not being on a ratio scale. Of the remaining traits, 13 (28.3%) were removed because of a lack of enough samples ($N < 4$ per colony). A total of eight traits (17.3%) were kept and used in all further analyses (fig. 2; table 3; see the previous section for more details on why we had to exclude the majority of the traits; the excluded traits are listed in table S1 [tables S1, S2 are available online]). All eight traits are at least approximately normally distributed when mean scaled (fig. S2; figs. S1–S5 are available online). However, to include as many traits as possible we also redid the evolutionary rate analyses (see below) where we included four ratios of continuous traits (published in Cheetham et al. 2007) that fulfilled the two above-mentioned criteria. The results of these analyses are qualitatively very similar to the analyses of the eight traits and are reported in the supplemental PDF (figs. S3, S4), but they are not discussed further in the main text.

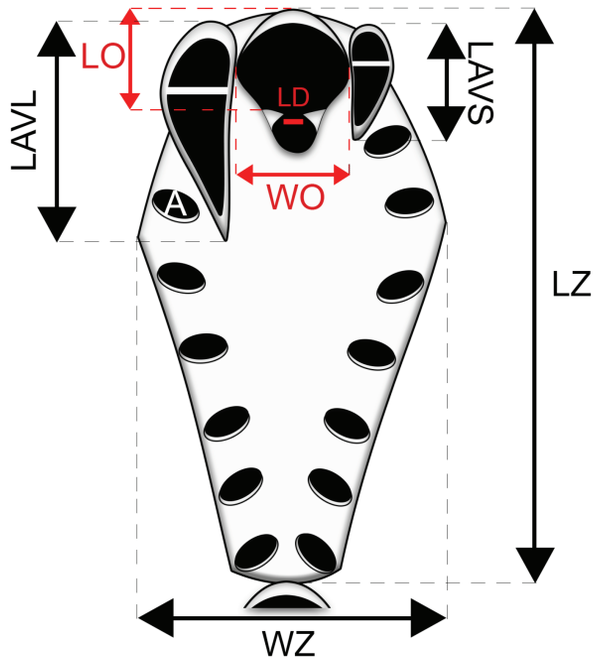

Figure 2: Schematic representation of an ordinary autozooid of *Metrarabdotos*. The traits analyzed in the current study are indicated. Trait abbreviations are defined and described in table 3. All traits are length measurements except NUMA, which refers to the count of areolae, indicated with an A in the sketch. Note that it is unclear whether LD was measured from the base or the tips of the denticles. In this sketch, LD is represented as the distance between the denticle tips.

Table 3: Morphological characters analyzed in the *Metrarabdotos* species

Abbreviation	Description
LZ	Zooid length: distalmost point on completed peristome to same point on next proximal zooid
WZ	Zooid width: maximum distance between lateral margins of frontal shield perpendicular to length
LO	Length of secondary orifice, measured at the level of proximal denticles
WO	Width of secondary orifice, measured perpendicular to orifice length
LD	Distance between lateral oral denticles
LAVS	Length of shorter avicularium
LAVL	Length of longer avicularium
NUMA	Number of areolae, counted on the upper surface of the frontal shield
Traits included in a separate analysis, extending the data set to 12 traits:	
LZ/WZ	Zooid length-width ratio
LO/WO	Orifice length-width ratio
LD/WO	Distance between lateral denticles in proportion to orifice width
LAVS/LAVL	Ratio of lengths of shorter and longer avicularia of a pair

Rates of Morphological Evolution during Anagenesis and Cladogenesis

The core of the punctuated equilibrium hypothesis is the prediction of faster rates of evolution during speciation events compared with within-lineage (anagenetic) evolution. Rates of evolution during anagenetic and cladogenetic events were estimated the following ways: rates of anagenetic evolution were calculated by (i) taking the difference between vectors of trait means of a descendant and ancestral population within a lineage (population in the sense of a time-averaged assemblage of fossil samples), (ii) standardizing this multivariate distance vector by the trait means in the ancestral population to transform the evolutionary changes to percent change in trait means, (iii) calculating the vector norm, and (iv) dividing this value by the time interval separating the ancestor and descendant populations in order to obtain the rates of trait evolution (percent change in means per million years). Rates of evolution during cladogenetic events were estimated similarly, but here the ancestor and descendant populations belong to different lineages. The descendant population is always defined as the first fossil sample in a new lineage, but the time of lineage splitting is unknown. Cheetham et al. (1994) assumed the timing of speciation happened at the time point in the ancestor species that was closest in time to the first appearance of the descendant species, while Cheetham (1986) assumed that sister species started to diverge at the time of the first appearance of the ancestral lineage. To cover both of these assumptions, we therefore estimated rates of evolution between the first population (oldest fossil sample) in the descendant lineage and all populations in the ancestral lineage that are older than the descendant population (see fig. 3 for an illustration). We evaluated the statistical significance of observed differences in rates between anagenetic and cladogenetic evolution using random permutation tests implemented in the lmPerm (ver. 2.1.0) R package (Wheeler 2016).

Following Cheetham (1986), we also performed a linear discriminant function analysis on the eight traits to study the multivariate distribution of cladogenetic and anagenetic changes within *Metrarabdotos*. The discriminant analysis was conducted using maximum likelihood estimation as implemented in the lda function of the MASS R package (Venables and Ripley 2002), where prior probabilities of class memberships were set to equal class proportions in the original data and individuals with missing values were omitted.

Estimating Lineage-Specific $\boldsymbol{G}$ *Matrices*

Individual zooids within a bryozoan colony are clones, which means that within-colony trait variation is due to environmental effects while among colony variation is due to both genetic and environmental effects. Treating within-colony zooids as replicate measurements of the same individual (colony), the within- and among-colony components of phenotypic variance were obtained by fitting mixed models using the R package MCMCglmm (Hadfield 2010). In these mixed models, we controlled for colony age effects by adding the stratigraphic layer as a random effect in the mixed model. Although the clonal biology of bryozoans represents a unique opportunity to estimate quantitative genetic parameters from the fossil record, we would like to emphasize some potential caveats. We lack the necessary data to control for the effects of dominance and epistasis on the estimated variances and

