Long-term origination-rates are re-set only at mass extinctions

Andrew Z. Krug, David Jablonski
Department of Geophysical Sciences, University of Chicago, 5734 S. Ellis Avenue, Chicago, Illinois 60637, USA.

PART 1: DATA AND METHODS

The stratigraphic ranges of genera were taken from a highly revised version of Sepkoski's (2002) compendium of marine genera (Jablonski et al., 2003; 2006). In the 28 November 2011 version of this database, our revisions raise the total number of fossil bivalve genera from 2861 to 3207, with 1589 additions, deletions, alterations, and refinements of stratigraphic ranges. The absolute ages of genera were assigned using the geologic timescale of Gradstein et al. (2004).

Taxon-age distributions can be used to estimate rates of origination and extinction from the fossil record (Van Valen, 1973; Raup, 1975b, 1978; Van Valen, 1979). Such estimates are generally referred to as *survivorship analyses* (Raup, 1978) because they track the fate of taxa forward or backward in time from a point of reference. Survivorship analyses differ from the more standard interval-by-interval tabulations of first and last appearances in the fossil record in their efforts to determine intrinsic or stochastically constant rates for a given clade (Raup, 1978). Here, the taxonomic rate in question refers to the probability of an event given some amount of elapsed time from a reference time-plane. Interval-by-interval variations in rates are considered stochastic fluctuations about a mean value, whereas changes in the probability of origination/extinction are expressed as inflections in the slopes of survivorship curves (Raup, 1978). Such shifts can be interpreted either as extrinsic forcing (e.g., a shift in the environmental or tectonic conditions in which evolution is occurring) or as an intrinsic biological process related to the age of the cohort [e.g., the probability of branching is higher in younger genera (Raup, 1978)]. These scenarios can be teased apart by analyzing a succession of cohorts. If parallel changes in slope occur in all cohorts that cross a given geologic time interval, the cause is likely related to events within that interval. If inflection points instead occur at similar positions within the age distribution of each cohort, age-dependent processes are supported (Raup, 1978). The analyses presented in the manuscript aim to test for long-term shifts in the intrinsic rate of origination for successions of bivalve cohorts that coexist at stage boundaries. For this purpose, survivorship analyses are preferable to analyses of smaller-scale fluctuations in origination rates associated with a given interval.

Backwards survivorship curves (BSCs) were constructed following Raup (1975a; 1978) and Foote (2001), using the first occurrences of all bivalve genera in the fossil record. Assuming rates are time-specific and taxonomically homogeneous, backwards survivorship curves (BSC's) define an exponential probability function whose slope is governed by the origination rate of the cohort (Foote, 2001). Cohorts in these analyses are defined as the set of genera whose stratigraphic ranges cross a stage boundary

(boundary-crossing cohorts of (Foote, 2001), ensuring that genera within a cohort coexisted. Genera can have stratigraphic ranges that encompass multiple stage boundaries, so rate calculations for successive cohorts will not be completely independent. The taxonomic overlap among cohorts decreases as the amount of time between them increases, so that the taxa shared among cohorts can range from over 90% to under 20%. Despite this wide range of taxonomic similarity, the rate calculations for Cenozoic and Mesozoic are largely invariant, suggesting that non- independence is not determining rates. However, to verify that the lack of independence among the age-frequency distributions is not biasing rate shifts, we calculated stage-level per-taxon rates (Foote, 2000) for all post-Paleozoic stages (Induan to Pleistocene) and averaged the values between major inflection points. These values were then compared directly to the rates calculated from BSC's. Because the estimated rate for the Mesozoic differs depending on whether Cenozoic or Mesozoic cohorts are being analyzed, stage-level per-taxon rates for the Mesozoic were calculated twice, once using only living families (for comparison to rates determined from BSC's of Cenozoic cohorts) and the second using all families that cross the Santonian-Campanian stage boundary (Late Cretaceous, to estimate the rates determined from Mesozoic BSC's). Spearman's rank-order test shows the rates from these 5 intervals to be perfectly correlated (*Spearman rho = 1, p = .017*).

