The Implications of Stratigraphic Compatibility for Character Integration among Fossil Taxa

PETER J. WAGNER1,∗ AND GEORGE F. ESTABROOK2,†

1Department of Paleobiology, National Museum of Natural History, Smithsonian Institution MRC-121, PO Box 37012, Washington, DC 20013, USA and 2Department of Ecology and Evolutionary Biology, University of Michigan, 313 North University, Ann Arbor, MI 48109, USA

∗Correspondence to be sent to: Department of Paleobiology, National Museum of Natural History, Smithsonian Institution MRC-121, PO Box 37012, Washington, DC 20013, USA; E-mail: wagnerpj@si.edu.

†Deceased.

Received 24 January 2015; reviews returned 2 April 2015; accepted 9 June 2015

Abstract.—Two characters are stratigraphically compatible if some phylogenies indicate that their combinations (state-pairs) evolved without homoplasy and in an order consistent with the fossil record. Simulations assuming independent character change indicate that we expect approximately 95% of compatible character pairs to also be stratigraphically compatible over a wide range of sampling regimes and general evolutionary models. However, two general models of rate heterogeneity elevate expected stratigraphic incompatibility. “Early burst” models, where rates of change are higher among early members of a clade than among later members of that clade, and “integration” models, where the evolution of characters is correlated in some manner. Both models have important theoretical and methodological implications. Therefore, we examine 259 metazoan clades for deviations from expected stratigraphic compatibility. We do so first assuming independent change with equal rates of character change through time. We then repeat the analysis assuming independent change with separate “early” and “late” rates (with “early” = the first third of taxa in a clade), with the early and late rates chosen to maximize the probability of the observed compatibility among the early taxa and then the whole clade. We single out Cambrian trilobites as a possible “control” group because morphometric studies suggest that integration patterns are not conserved among closely related species. Even allowing for early bursts, we see excess stratigraphic incompatibility (i.e., negative deviations) in significantly more clades than expected at 0.50, 0.25, and 0.05 P values. This pattern is particularly strong in chordates, echinoderms, and arthropods. However, stratigraphic compatibility among Cambrian trilobites matches the expectations of integration studies, as they (unlike post-Cambrian trilobites) do not deviate from the expectations of independent change with no early bursts. Thus, these results suggest that processes such as integration strongly affect the data that paleontologists use to study phylogeny, disparity, and rates. [Cambrian; correlated change; early bursts; evolvability; fossils; integration; stratigraphic compatibility.]

The simplest models of morphological evolution posit equal rates of independent change among characters over time and across phylogeny. However, it is well established that morphological evolution is much more complex than that. These complexities encompass many basic issues, including simple shifts in overall rates of character evolution (Erwin 1992), “evolvability” making new sets of characters available to some clade members (Wagner and Altenberg 1996), and linkages from common selective factors or the genetic, developmental, and/or functional integration of characters (Olson and Miller 1958). These seemingly distinct issues share one factor in common: rates of change for some characters differ in different portions of a phylogeny. On the one hand, this means that morphological character data offer rich fodder for testing a variety of macroevolutionary hypotheses predicting such rate heterogeneities (i.e., heterotachy sensu Heath and Moore 2014). On the other hand, variable rates and correlated character change induce incorrect phylogeny reconstruction in simulations (Kuhner and Felsenstein 1994; Wagner 1998; Huelsenbeck and Nielsen 1999; Wright and Hillis 2014). Even if we reconstruct general phylogenetic relationships correctly, then rate heterogeneities will complicate tip-dating analyses using morphological data (e.g., Pyron 2011; Ronquist et al. 2012; Heath and Moore 2014). Thus, the ability to recognize complex modes of character evolution serves multiple ends.

Recognizing the effects of rate heterogeneities is particularly important with fossil data. Fossil taxa provide many examples of early bursts in which inferred rates of character evolution decrease over time (Cloutier 1991; Anstey and Pachut 1995; Wagner 1997; Ruta et al. 2006; Brusatte et al. 2008; Lloyd et al. 2012). The associated pattern of high early disparity within clades is an even more general occurrence (Hughes et al. 2013). Fossil data also provide evidence that sets of characters evolve in correlated suites. In some cases, new partitions of character space become available to derived clades, elevating evolvability for the whole clade (Eble 2000; Wagner et al. 2006). Fossils also can document how the conservation of integration patterns changes over time (Webster and Zelditch 2011a, 2011b; Gerber 2013), across phylogeny (Eble 2005; Goswami 2006b; Bell et al. 2011), and among character types (Polly 1998; Webster 2015). Fossils also provide cases connecting integration and evolvability when linked changes increase the subsequent probability of change in previously static characters (Goswami et al. 2014). Although most integration studies rely on morphometric analyses, the models themselves extend to and make predictions about discrete characters (O’Keefe and Wagner 2001; Webster 2007; Dávalos et al. 2014). Moreover, stratigraphic distributions provide information about temporal sequences, which might be a powerful tool if it can be shown
that models such as early bursts and integration make different predictions about the stratigraphic distributions of characters and character combinations than do null models of homogeneous rates and independent change.