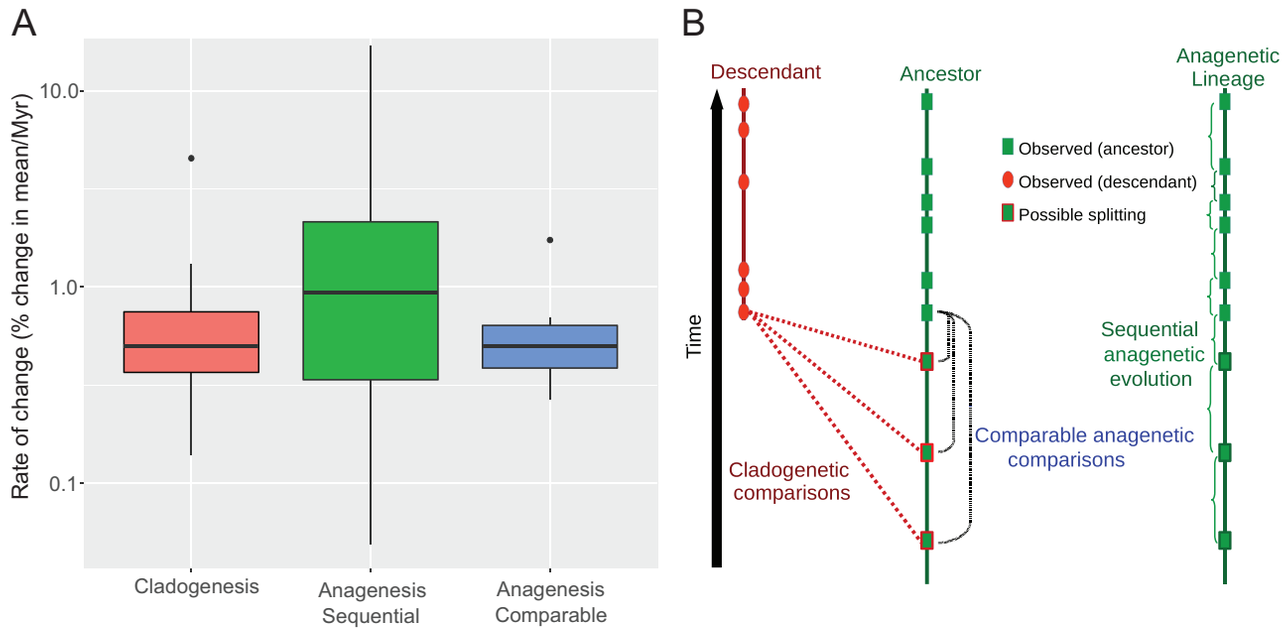

Figure 3: Rates of evolution within lineages (anagenetic evolution) and during speciation events (cladogenesis). *A* shows rates of evolution (percent change in trait mean per million years) during cladogenesis and anagenesis within *Metrarabdotos*. We did not detect any systematic difference in rates of evolution during anagenetic and cladogenetic evolution (table 4). Because of unknown time of splitting events between lineages, cladogenetic rates are estimated on the basis of evolutionary changes between the first population (oldest fossil sample) in the descendant lineage and all populations in the ancestral lineage that are older than the descendant population (*B*). Sequential anagenetic rates of evolution are computed between consecutive ancestor-descendant population pairs within a lineage. Comparable anagenetic rates are rates of anagenetic evolution between the population in the ancestral lineage that is older but closest in time to the first appearance of the descendant population and all older populations in the ancestral lineage. These rates are estimated across time intervals that are similar (often identical) to those used for the rates calculated during cladogenetic events and are therefore more directly comparable to the cladogenetic rates.

for colony-specific environmental effects, which may be substantial. Furthermore, the *Metrarabdotos* data are stratigraphically well constrained, but colonies from the same stratigraphic level are not from a single generation. The effect of time averaging across fossil samples means that anagenetic microevolution may contribute to some of the observed within-lineage population variation at a given time point in the fossil record (Hunt 2004*a*, 2004*b*). The estimated lineage-specific variance-covariance matrices are therefore most appropriately referred to as broad-sense **G** matrices that are potentially upwardly biased because of anagenetic evolution and plasticity.

We estimated the (broad-sense) genetic variance-covariance matrix for five lineages. These five species are ancestors to descendant lineages, and all have a colony number of at least 29 (94 = *M. colligatum*, 68 = *M. auriculatum*, 48 = *M. lacrymosum*, 29 = *M. boldi*, 44 = *M. coatesi*). The choice of minimum sample size is based on a large-scale analysis of sampling error in high-dimensional covariance matrices described in Grabowski and Porto (2017). At this sample size, 8 × 8 structured covariance matrices are expected to be well estimated, with the most common matrix descriptors predicted to be around 2%–7% away from their true value (for the prediction algorithm, see Grabowski and Porto 2017). The priors for the Bayesian mixed models MCMCglmm were a crude guess based on the phenotypic variance matrix of the traits for each species. Sensitivity of parameter estimates based on the priors were therefore thoroughly tested by varying the priors. All models were robust against changes in the priors. Each model ran for 1,500,000 iterations, with a thinning interval of 1,000. We discarded 500,000 iterations as burn-in. We assessed pairwise convergence of multiple independent chains (with diverse priors) using Gelman and Rubin's (1992) convergence diagnostic (psrf) implemented in the coda (ver. 0.19-2) R package (Plummer et al. 2006). This convergence diagnostic compares the estimated between-chain and within-chain variances for each model parameter, and whenever large differences between these variances are present it is used to indicate nonconvergence. For all matrices, multivariate psrf factors between 1 and 1.2 were considered as indicating good convergence. Matrices that did not converge were removed from further analyses.

Since some estimated **G** matrices were negative semidefinite, we controlled for inverse matrix noise (Marroig et al. 2012) before calculating any statistic that depends

on matrix inversion. Noise control techniques are used here to prevent matrix noise (estimation error) from dominating the inverted **G** matrices used in, for example, the calculation of conditional evolvabilities (see below). We used the extension approach proposed by Marroig et al. (2012), in which the smallest eigenvalues of **G** (i.e., the largest contribution to the inverted matrix) are substituted by the last reliable eigenvalue (for details, see Marroig et al. 2012), in an attempt to prevent noise from dominating any evolutionary statistic in the downstream analyses.

Once properly estimated, we compared estimated **G** matrices using random skewers (Marroig and Cheverud 2001). In brief, we applied 10,000 random selection vectors of unit length to both matrices and used the average vector correlation between the evolutionary responses produced by such matrices as a measurement of similarity. Observed similarity values were then compared with a null distribution of values that would be expected given identical matrices that are undersampled (Grabowski and Porto 2017).

Drift and Selection

The evolutionary mechanisms causing or constraining evolutionary changes during anagenetic and cladogenetic evolution within *Metrarabdotos* are not known. Cheetham et al. (1994) suggested drift as an adequate explanation for the observed morphological differentiation during speciation events but noted that directional selection remained a plausible alternative for at least some cladogenetic events. Here, we test whether drift can account for the observed morphological evolution within lineages and at speciation events. Since directional selection cannot be rejected as a potential evolutionary mechanism during speciation, we also investigate whether the selection needed to explain changes at cladogenetic events are stronger (i.e., require steeper selection gradients) compared with the selection needed to explain changes between consecutive ancestor and descendant populations within lineages (i.e., anagenetic evolution). Again, a population is in this context referring to a time-averaged assemblage of fossil samples.

Under purely stochastic evolution (i.e., drift), the between-population trait variances should be proportional to the within-population trait variances of the ancestral lineage (following Lande 1979). In logarithmic form, this relationship can be expressed as $\log B_i = \log(t/N_e) + \beta(\log W_i)$, where B_i is the between-population variance and W_i is the within-population variance for the ith eigenvector, t is the time in generations, N_e is the effective population size, and β is the regression slope. The terms B and W are considered strictly proportional when β is equal to 1. A neutrality test can be implemented, therefore, by regressing B on W and testing whether the 95% confidence interval of the regression slope includes the value of 1 (i.e., cannot discard drift) or not (i.e., evidence of selection; Ackermann and Cheverud 2002). Here, we performed neutrality tests for all five **G** matrices and their corresponding lineages to investigate whether changes within lineages are compatible with a purely stochastic evolutionary process. We also test whether drift can explain changes during cladogenesis by performing the neutrality test on the average **G** matrix (based on the five species-specific **G** matrices) and the vectors of trait means in each of the five descendant populations in each ancestor-descendant pair.

Retrospective net selection gradients for anagenetic and cladogenetic events were calculated using the five estimated **G** matrices. Morphological differences between consecutive ancestral-descendant populations (populations in the sense of time-averaged assemblages of fossil samples) within lineages were calculated by (i) taking the difference between vectors of trait means of the two populations before (ii) standardizing this multivariate distance vector by the trait means in the ancestral population, which transforms the evolutionary changes to percent change in trait means. Ancestral populations in each cladogenetic event were chosen in the same way as when we estimated rates of evolution (see above and fig. 3). In all cases, we used the estimated lineage-specific **G** matrix and Lande's (1979) equation of multivariate evolution ($\beta = \mathbf{G}^{-1}[z_i - z_j]$) to estimate mean-standardized directional selection gradients (β_μ) between an ancestral (z_i) and a descendant (z_j) population. The fact that our reconstruction of **G** is based on fossils, not extant species, allows the use of (potentially upwardly biased) broad-sense genetic matrices (and not phenotypic matrices) for the ancestral lineage, a rare possibility in retrospective selection analyses. We evaluated the statistical significance of observed differences in the strength of directional selection between anagenetic and cladogenetic events using random permutation tests implemented in the lmPerm (ver. 2.1.0) R package (Wheeler 2016). Permutation tests are a nonparametric type of statistical significance test in which the distribution of the test statistic under the null hypothesis is approximated by randomly rearranging the labels of the observed data points. In our case, the labels refer to the type of event (cladogenetic vs. anagenetic), and the observed data points refer to the estimates of directional selection.