The age of a genus within a cohort is simply the geologic age (in millions of years [Myr]) of the first appearance of the genus minus the age of the stratigraphic boundary defining the cohort. We tallied the proportion of these occurrences found in each geologic time interval proceeding backwards from the present. Plotting these proportions against time produces a curve (BSC) representing the age-frequency distribution of bivalve genera. Assuming that rates are time-specific and taxonomically homogeneous, a BSC defines an exponential probability function governed only by the origination rate (λ) of the cohort (Foote, 2001). The effects of these assumptions have been discussed in depth by various authors (Van Valen, 1973; Raup, 1975b, 1978; Van Valen, 1979; Patzkowsky, 1995; Foote, 2001; Finnegan et al., 2008), including potential influences on the shapes of survivorship curves and estimates of taxonomic rates.

To evaluate the statistical support for potential inflection points within each BSC over a single exponential function, 1-parameter, 3-parameter, and 5-parameter probability functions (equations given below) were fit to the global data using the *optim* function in the software package R (2010). The 3-parameter model involved fitting two separate exponential functions, with independent λ values, separated by an inflection point. The age-frequency distribution density function will therefore be represented by

$$\lambda_1 e^{-\lambda_1 * t} \qquad \text{if } t < t_{crit} \qquad (1)$$

$$\left(e^{-\lambda_1 * t_{crit}} \right)\left(\lambda_2 e^{-\lambda_2 \left(t - t_{crit} \right)} \right) \qquad \text{if } t \geq t_{crit} \qquad (2)$$

where λ_1 and λ_2 are rate estimates (genera per lineage per Myr) and t_{crit} is the inflection point (age in Myr) separating the two functions. The 1-parameter function is represented only by equation (1), and therefore involves no inflection point (Foote, 2001).

For several Cenozoic cohorts and all Mesozoic cohorts analyzed in this study, more than one stage boundary was a local minimum in the time series of supports (Fig 1B). Where two inflection points were apparent, a 5-parameter model was used to re-evaluate the two most supported inflection points and to estimate the origination rates for the 3 segments of the BSC. The density equations are as above, except that a third rate (λ_3) and second inflection point (t_{crit2}) were used, so that for $t \geq t_{crit2}$, the density function would equal

$$\left(e^{-\lambda_1 * t_{crit1}}\right)\left(e^{-\lambda_2\left(t_{crit2} - t_{crit1}\right)}\right)\left(\lambda_3 e^{-\lambda_3\left(t - t_{crit2}\right)}\right) \quad (3)$$

 Any additional inflection points over the two most supported could not be analysed, as the sample sizes for each segment of the BSC becomes too small to estimate rates accurately. Once the maximally supported inflection points were determined, the rates associated with those inflection points were taken as the best estimate of true intrinsic origination rates. Confidence intervals on estimated rates were calculated by bootstrap analysis, resampling from the list of genus ages within each cohort while holding the maximally supported inflection points constant. 1000 repetitions were performed.

Only cohorts containing > 100 genera were analyzed to enhance statistical power in determining inflection points, limiting the analysis to cohorts that cross stage boundaries ranging from the Pleistocene to the Middle Jurassic (Aalenian-Bajocian boundary). Each BSC was first evaluated using the 3-parameter model, which involved optimizing equations 1 and 2 to the raw lists of genus ages separated by an inflection point (t_{crit}). Every stage boundary between the ages of the youngest and oldest genus within the cohort were analyzed (excluding the oldest 1% of the genera), the rates surrounding the inflection point determined, and the support for the model assessed. This analysis resulted in a time series of supports for each inflection point. Because the *optim* function intrinsically minimizes probability functions to a dataset, the local minima in the time series of support values actually represent the stage boundaries with the most support for an inflection point. All local minima that stood statistically outside the distribution of supports for each cohort were used to determine shifts in origination rates, and corrected Akaike Information Criteria (AICc) values were used to evaluate the relative support for the 1-parameter, 3-parameter and 5-parameter models (Burnham and Anderson, 2002).