Evolvability and Directions of Divergence

We calculate different evolvability measures for the estimated **G** matrices following the approaches in Hansen and Houle (2008) using the evolqg 0.2-5 R package (Melo et al. 2016). Each **G** matrix was mean standardized before we calculated evolvability parameters. Evolvability (e)

represents the ability of a population to evolve in the direction of selection in the absence of stabilizing selection. The evolvability (e) of a trait can be measured as the expected proportional change in the trait per generation to linear directional selection of unit strength (Hansen et al. 2003; Hansen and Houle 2008). A trait with an evolvability of 0.1 will accordingly have a predicted 10% change in trait mean per generation per unit directional selection (i.e., the selection gradient is 1). The average evolvability (e_{mean}) corresponds to the expected evolvability in a random direction in phenotype space and was calculated as the average of the eigenvalues of the **G** matrix. Conditional evolvability (c) represents the ability of a population to respond to directional selection when all other traits in the **G** matrix are under stabilizing selection (i.e., all other traits are not allowed to change). The average conditional evolvability (c_{mean}) was calculated on the basis of the conditional evolvability in 10,000 random directions (selection gradients) in phenotype space. Each element in each of the 10,000 vectors was randomly sampled from a multivariate normal distribution with zero mean and unit variance, divided by its norm to standardize it to unit length.

We estimated the evolvability and conditional evolvability in the directions in which evolution took place during anagenesis and cladogenesis to assess whether the structure of the ancestral **G** matrix influenced morphological divergence during speciation and to investigate whether evolution happened in directions with different levels of evolvability during anagenetic and cladogenetic evolution. Evolvability and conditional evolvability during cladogenesis were calculated using the same set of potential ancestral populations as when we estimated rates of evolution during speciation (see fig. 3). Each element in the vector describing the direction of divergence was calculated by subtracting the average trait values of the first population in the descendant lineage from the average trait values of the ancestor population, standardized by the ancestral mean. The vector representing divergence was normalized to unit length. Similarly, evolvability and conditional evolvability statistics during anagenetic evolution were calculated between consecutive sample means within all lineages with an estimated **G** matrix.

We evaluated the statistical significance of observed differences in the distribution of evolvabilities and conditional evolvabilities during anagenetic and cladogenetic evolution using random permutation tests implemented in the lmPerm (ver. 2.1.0) R package (Wheeler 2016).

Results

Species Differences

The eight traits we included in our analyses are correlated with species differences (table 2), and more than 65% of the traits differ significantly across the five ancestor-descendant species pairs we analyzed when standardized by the ancestral mean (mean calculated on the basis of all populations across all time intervals where the ancestor is present, both before and after the hypothesized descendant linage is present in the fossil record) for each ancestor-descendant pair. The number of significant trait differences ranged from eight (between *M. auriculatum* and *M. boldi*) to one (between *M. tainorum* and *M. jungi*). All traits differed across at least one ancestor-descendant species pair, while the length of the shorter avicularium (LAVL) differed significantly across four of five ancestor-descendant species pairs.

Trait Dynamics and Rates of Evolution during Anagenesis and Cladogenesis

The multivariate distribution of cladogenetic and anagenetic changes is illustrated through a 3D phylomorphospace based on the average discriminant score for each lineage in each time interval where it is present (fig. 4*A*; see the supplemental material, available online, for a rotating 3D version of this figure). Discriminant functions 1 and 2 explain 67.3% and 19.4% of the variation, respectively. Lineages show large fluctuations in multivariate trait space and tend to occupy different parts of the morphospace at different time points. All lineages occupy overlapping ranges in this multivariate morphospace except *M. tainorum* and *M. jungi*, which are located in a separate part of the multivariate morphospace. The original presentation of multivariate evolution within *Metrarabdotos* (from Cheetham et al. 1994) is shown in figure 4*B* for comparison.

Second, we compared the rates of evolution during cladogenetic and anagenetic events using mean-standardized measurements of trait evolution following Hansen and Houle (2008). We find no evidence for differences in rates of cladogenetic versus anagenetic change (fig. 3*A*; table 4). The range of observed rates is larger for the full anagenetic evolution data set, but this may be due to the fact that the number of rates for possible cladogenetic events ($N = 8$) is smaller than the number of rates for observed anagenetic events ($N = 99$). To more directly compare rates of anagenetic and cladogenetic evolution, we therefore calculated rates of anagenetic evolution between the population in the ancestral lineage that is older but closest in time to the first appearance of the descendant population and all older populations in the ancestral lineage (see the diagram in fig. 3*B* for details). These anagenetic rates are therefore calculated on the basis of time intervals that are very similar (most often identical) to the ones used for the rates calculated during cladogenetic events, but these are also indistinguishable from the cladogenetic rates (fig. 3*A*; table 4). In the supplemental PDF (table S2), we provide

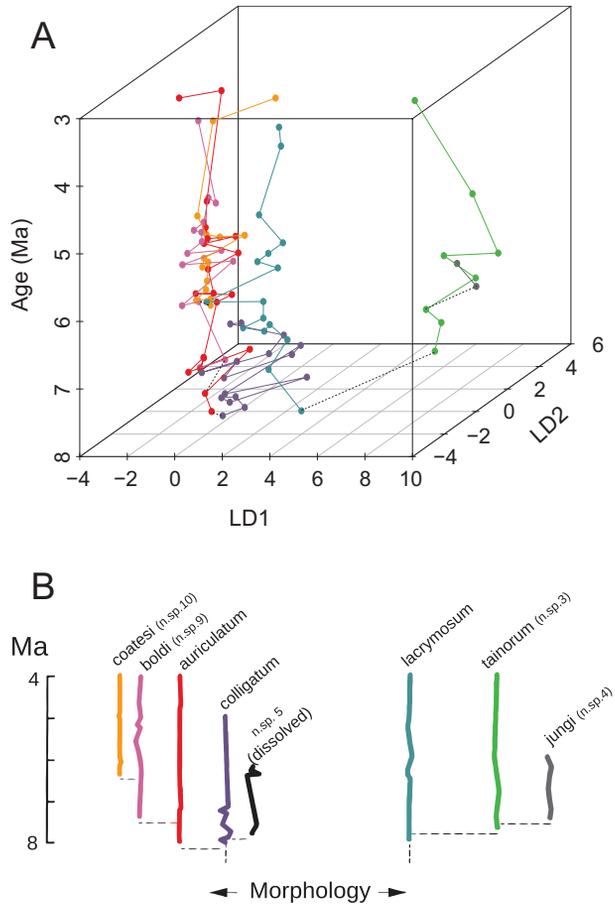

Figure 4: Comparison of multivariate phylomorphospaces of *Metrarabdotos* lineages. *A*, The phylomorphospace calculated on the basis of a discriminant function analysis of the curated set of eight traits analyzed in the current study shows how the lineages evolve through time. Points represent the average scores for each species at each sampling time point. Black dashed lines represent cladogenetic events. The first discriminant function describes 67.3% of the variation, while the second function describes 19.4%. A rotating 3D version of this figure is available in the supplemental material. *B*, Original illustration of the phylomorphospace of *Metrarabdotos* (Cheetham 1986; Cheetham et al. 1994). Colors are comparable across the two plots, allowing a direct comparison of the illustrated dynamics for each lineage. The morphology axis represents a composite variable based on canonical axes of variation from a discriminant function analysis. Differences between species that are not part of a particular ancestor-descendant pair are not comparable, as they are on an arbitrary scale.

the calculated rates of evolution between every possible ancestor-descendant sample in each of the five cladogenetic events, along with rates during "comparable" anagenetic evolution. There is no systematic trend in the rates of evolution regarding which sample that is treated as ancestral in a cladogenetic event, but note that the number of possible ancestral samples for each cladogenetic event is small and never larger than three.

Genetic Drift, Selection Gradients, and Patterns of Evolvability

The estimated (broad-sense) **G** matrices for five species of *Metrarabdotos* indicate that genetic variation accounts, on average, for 50.2% of the total variation in the zooid morphology. The five species seem to have above-average trait evolvability, as the average trait evolvability and conditional evolvability were 1.18% and 0.49%, respectively, across all five **G** matrices, both larger than the median evolvability of size traits reported in Hansen et al. (2011). We note that a potential reason for this high evolvability is the fact that the estimated **G** matrices contain not only additive genetic variances but also dominance and epistatic variance components, colony-specific environmental effects, and potential temporal variance caused by anagenetic evolution and plasticity. Most of the variation in each of the estimated lineage-specific **G** matrices is concentrated among the first four principal components, which account for more than 90% of the total variation in each **G** (fig. 5*A*). We do not find evidence of any differences among the five **G** matrices, as measured by random skewers (Marroig and Cheverud 2001), that cannot be explained by sampling error alone (fig. 5*B*). This suggests that differences in the patterns of genetic association within species are sufficiently small not to be detected given the sample sizes we have available. Phenotypic and genetic variation is distributed unevenly in morphospace (fig. 5*C*, 5*D*).