The raw vector of genus ages, minus the oldest 1% of the genera, was used instead of the cumulative frequency of the ages to avoid biases introduced by the non-independence of different sections of the BSC in the cumulative dataset. Because cumulative proportions must sum to 1, the slope of a BSC in one time interval can affect the slope in another. Fitting the models to the raw list of genus ages avoids this bias, but makes the optimization more sensitive to the resolution of the data. For example, a gap in observed genus first appearances between 64 and 67 Ma in the global dataset, coupled with several originations observed between 67 and 71 Ma, shifts the apparent age of the inflection point in bivalve origination rates slightly from the K/Pg boundary, maximally supporting a t_{crit} at the base of the Maastrichtian Stage at the end of the Cretaceous Period.

Rate shifts following the end-Cretaceous mass extinction were also analyzed for individual bivalve clades, usually ranked as orders (following the revised classification of Bouchet et al. (2010). For these analyses only the maximally supported inflection point of 70.6 Ma (as determined by the analysis of the whole bivalve faunas) was used, as the goal was to dissect the underlying drivers of the major inflection points. Because the phylogenetic partition of the data significantly reduces sample sizes, the analysis was limited to Cenozoic cohorts with 25 or more genera, which limited the analysis to cohorts occurring near the Recent. The Cenozoic and pre-K-Pg rates among included cohorts were then averaged for each bivalve order. However, because of the large size of the order Venerida, its families were informally subdivided according to their phylogenetic affinities (Mikkelsen et al., 2006; Taylor et al., 2007). Venerida1 includes the derived portions of the tree, including the superfamilies Arcticoidea, Chamoidea, Gaimardioidea, Mactroidea, Pholadoidea, Veneroidea, and Ungulinoidea. This subdivision corresponds to the Neoheterodontei division of figure 2 of Taylor et al. (2007). The remaining, more basal branches constitute Venerida2, including the superfamilies Cardioidea, Cyamioidea, Galeommatoidea, Gastrochaenoidea, Glossoidea, Hiatelloidea, Solenoidea, Tellinoidea. The Late Triassic mass extinction could not be analyzed at the level of orders, as order-level cohorts for the Mesozoic were too small to determine robust rates using survivorship analyses.

The stratigraphic durations of genera within intervals were determined by calculating the mean and median durations for bivalves originating within a geological substage. Stratigraphic duration was determined as the separation in millions of years between the first and last appearance of a genus, with the age of each substage taken at the midpoint of the substage. The statistical significance of peaks in stratigraphic duration was tested using a randomization procedure as in previous work (Miller and Foote, 2003). The stratigraphic durations of all genera used for the analyses were pooled and randomly assigned, without replacement, to the list of genera. The analyses were then repeated using the randomized durations. This procedure was repeated 10,000 times and the mean and confidence intervals were computed. Both mean and median durations were tested using this randomization procedure, and the results are comparable.

The number of long-lived genera within a stage is defined as all genera originating within a timebin whose stratigraphic duration is greater than the upper 95% CI determined for that bin by the randomization procedure described above.

PART 2: TABLES AND FIGURES

Table DR1. List of statistically supported inflection points for Cenozoic and Mesozoic cohorts, labelled by the age of the lower boundary of the stratigraphic stage that defines the cohort, using the time scale of Gradstein et al. (2004).

Cenozoic Cohort Age (Ma)	$Tcrit_1$	$Tcrit_2$	ΔAIC_0	ΔAIC_1	ΔAIC_2	Mesozoic Cohort Age (Ma)	$Tcrit_1$	$Tcrit_2$	$Tcrit_3$	ΔAIC_0	ΔAIC_1	ΔAIC_2
2.59	70.6	-	340	0	-	70.6	237	196.5	-	761	0	-3
3.6	70.6	13.65	361	0	-110	83.5	237	196.5	-	813	0	-3
5.33	70.6	13.65	436	0	-107	85.8	237	196.5	99.6	728	0	-2
7.25	70.6	13.65	534	0	-70	89.3	237	196.5	99.6	722	0	-3
11.61	70.6	-	732	0	-	93.5	237	196.5	99.6	669	0	-2
13.65	70.6	-	828	0	-	99.6	237	196.5	-	664	0	-1
15.97	70.6	-	837	0	-	112	237	196.5	-	659	0	-2
20.43	70.6	-	797	0	-	125	237	196.5	145.5	692	0	-2
23.03	70.6	-	819	0	-	130	237	196.5	145.5	656	0	-2
28.4	70.6	-	864	0	-	136.4	237	196.5	145.5	573	0	-2
33.9	70.6	-	884	0	-	140.2	237	196.5	-	527	0	-1
37.2	70.6	-	1015	0	-	145.5	237	196.5	-	586	0	-2
40.4	70.6	-	1067	0	-	150.8	237	196.5	-	683	0	1
						155.7	237	196.5	-	706	0	0
						161.2	237	196.5	-	668	0	-2
						164.7	237	196.5	-	761	0	-3

$Tcrit_1$ designates the oldest inflection point, with $Tcrit_2$ and $Tcrit_3$ representing progressively younger inflection points. Stage boundaries having inflection points include A. Anisian-Ladinian at 237 Ma, B) Hettangian-Sinemurian at 196.5 Ma, C) Tithonian-Berriasian at 145.5 Ma, D) Albian-Cenomanian at 99.6 Ma, and E) Campanian-Maastrichtian at 70.6 Ma. $\Delta AICc$ values show the relative support of models that incorporate zero inflection points ($AICc_0$), 1 inflection point ($AICc_1$), and two inflection points ($AICc_2$). Values for $AICc_0$ and $AICc_2$ are subtracted from $AICc_1$ (shaded column) to show the difference between the support for each model. Note that a single inflection point is always supported over no inflection points, but the relative support for
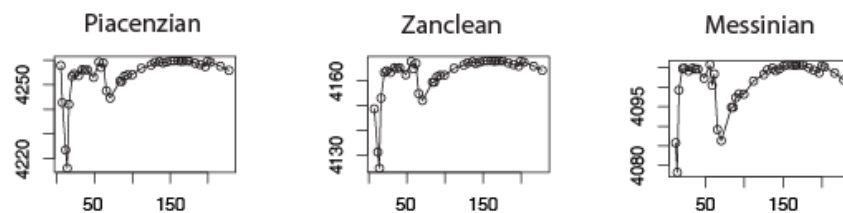
a second inflection point varies among cohorts, with far less added support in Mesozoic cohorts. All Cenozoic $AICc_1$ values were determined for a Tcrit of 70.6 Ma. All Mesozoic $AICc_1$ were determined for a Tcrit of 237. Calculating $AICc_1$ using other statistically defined inflection points does not appreciably change the calculated $AICc_1$ for Mesozoic cohorts, so these values are used for consistency across cohorts. Rates could not be calculated for more than 2 inflection points, so no AICc values are given for the addition of $Tcrit_3$ in Mesozoic cohorts. See model description above for subscript definitions.

Table DR2. Corrected AIC values for linear models correlating the shift in origination rates following the K-Pg boundary with variables related to the number of genera within recovery intervals.
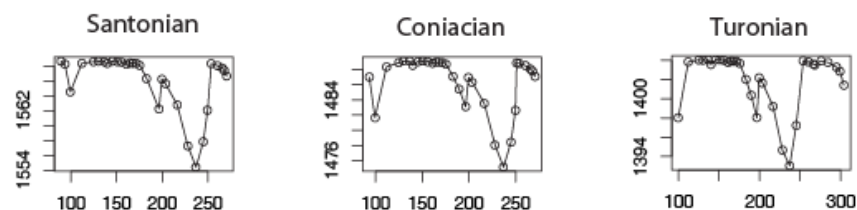
Variables (Shift in Origination =)	AICc
Survivors + Recovery Genera	-62.44
Survivors + Long-Lived Genera	-63.27
Recovery Genera Only	-65.71
Survivors Only	-65.997
Long-Lived Genera Only	-67.06