Using Lande's (1979) equation of multivariate evolution under drift, we find that the observed morphological differentiation within lineages and at speciation events are compatible with a purely stochastic evolutionary process as revealed by the regression tests (95% confidence intervals of slopes all include 1; table 5). Drift can therefore not be excluded as the mechanism explaining both within-lineage evolution and morphological divergence during speciation. However, the 95% confidence intervals of each slope estimate are large (the sample size is small for each regression analysis). Also, finding that the slope is not statistically significantly different from 1 does not exclude other evolutionary mechanisms as potential drivers of the evolutionary changes.

Under the assumption that directional selection drove the observed trait dynamics, we used Lande's (1979) equation of multivariate evolution to reconstruct mean-standardized directional selection gradients (β_μ) based on differences between successive time points in time series (anagenetic selection gradients) and changes in morphology associated with lineage splits (cladogenetic selection gradients). We find no difference in the strength of selection during anagenesis compared with cladogenesis (fig. 6; table 6). We also find that evolution happened almost exclusively in directions with higher-than-average

Table 4: Comparing rates of evolution during anagenesis and cladogenesis

	df	SS	MS	No. iterations	Pr(Prob)
Rates of evolution:					
Anagenesis sequential vs. cladogenetic:					
Factor	1	3.40	3.40	119	.4622
Residuals	105	615	5.86		
Comparable anagenesis vs. cladogenetic:					
Factor	1	.650	.650	56	.6429
Residuals	14	16.5	1.18		
Evolvability:					
Factor	1	.000386	.000386	88	.5341
Residuals	82	.0239	.000292		
Conditional evolvability:					
Factor	1	.0000450	.0000450	51	1.0000
Residuals	82	.00590	.0000720		

Note: Shown are results from the random permutation procedure aimed at testing whether the rates of evolution and/or evolvability statistics in the direction of evolution are different during anagenesis ("comparable anagenetic" and "anagenesis sequential"; see fig. 3) and cladogenesis.

evolvability and conditional evolvability in multivariate space (fig. 7; table 7) compared with random directions in morphospace ($P < .001$). Evolvability and conditional evolvability are not different in directions traveled during anagenetic and cladogenetic evolution (table 4).

Discussion

Given a morphological species concept, lineage splitting is necessarily correlated with morphological change, but to what extent speciation is correlated with larger than usual changes in phenotypes is debated (Lieberman and Eldredge 2014; Venditti and Pagel 2014; Pennell et al. 2014*a*, 2014*b*). Caribbean species of *Metrarabdotos* are generally considered the best example of a punctuated mode of evolution (e.g., Jackson and Cheetham 1999; Gould 2002; Hunt and Slater 2016), and the phylogeny from Cheetham (1986) showing minimal phenotypic change within lineages but large changes during speciation appears in several textbooks for evolutionary biology, paleobiology, and paleontology as an example of what the fossil record teaches us about trait dynamics during anagenesis and cladogenesis (e.g., Ridley 2003; Foote and Miller 2007; Benton and Harper 2009; Futuyma and Kirkpatrick 2017).

We failed to find differences in evolutionary rates between anagenesis and cladogenesis in our reanalysis of a subset of the *Metrarabdotos* data. Separate lineages are not occupying distinct places in multivariate space but show large fluctuations that overlap extensively. Substantial evolution happens during speciation events in *Metrarabdotos*, but the rates of evolution during within-lineage (anagenetic) evolution are overall very similar to the rates of evolution across ancestor and descendant species pairs (cladogenesis). Our finding that both within-lineage evolution and phenotypic diversification during speciation events are compatible with drift supports homogeneity of evolutionary processes during anagenesis and cladogenesis. For the sake of completeness, we also investigated whether we could find indications of differences in the strength of selection (steepness of selection gradients) during anagenesis and cladogenesis. We find that the distributions of selection gradients calculated for cladogenetic and anagenetic events are very similar, indicating that if selection explained these evolutionary events, is not necessary to invoke stronger selection during either of the two modes of evolution to explain the data. Overall, the results of the current study contradict the claim by Cheetham et al. (1994, p. 373) that "agents of speciation are different from the pervasive stabilizing selection required to explain phenotypic stasis with species . . . decoupling speciation from forces acting within species."

The lack of a punctuated mode of evolution in our reanalysis of a subset of the original trait data on *Metrarabdotos* supports the hypothesis that the claimed punctuated mode of evolution within the genus (Cheetham 1986, 1987; Cheetham et al. 1993, 1994, 2007; Jackson and Cheetham 1994) is due to violations of measurement-theoretic principles and other methodological issues, such as the replacement of missing data by the species' means. The measurement theoretical issues include the assignment of arbitrary numerical values to character states of traits on a nominal scale (e.g., position of a structure in the zooid) and treating them as if the numerical magnitude between the attributes had meaning (Cheetham 1986, 1987; Cheetham et al. 1994). Such nominal traits were used in