Lower AICc values indicate greater support for the model. Survivors: number of genera crossing the K-Pg boundary. Recovery genera: all genera that originate within the post-extinction recovery interval. Long-lived genera: number of recovery genera whose stratigraphic duration exceeds the 95% CI for the Danian stage in Figure 3.
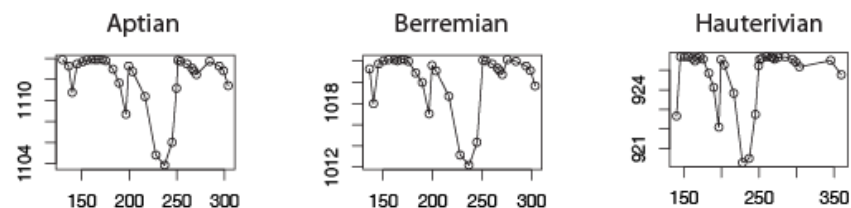
Figure DR1. Time-series of supports for potential inflection points within successions of cohorts that support transient inflection points at the A) Mid-Miocene (13.65 Ma), B) Cenomanian (99.6 Ma), and C) Berriasian (145.5 Ma). Cohorts that cross the stages before and after each of these series do not support the respective inflection points.
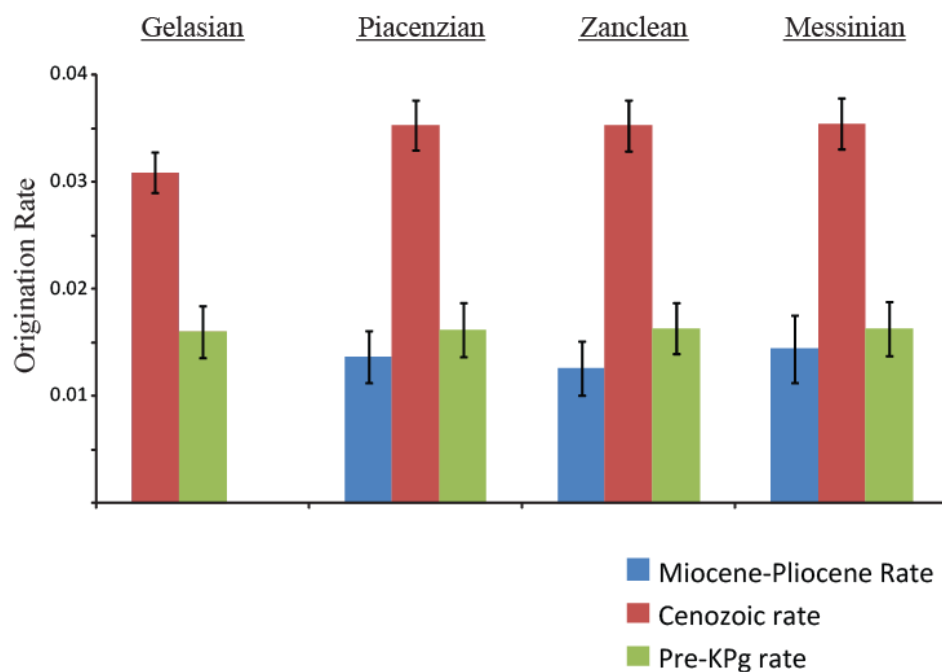
Figure DR2. Rate estimates incorporating a second inflection point at 13.65 Ma for supported Miocene and Pliocene cohorts. The Late Pliocene Gelasian Stage, which does not support the mid-Miocene inflection point, is included to show that the cohort of bivalves that reach the Gelasian stage no longer carry a significant evolutionary signature of decreased mid-Miocene origination. All stage names designate the base of the stage that defines the cohort.
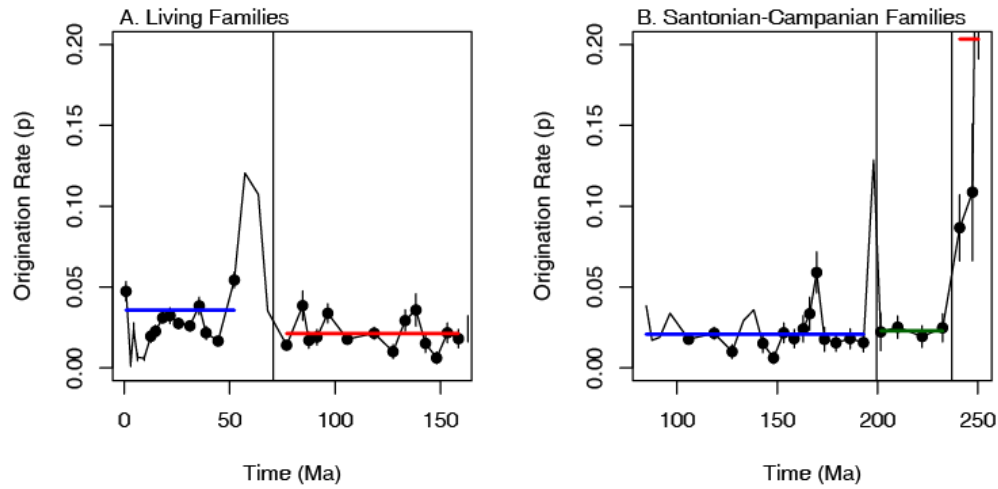
Figure DR3. Stage-level per-taxon origination rates for (A) living families and (B) families extant at the Santonian-Campanian boundary in the Cretaceous. Error bars are 95% CI derived from 1000 bootstrap estimates. Dashed color lines represent the average origination rate for the segment of time encompassed by the line, excluding recovery intervals and other stages that support transient rate shifts in BSCs (portions of line lacking a symbol, Fig DR1). Vertical lines represent stage boundaries designated as inflection points in BSCs. First Triassic stage (Induan) included in time series but not scene due to scaling of the y-axis.

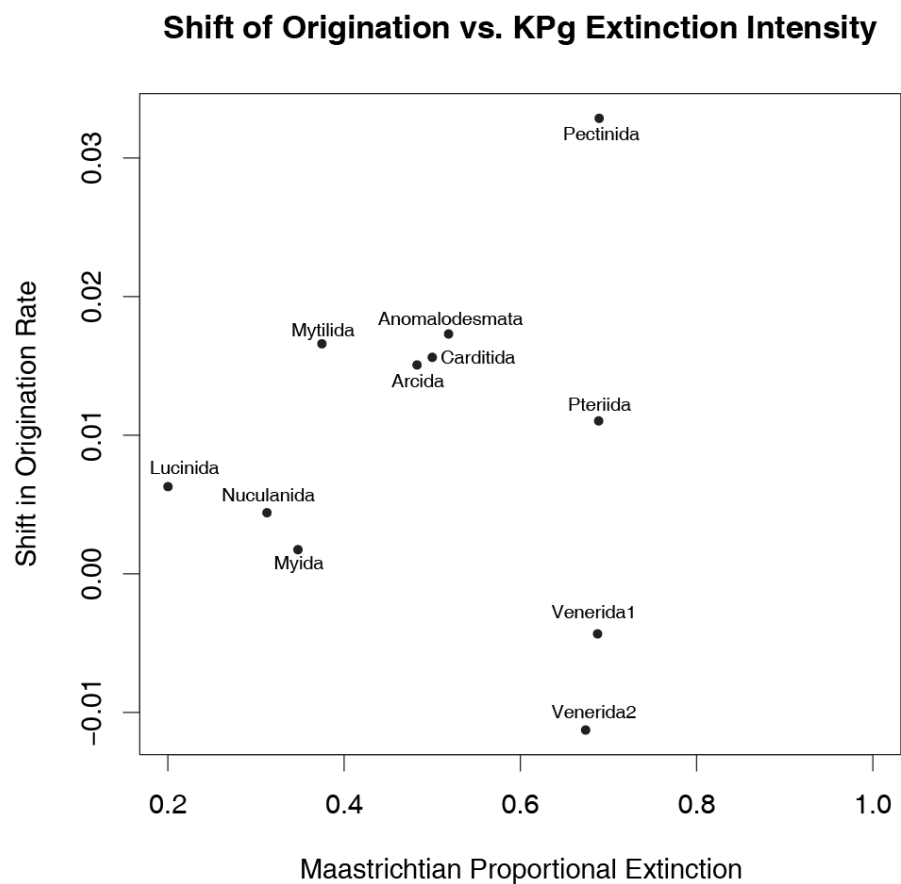**Shift of Origination vs. KPg Extinction Intensity**