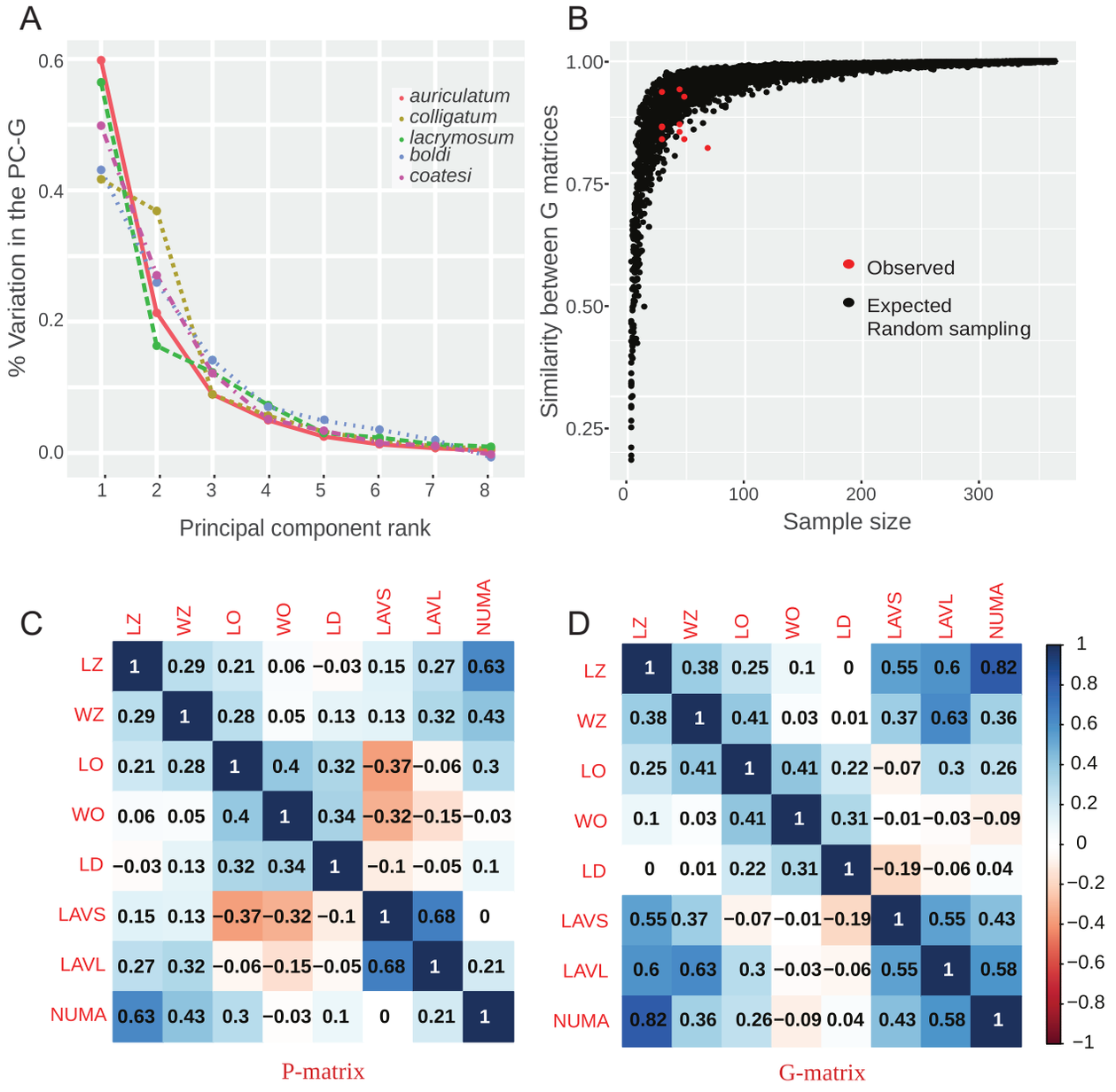

Figure 5: Genetic variation is concentrated in a few axes of the morphospace. *A*, Percent variance explained by each principal component of each of the five species-specific **G** matrices. Each line represents the distribution of genetic variation for each of the five species for which we were able to estimate a broad-sense **G** matrix. *B*, Pairwise similarity, measured through random skewers, between all five **G** matrices (red points) compared with the similarity that would be expected due to sampling (black points). *C*, *D*, Phenotypic (*C*) and genetic (*D*) correlation matrix for the eight morphological traits in the species with highest sample size (*Metrarabdotos auriculatum*). Note that we are using correlation matrices solely to illustrate the overall similarities in trait associations between **P** and **G**. All calculations in the main text were done using species-specific mean-standardized variance-covariance matrices. Trait abbreviations are defined and described in table 3.

discriminant function analyses to delimit species and investigate rates of morphological evolution (Cheetham 1986, 1987) and in estimating variance-covariance matrices (Cheetham et al. 1994). Cheetham et al. (2007, p. 2) argued that the clonal nature of bryozoans produced individuals within colonies (zooids) with enough variation to make traits on a nominal scale behave as continuous variables. However, a trait on a nominal scale does not "behave" as a trait on a ratio scale even though it is treated as one. To avoid making meaningless statements regarding

Table 5: Neutrality tests

Samples, species	*N*	Slope estimate (95% CI)
Anagenetic samples:		
M. auriculatum	19	.96 (.71–1.22)
M. colligatum	15	.83 (.30–1.36)
M. lacrymosum	15	.89 (.68–1.09)
M. boldi	14	.83 (.20–1.47)
M. coatesi	13	.89 (.27–1.52)
Cladogenetic samples:		
Multiple	8	.89 (.26–1.54)

Note: Shown are results from the regression of between-population variances on within-population variances aimed at testing whether genetic drift alone can explain the patterns of evolutionary diversification within lineages and at speciation events. Slope estimate are regression coefficients. CI = confidence interval.

trait differences among populations within and among lineages of *Metrarabdotos*, we were therefore able to study only eight traits from the original data set. While some of the characters we had to eliminate from the data are so-called species-defining traits, the traits we analyze also represent important characters that define and separate *Metrarabdotos* species (Cheetham 1986; Cheetham et al. 1994). Size-related traits are generally considered important characters for bryozoan taxonomy (McKinney and Jackson 1989).

Investigations of a potential link between species diversification and trait evolution have recently enjoyed a revival with the development of new phylogenetic comparative methods. One approach has been to develop models to investigate whether morphological evolution appears gradual (related to time) or punctuated (related to lineage-splitting events) within clades (Bokma 2002; Ingram 2011; Hunt 2013). Variants of such models have been fit to different groups of taxa, but these studies have so far given mixed support of the relative importance of anagenetic and cladogenetic change within clades (Bokma 2002, 2008; Monroe and Bokma 2009; Ingram 2011; Hunt 2013; Ingram et al.

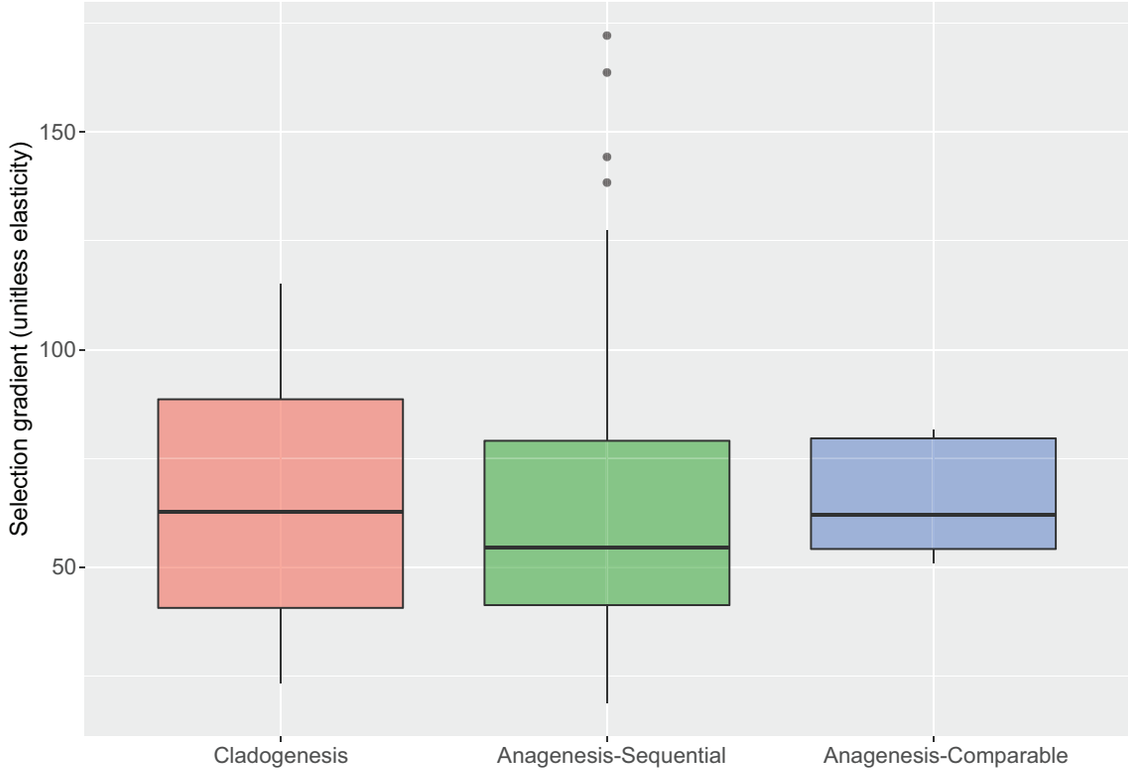

Figure 6: Selection gradients during anagenetic and cladogenetic evolution. Permutation tests revealed no significant differences between selection gradients calculated for anagenetic and cladogenetic evolution (table 6). Selection gradients for sequential anagenetic changes are computed between consecutive ancestor-descendant population pairs within a lineage (see fig. 3*B*). Cladogenetic selection gradients are estimated on the basis of evolutionary changes between the first population (oldest fossil sample) in the descendant lineage and all populations in the ancestral lineage that are older than the descendant population. Selection gradients for "Anagenesis-Comparable" are computed between the population in the ancestral lineage that is older but closest in time to the first appearance of the descendant population and all older populations in the ancestral lineage; they are therefore more or less directly comparable to the selection gradients for the cladogenetic events, as these rates are estimated across time intervals that are similar (often identical) to those used for the rates calculated during cladogenetic events.

Table 6: Comparisons of selection gradients and evolvabilities during anagenesis and cladogenesis

	df	SS	MS	No. iterations	Pr(Prob)
Strength of selection:					
Factor	1	.100	.0600	51	1.000
Residuals	14	903	645		
Evolvability:					
Factor	1	.0000161	.0000161	210	.324
Residuals	88	.00636	.000295		
Conditional evolvability:					
Factor	1	.0000161	.0000161	51	.863
Residuals	88	.00636	.0000722		

Note: Shown are results from the random permutation procedure aimed at testing whether the strength of selection and/or evolvability in the direction of selection are different during anagenesis ("comparable anagenetic"; see fig. 3) and cladogenesis.