Figure DR4. Shift in per-taxon origination rates for bivalve orders following the K-Pg mass extinction versus the proportion of genera within each order that went extinct at the K-Pg boundary. Relationship is statistically insignificant (Spearman *rho* = .2; *p* = .56).

PART 3: REFERENCES

R Developmental Core Team, 2008, R: a language and environment for statistical computing, R Foundation for Statistical Computing, http://www.r-project.org.

Bouchet, P., Rocroi, J., Bieler, R., Carter, J.G., and Coan, E.V., 2010, Nomenclator of bivalve families with a classification of bivalve families: Malacologia, v. 52, p. 1-184.

Burnham, K.P., and Anderson, D.R., 2002, Model Selection and Multimodel Inference: New York, Springer, 488 p.

Finnegan, S., Payne, J.L., and Wang, S.C., 2008, The Red Queen revisited: reevaluating the age selectivity of Phanerozoic marine genus extinctions: Paleobiology, v. 34, p. 318-341.

Foote, M., 2000, Origination and extinction components of taxonomic diversity: general problems: Paleobiology, v. 26 Suppl., p. 74-102.

—, 2001, Evolutionary rates and the age distributions of living and extinct taxa, *in* Jackson, J.B.C., Lidgard, S., and McKinney, M.L., eds., Evolutionary Patterns: Chicago, University of Chicago Press, p. 245-294.

Gradstein, F.M., Ogg, A.J., and Smith, A.G., 2004, A Geologic Time Scale 2004: Cambridge, Cambridge University Press, p. 610.

Jablonski, D., Roy, K., and Valentine, J.W., 2006, Out of the tropics: evolutionary dynamics of the latitudinal diversity gradient: Science, v. 314, p. 102-106.

Jablonski, D., Roy, K., Valentine, J.W., Price, R.M., and Anderson, P.S., 2003, The impact of the Pull of the Recent on the history of bivalve diversity: Science, v. 300, p. 1133-1135.

Mikkelsen, P.M., Bieler, R., Kappner, I., and Rawlings, T.A., 2006, Phylogeny of Veneroidea (Mollusca: Bivalvia) based on morphology and molecules: Zoological Journal of the Linnean Society, v. 148, p. 439-521.

Miller, A.I., and Foote, M., 2003, Increased longevities of Post-Paleozoic marine genera after mass extinctions: Science, v. 302, p. 1030-1032.

Patzkowsky, M.E., 1995, A hierarchical branching model of evolutionary radiations: Paleobiology, v. 21, p. 440-460.

Raup, D.M., 1975a, Taxonomic diversity estimation using rarefaction: Paleobiology, v. 1, p. 333-342.

—, 1975b, Taxonomic survivorship curves and Van Valen's Law: Paleobiology, v. 1, p. 82-96.

—, 1978, Cohort analysis of generic survivorship: Paleobiology, v. 4, p. 1-15.

Sepkoski, J.J., 2002, A compendium of fossil marine animal genera: Bulletins of American Paleontology, v. 363, p. 1-563.

Taylor, J.D., Williams, S.T., Glover, E.A., and Dyal, P., 2007, A molecular phylogeny of heterodont bivalves (Mollusca: Bivalvia: Heterodonta): new analyses of 18S and 28S rRNA genes: Zoologica Scripta, v. 36, p. 587-606.

Team, R.D.C., 2010, R: a language and environment for statistical computing.

Van Valen, L., 1973, A new evolutionary law: Evolutionary Theory, v. 1, p. 1-30.

—, 1979, Taxonomic survivorship curves: Evolutionary Theory, v. 4, p. 129-142.