2016; for a study where pulses of evolution are not necessarily coupled to cladogenesis but may occur at any point in time, which can be considered a weak version of the punctuated equilibrium hypothesis, see also Landis and Schraiber 2017). An increased rate of evolution during speciation predicts a positive correlation between species diversification and morphological evolution (Ricklefs 2004), a prediction supported by some studies (Rabosky and Adams 2012; Rabosky et al. 2013) but not others (Adams et al. 2009; Venditti et al. 2011). However, if lineages within clades differ in their general ability to adapt and fill ecological niches, differences in such "evolvability" of lineages among clades can potentially explain positive correlations between species diversification and morphological evolution without elevated rates of evolution during speciation (Rabosky 2012). Differences in clade "evolvabilities" can also potentially explain why we observe large differences in species richness among clades. Our results support the idea that trait evolution during and around the time of speciation does not deviate from regular, anagenetically paced evolutionary

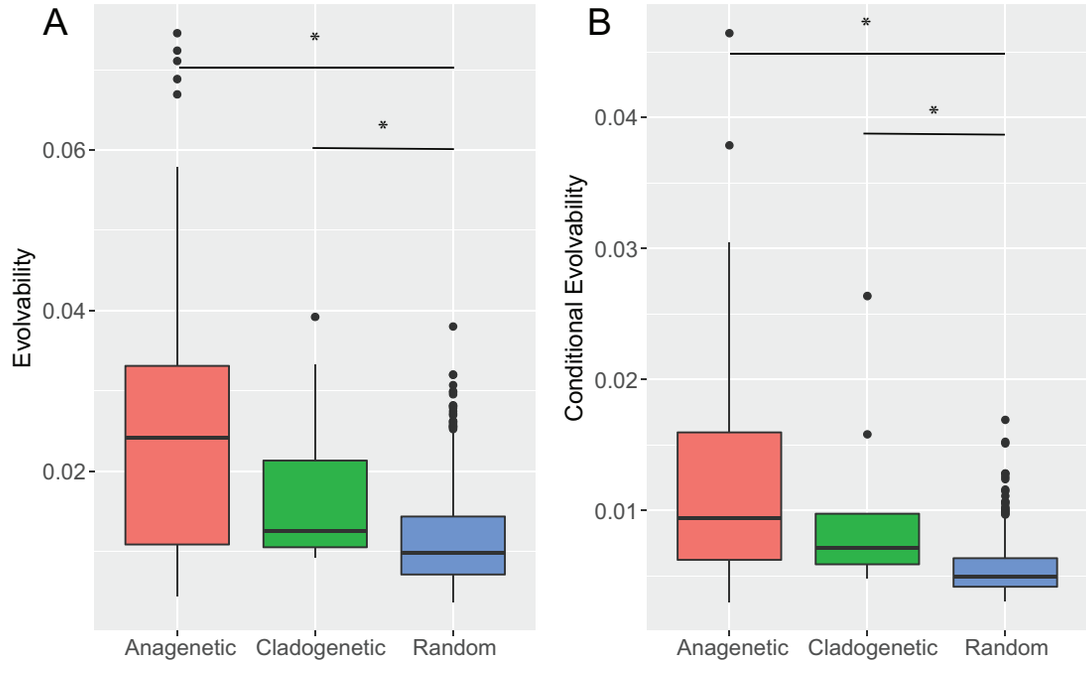

Figure 7: Anagenetic and cladogenetic changes occur in directions of higher than average evolvability. *A*, *B*, Comparison of evolvability (*A*) and conditional evolvability (*B*) along directions of anagenetic (red), cladogenetic (green), and random (blue) directions in the five lineage-specific **G** matrices. Note that for cladogenesis, the selected **G** matrix is the one calculated for the ancestral species. Both anagenetic and cladogenetic changes happen in directions of above-average evolvability. Asterisks represent statistically significant differences (table 7).

Table 7: Comparing evolvability in directions of observed evolution to random directions in morphospace

	df	SS	MS	No. iterations	Pr(Prob)
Cladogenesis vs. random direction:					
Evolvability:					
Factor	1	.0003124	.000312	5,000	.007
Residuals	406	.01675	.0000412		
Conditional evolvability:					
Factor	1	.0001535	.000154	5,000	$<2.2 \times 10^{-16}$
Residuals	406	.002089	.00000515		
Anagenesis vs. random direction:					
Evolvability					
Factor	1	.011888	.0119	5,000	$<2.2 \times 10^{-16}$
Residuals	474	.035119	.0000741		
Conditional evolvability:					
Factor	1	.00306	.00306	5,000	$<2.2 \times 10^{-16}$
Residuals	474	.00721	.0000152		

Note: Shown are results from the random permutation procedure aimed at testing whether the unconditional and conditional evolvabilities of empirical **G** matrices in directions of anagenesis and/or cladogenesis are significantly different from the evolvability of **G** in random directions.

process, which might happen at different speeds in different clades.

Cheetham et al. (1993, 1994, 1995) pioneered the application of quantitative genetics on data from the fossil record, capitalizing on the fact that colonies of bryozoans consist of multiple genetically identical units (zooids). The five (broad-sense and potentially upwardly biased) genetic variance-covariance (**G**) matrices we estimated on the basis of a subset of the original data showed that evolutionary changes during anagenesis and cladogenesis happen almost exclusively in directions with above-average evolvability, while the evolvability in directions traveled during anagenesis and cladogenesis do not differ. Phenotypic evolution during speciation therefore seems to represent "more of the same" as observed during anagenetic change. A possible interpretation of the observation that the directions of evolution align with directions of large genetic variation is that the genetic variance-covariance structure puts constraints on long-term phenotypic evolution (Schluter 1996; Marroig and Cheverud 2005; Porto et al. 2009; Hansen and Voje 2011; Grabowski 2016). Breaking up genetic constraints—or genetic revolutions—have been suggested as important mechanisms facilitating speciation (Mayr 1954; Carson 1975; Templeton 1980). One potential prediction from these hypotheses of speciation via genetic revolutions is that evolution should travel in a direction less influenced by the ancestral **G** matrix during a speciation event compared with the constrained evolution within lineages. We do not find any support for that prediction, as both cladogenetic and anagenetic events happen in directions of above-average evolvability. We note, however, that an alternative explanation for the observation that cladogenetic and anagenetic events happen in directions of above-average evolvability may be due to the fact that the colonies used to estimate trait means and variances at a certain point in time in the fossil record are not from the exact same generation. The broad-sense **G** matrices we estimated may therefore contain variation caused by microevolutionary changes among colonies sampled from the same geological time point, which will bias **G** toward directions where this anagenetic evolution occurred. This might give the impression that anagenetic and cladogenetic changes among different population means also happened in directions with high evolvability, but if they do, however, this can also be interpreted as an argument for constraints on short-term evolution that extends to constraints on macroevolutionary timescales.

We would like to highlight some limitations to our approach, which are shared by most of the original work on *Metrarabdotos*. First, our comparisons of evolutionary dynamics during anagenesis and cladogenesis are based on ancestor-descendant relationships as defined in the phylogenetic hypothesis from the original work on *Metrarabdotos* (Cheetham 1986). Analyzing the evolutionary dynamics of the same traits that were used to build the phylogenetic hypothesis is problematic because of circularity issues; analyzing and comparing evolutionary changes in the same traits that were used to delimit species may create spurious relationships regarding the amount of evolution happening at cladogenesis and during anagenesis. If species are defined on the basis of a set of traits, it is not surprising if those exact traits on average change faster during cladogenesis than during anagenesis. A better approach would have been to establish a phylogenetic hypothesis of *Metrarabdotos* based on molecular sequence and/or morphological data that were not directly used in the study of rates of morphological evolution. Since there are no independent data available to create a phylogeny,

however, we chose to take the original phylogenetic hypothesis at face value. Second, the fossil record on *Metrarabdotos* allows estimation of (broad-sense) **G** matrices of ancestors in hypothesized ancestor-descendant species pairs. We were able to estimate only a single **G** matrix per lineage because of low sample sizes, which means that all results and interpretations based on these matrices hinge on the assumption that they stayed constant over the lifetime of a species. This is a strong assumption, but the failure to detect differences among the five species-level **G** matrices may indicate a rather stable variance-covariance structure within lineages of *Metrarabdotos*. We note, however, that lack of statistically significant differences among the five matrices does not mean that the matrices are truly identical.

Because of the large number of traits we excluded from our reanalysis, we suspect that many will think we have a weak case for claiming lack of evidence for a punctuated mode of evolution within *Metrarabdotos*. We had hoped to be able to analyze more data from the *Metrarabdotos* data set, but an important principle within measurement theory is that scale types need to be respected. The quantitative treatment of the *Metrarabdotos* data in much of the original work that explored evolutionary tempo and mode in the genus assumed that magnitudes between character states had meaning (e.g., Cheetham 1986; Cheetham et al. 1994), but magnitudes between character states within nominal traits—the scale type of the majority of traits in the original *Metrarabdotos* data set—are nonsensical (Houle et al. 2011). Our study should therefore not be understood as a criticism of the *Metrarabdotos* data set in itself, which undeniably represents an important and impressive contribution. Our concern is the validity of the statistical analyses of these data in the context of tempo and mode of morphological evolution. These concerns, together with the extensive problem of missing data, severely reduced the number of traits we were able to reanalyze.

We would also like to make it clear that we do not argue against analyses of traits on nominal (discrete and unordered traits) and ordinal (discrete and ordered traits) scales or that such traits are unimportant in the study of tempo and mode in evolution. Their obvious importance makes it essential to respect the constraints these scale types put on how they can be analyzed in a meaningful way. Methods for calculating rates of morphological evolution of nominal and ordinal traits have been developed (e.g., Pagel 1994; Lewis 2001; Pagel and Maede 2006; Lloyd 2016), but how such rates can be compared in a sensible way to rates of evolution of continuous traits is far from clear. One option is therefore to avoid putting traits on a nominal scale (if possible) when the research question (theoretical context) demands calculating and interpreting statistics such as means and variances. For a positional trait it would be possible to measure the distance from a homologous point on the zooid to a homologous point on the structure itself, which will put the position of the structure on a ratio scale. Another possibility is to conduct a separate analysis of discrete traits on the *Metrarabdotos* lineages, but the challenge regarding the amounts of missing data also applies to this trait category. Whether an analysis of discrete traits can resurrect the genus as a textbook example of a punctuated mode of evolution is an open question. What seems clear, however, is that the traits analyzed in the current study do not suggest a strong link between morphological divergence and speciation.

Acknowledgments

We are indebted to Gene Hunt for providing the raw data on *Metrarabdotos* from the work of Alan H. Cheetham and colleagues. We thank JoAnn Sanner for the scanning electron micrographs of the seven species of *Metrarabdotos* analyzed in this study (fig. 1). The manuscript benefited from comments from Jarrod D. Hadfield, Steven J. Hageman, Thomas F. Hansen, David Houle, Gene Hunt, Lee Hsiang Liow, Fábio Machado, Michael B. Morrissey, Christophe Pélabon, and associate editor Scott J. Steppan. This work was supported by grant 249961 from the Norwegian Research Council to K.L.V.

Data and Code Availability

Data and R code for these analyses have been deposited in the Dryad Digital Repository (Voje et al. 2019; https://doi.org/10.5061/dryad.t4b8gthxm).

Literature Cited

Ackermann, R. R., and J. M. Cheverud. 2002. Discerning evolutionary processes in patterns of tamarin (genus *Saguinus*) craniofacial variation. American Journal of Physical Anthropology 117:260–271.

Adams, D. C., C. M. Berns, K. H. Kozak, and J. J. Wiens. 2009. Are rates of species diversification correlated with rates of morphological evolution? Proceedings of the Royal Society B 276:2729–2738.

Bell, M. A., J. V. Baumgartner, and E. C. Olson. 1985. Patterns of temporal change in single morphological characters of a Miocene stickleback fish. Paleobiology 11:258–271.

Bell, M. A., M. P. Travis, and D. M. Blouw. 2006. Inferring natural selection in a fossil threespine stickleback. Paleobiology 32:562–577.

Benton, M. J., and D. A. T. Harper. 2009. Introduction to paleobiology and the fossil record. Wiley, Oxford.

Benton, M. J., and P. N. Pearson. 2001. Speciation in the fossil record. Trends in Ecology and Evolution 16:405–411.

Bokma, F. 2002. Detection of punctuated equilibrium from molecular phylogenies. Journal of Evolutionary Biology 15:1048–1056.

———. 2008. Detection of "punctuated equilibrium" by Bayesian estimation of speciation and extinction rates, ancestral character states, and rates of anagenetic and cladogenetic evolution on a molecular phylogeny. Evolution 62:2718–2726.

Carson, H. L. 1975. The genetics of speciation at the diploid level. American Naturalist 109:83–92.

Charlesworth, B., R. Lande, and M. Slatkin. 1982. A neo-Darwinian commentary on macroevolution. Evolution 36:474–498.

Cheetham, A. H. 1986. Tempo of evolution in a Neogene bryozoan: rates of morphologic change within and across species boundaries. Paleobiology 12:190–202.

———. 1987. Tempo of evolution in a Neogene bryozoan: are trends in single morphologic characters misleading? Paleobiology 13:286–296.

Cheetham, A. H., and L.-A. C. Hayek. 1988. Phylogeny reconstruction in the Neogene bryozoan *Metrarabdotos*: a paleontologic evaluation of methodology. Historical Biology 1:65–83.

Cheetham, A. H., J. B. C. Jackson, and L.-A. C. Hayek. 1993. Quantitative genetics of bryozoan phenotypic evolution. I. Rate tests for random change versus selection in differentiation of living species. Evolution 47:1526–1538.

———. 1994. Quantitative genetics of bryozoan phenotypic evolution. II. Analysis of selection and random change in fossil species using reconstructed genetic parameters. Evolution 48:360–375.

———. 1995. Quantitative genetics of bryozoan phenotypic evolution. III. Phenotypic plasticity and the maintenance of genetic variation. Evolution 49:290–296.

Cheetham, A. H., J. Sanner, and J. B. C. Jackson. 2007. *Metrarabdotos* and related genera (Bryozoa: Cheilostomata) in the late Paleogene and Neogene of tropical America. Memoir (The Paleontological Society) 67:1–91.

Chiba, S. 1996. A 40,000-year record of discontinuous evolution of island snails. Paleobiology 22:177–188.

Eldredge, N., and S. J. Gould. 1972. Punctuated equilibria: an alternative to phyletic gradualism. Pages 82–115 *in* T. J. M. Schopf and J. M. Thomas, eds. Models in paleobiology. Freeman Cooper, San Francisco.

Foote, M., and A. I. Miller. 2007. Principles of paleontology. W. H. Freeman, New York.

Futuyma, D. J., and M. Kirkpatrick 2017. Evolution. 4th ed. Sinauer, Sunderland, MA.

Gelman, A., and D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7:457–472.

Gingerich, P. D. 1976. Paleontology and phylogeny: patterns of evolution at the species level. American Journal of Science 276:1–28.

———. 1984. Punctuated equilibria—where is the evidence? Systematic Zoology 33:335–338.

———. 2009. Rates of evolution. Annual Review of Ecology, Evolution, and Systematics 40:657–675.

———. 2019. Rates of evolution: a quantitative synthesis. Cambridge University Press, New York.

Gould, S. J. 1980. Is a new and general theory of evolution emerging? Paleobiology 6:119–130.

———. 2002. The structure of evolutionary theory. Harvard University Press, Cambridge, MA.

Gould, S. J., and N. Eldredge. 1977. Punctuated equilibria: the tempo and mode of evolution reconsidered. Paleobiology 3:115–151.

Grabowski, M. 2016. Bigger brains led to bigger bodies? the correlated evolution of human brain and body size. Current Anthropology 57:174–196.

Grabowski, M., and A. Porto. 2017. How many more? sample size determination in studies of morphological integration and evolvability. Methods in Ecology and Evolution 8:592–603.

Hadfield, J. D. 2010. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. Journal of Statistical Software 33:1–22.

Hansen, T. F., and D. Houle. 2008. Measuring and comparing evolvability and constraint in multivariate characters. Journal of Evolutionary Biology 21:1201–1219.

Hansen, T. F., C. Pélabon, W. S. Armbruster, and M. L. Carlson. 2003. Evolvability and genetic constraint in *Dalechampia* blossoms: components of variance and measures of evolvability. Journal of Evolutionary Biology 16:754–766.

Hansen, T. F., C. Pélabon, and D. Houle. 2011. Heritability is not evolvability. Evolutionary Biology 38:258–277.

Hansen, T. F., and K. L. Voje. 2011. Deviation from the line of least resistance does not exclude genetic constraints: a comment on Berner et al. (2010). Evolution 65:1821–1822.

Hopkins, M. J., and S. Lidgard. 2012. Evolutionary mode routinely varies among morphological traits within fossil species lineages. Proceedings of the National Academy of Sciences of the USA 109:20520–20525.

Houle, D., C. Pélabon, G. P. Wagner, and T. F. Hansen. 2011. Measurement and meaning in biology. Quarterly Review of Biology 86:3–34.

Hunt, G. 2004*a*. Phenotypic variance inflation in fossil samples: an empirical assessment. Paleobiology 30:487–506.

———. 2004*b*. Phenotypic variation in fossil samples: modeling the consequences of time-averaging. Paleobiology 30:426–443.

———. 2007. The relative importance of directional change, random walks, and stasis in the evolution of fossil lineages. Proceedings of the National Academy of Sciences of the USA 104:18404–18408.

———. 2013. Testing the link between phenotypic evolution and speciation: an integrated palaeontological and phylogenetic analysis. Methods in Ecology and Evolution 4:714–723.

Hunt, G., M. J. Hopkins, and S. 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 USA 112:4885–4890.

Hunt, G., and G. Slater. 2016. Integrating paleontological and phylogenetic approaches to macroevolution. Annual Review of Ecology, Evolution, and Systematics 47:189–213.

Ingram, T. 2011. Speciation along a depth gradient in a marine adaptive radiation. Proceedings of the Royal Society B 278:613–618.

Ingram, T., A. Harrison, D. L. Mahler, M. del Rosario Castañeda, R. E. Glor, A. Herrel, Y. E. Stuart, et al. 2016. Comparative tests of the role of dewlap size in *Anolis* lizard speciation. Proceedings of the Royal Society B 283:20162199.

Jackson, J. B. C., and A. H. Cheetham. 1990. Evolutionary significance of morphospecies: a test with cheilostome Bryozoa. Science 248:579–583.

———. 1994. Phylogeny reconstruction and the tempo of speciation in cheilostome Bryozoa. Paleobiology 20:407–423.

———. 1999. Tempo and mode of speciation in the sea. Trends in Ecology and Evolution 14:72–77.

Jones, D. 2009. Directional evolution in the conodont *Pterospathodus*. Paleobiology 35:413–431.

Kellogg, D. E. 1983. Phenology of morphologic changes in radiolarian lineages from deep-sea cores: implications for macroevolution. Paleobiology 9:355–362.

Kovarovic, K., L. C. Aiello, A. Cardini, and C. A. Lockwood. 2011. Discriminant function analyses in archaeology: are classification rates too good to be true? Journal of Archaeological Science 38:3006–3018.

Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. Evolution 33:402–416.

Landis, M. J., and J. G. Schraiber. 2017. Pulsed evolution shaped modern vertebrate body sizes. Proceedings of the National Academy of Sciences of the USA 114:13224–13229.

Lazarus, D. 1986. Tempo and mode of morphologic evolution near the origin of the radiolarian lineage *Pterocanium prismatium*. Paleobiology 12:175–189.

Levinton, J. S. 1988. Genetics, paleontology, and macroevolution. Cambridge University Press, New York.

Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Systematic Biology 50:913–925.

Lieberman, B. S., and N. Eldredge. 2014. What is punctuated equilibrium? What is macroevolution? A response to Pennell et al. Trends in Ecology and Evolution 29:185–186.

Lloyd, G. T. 2016. Estimating morphological diversity and tempo with discrete character-taxon matrices: implementation, challenges, progress, and future directions. Biological Journal of the Linnean Society 118:131–151.

Malmgren, B. A., and J. P. Kennett. 1981. Phyletic gradualism in a Late Cenozoic planktonic foraminiferal lineage; DSDP site 284, southwest Pacific. Paleobiology 7:230–240.

Marroig, G., and J. M. Cheverud. 2001. A comparison of phenotypic variation and covariation patterns and the role of phylogeny, ecology, and ontogeny during cranial evolution of New World monkeys. Evolution 55:2576–2600.

———. 2005. Size as line of least evolutionary resistance: diet and adaptive morphological radiation in New World monkeys. Evolution 59:1128–1142.

Marroig, G., D. A. R. Melo, and G. Garcia. 2012. Modularity, noise and natural selection. Evolution 66:1506–1524.

Mayr, E. 1954. Change of genetic environment and evolution. Pages 157–180 *in* J. Huxley, A. C. Hardy, and E. B. Ford, eds. Evolution as a process. Allen & Unwin, London.

McKinney, F. K., and J. B. C. Jackson. 1989. Bryozoan evolution. University of Chicago Press, Chicago.

Melo, D., G. Garcia, A. Hubbe, A. P. Assis, and G. Marroig. 2016. EvolQG—an R package for evolutionary quantitative genetics. F1000Research 4:925.

Monroe, M. J., and F. Bokma. 2009. Do speciation rates drive rates of body size evolution in mammals? American Naturalist 174:912–918.

Pagel, M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proceedings of the Royal Society B 255:37–45.

Pagel, M., and A. Meade. 2006. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. American Naturalist 167:808–825.

Pearson, P. N., and T. H. G. Ezard. 2014. Evolution and speciation in the Eocene planktonic foraminifer *Turborotalia*. Paleobiology 40:130–143.

Pennell, M. W., L. J. Harmon, and J. C. Uyeda. 2014*a*. Speciation is unlikely to drive divergence rates. Trends in Ecology and Evolution 29:72–73.

———. 2014*b*. Is there room for punctuated equilibrium in macroevolution? Trends in Ecology and Evolution 29:23–32.

Plummer, M., N. Best, K. Cowles, and K. Vines. 2006. CODA: convergence diagnosis and output analysis for MCMC. R News 6:7–11.

Porto, A., F. B. de Oliveira, L. T. Shirai, V. De Conto, and G. Marroig. 2009. The evolution of modularity in the mammalian skull I: morphological integration patterns and magnitudes. Evolutionary Biology 36:118–135.

Rabosky, D. L. 2012. Positive correlation between diversification rates and phenotypic evolvability can mimic punctuated equilibrium on molecular phylogenies. Evolution 66:2622–2627.

Rabosky, D. L., and D. C. Adams. 2012. Rates of morphological evolution are correlated with species richness in salamanders. Evolution 66:1807–1818.

Rabosky, D. L., F. Santini, J. Eastman, S. A. Smith, B. Sidlauskas, J. Chang, and M. E. Alfaro. 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. Nature Communications 4:1958.

Ricklefs, R. E. 2004. Cladogenesis and morphological diversification in passerine birds. Nature 430:338–341.

Ridley, M. 2003. Evolution. Blackwell Scientific, Oxford.

Schluter, D. 1996. Adaptive radiation along genetic lines of least resistance. Evolution 50:1766–1774.

Spanbauer, T. L., S. C. Fritz, and P. A. Baker. 2018. Punctuated changes in the morphology of an endemic diatom from Lake Titicaca. Paleobiology 44:1–12.

Templeton, A. R. 1980. Modes of speciation and inferences based on genetic distances. Evolution 34:719–729.

Theriot, E. C., S. C. Fritz, C. Whitlock, and D. J. Conley. 2006. Late Quaternary rapid morphological evolution of an endemic diatom in Yellowstone Lake, Wyoming. Paleobiology 32:38–54.

Venables, W. N., and B. D. Ripley. 2002. Modern applied statistics with S. 4th ed. Springer, New York.

Venditti, C., A. Meade, and M. Pagel. 2011. Multiple routes to mammalian diversity. Nature 479:393–396.

Venditti, C., and M. Pagel. 2014. Plenty of room for punctuational change. Trends in Ecology and Evolution 29:71–72.

Voje, K. L. 2016. Tempo does not correlate with mode in the fossil record. Evolution 70:2678–2689.

Voje, K. L., E. Di Martino, and A. Porto. 2019. Data from: Revisiting a landmark study system: no evidence for a punctuated mode of evolution in *Metrarabdotos*. American Naturalist, Dryad Digital Repository, https://doi.org/10.5061/dryad.t4b8gthxm.

Voje, K. L., J. Starrfelt, and L. H. Liow. 2018. Model adequacy and microevolutionary explanations for stasis in the fossil record. American Naturalist 191:509–523.

Wheeler, B. 2016. lmPerm: permutation tests for linear models. R package version 2.1.0. https://CRAN.R-project.org/package=lmPerm.

References Cited Only in the Online Enhancements

Wright, S. 1934. An analysis of variability in the number of digits in an inbred strain of guinea pigs. Genetics 19:506–536.

Associate Editor: Scott J. Steppan
Editor: Alice A. Winn