In this article, we use stratigraphic compatibility (Estabrook and McMorris 2006; Wagner and Estabrook 2014) to evaluate how frequently patterns in the fossil record deviate from the expectations of homogeneous rate models. Stratigraphic compatibility represents a particular case extension of general compatibility (e.g., Le Quesne 1969): compatible character pairs can be mapped onto some phylogenies with no homoplasy; stratigraphically compatible characters can be mapped onto some trees with no homoplasy and have state-pairs evolve in an order matching the order in which they appear in the fossil record (Fig. 1a,b). Assume that the oldest sampled species possesses the 00 combination (hereafter: state-pair). There are two possible patterns of stratigraphic compatibility: “divergent,” in which taxa with 10 and 01 combinations appear in younger strata (Fig. 1a), and “hierarchical,” in which taxa with the 01 (or 10) combination appear in younger strata and species with 11 combination appear in the youngest strata (Fig. 1b).

Stratigraphic incompatibility is the third alternative (Fig. 1c,e): the case where the intermediate state-pairs (e.g., 10 given 00, 10, and 11) is the last to appear in the fossil record. There are three simple explanations for this. One is that species with 10 evolved before species with 11 did, but we failed to sample early species with 10 (Fig. 1c). The other two explanations invoke both homoplasy and concentrated change early in clade history. In both cases, species with 11 evolve directly from species with 00 (concentrated change). Species possessing 10 evolve later, either from species with 00 (parallelism; Fig. 1d) or from species bearing 11 (reversal; Fig. 1e).

We can extend stratigraphic compatibility to multistate characters using binary breakdowns of those characters. Suppose that we observe taxa with 00, 10, 11, and 12 state-pair (Fig. 1f). There are two binary sets of three state-pairs (00, 10, 11, and 00, 10, 12) in which one state-pair is intermediate between the other two (10 in both cases). Taxa bearing 00 appear first and taxa bearing 11 and 12 appear last. Because the intermediate state-pair (10) appears before the final state-pair (11 and 12, respectively), both breakdowns are stratigraphically compatible. Suppose a different two sets with three combinations: 00, 02, 10, and 00, 10, 11 (Fig. 1g). Taxa with 00 again appear first and taxa with 02 and 10 appear last. The intermediate pair of the first set (00) precedes the other two pairs and therefore is stratigraphically compatible. However, the intermediate pair for the second set (10) appears last and therefore is stratigraphically incompatible.

Given the probabilistic nature of sampling, we expect stratigraphic incompatibility because of missing fossils (e.g., Fig. 1c) at some frequency. Moreover, the probabilistic nature of character evolution means that we expect joint change followed by homoplasy (e.g., Fig. 1d or e) at some frequency under even the simplest models of independent change. However, alternate models explicitly predict evolutionary patterns matching Figure 1d,e. Early bursts should encourage 00→11 transitions on early branches simply by making the probability of two changes relatively high. This sets the stage for subsequent 00→10 or 11→10 transitions. Both correlated change and elevated evolvability scenarios make similar predictions: if the probability of linked characters changing together is much higher than the probability of the individual characters changing independently, then 00→11 transitions preceding 00→10 or 11→10 transitions should be common (see, e.g., Goswami and Polly 2010; Goswami et al. 2014).
The purpose of the article is 2-fold. Our primary goal is to assess whether stratigraphic compatibility in metazoan clades indicates that integration and/or related processes strongly affect the distribution of anatomical character states. However, because the properties of stratigraphic compatibility have been little explored to date, we first must assess how sampling and a variety of alternative evolutionary models besides those matching integration models affect expected stratigraphic compatibility. This is the second aim of our article, which we do using a variety of simulations. We then use Monte Carlo analyses based on these simulations but tailored to individual data sets in order to assess whether we observe too little stratigraphic compatibility in fossil data even after using evolutionary and sampling parameters that minimize the expected stratigraphic compatibility given independent change models. Finally, independent studies indicate that integration patterns differ more among closely related Cambrian trilobite species than among closely related post-Cambrian trilobite species (Webster 2007; Webster and Zelditch 2011a, 2011b). This in turn predicts that stratigraphic compatibility should be greater among Cambrian trilobites than among post-Cambrian trilobites and thus allows us to use trilobites as a control group when assessing our results.

**METHODS**

**Establishing General Relationships between Sampling, Evolutionary Models, and Stratigraphic Compatibility**

Analytic solutions providing exact expectations for compatibility would require integrating over all possible phylogenies (Felsenstein 1981). This is impossible for any large number of taxa (e.g., Felsenstein 1978). Therefore, we use simulations first to assess the effects of different evolutionary and sampling parameters on expected stratigraphic compatibility, and then to assess whether published data sets deviate from null expectations. Paleobiologists have a long history of simulating morphological evolution over phylogenies in order to examine how different evolutionary models and sampling regimes affect patterns such as morphological disparity (e.g., Raup and Gould 1974; Foote 1991, 1994, 1996b). Following these examples, we simulate morphological evolution and phylogeny first to assess the effects of several parameters on expected stratigraphic compatibility and second to provide Monte Carlo tests of whether observed data sets deviate from the expectations of simple models of morphological change.

Our “assessment” simulations run until 32 species are sampled, with extinction rates set to 90% of the origination rates. Unless otherwise stated, sampling is set to 31% of the extinction rate, which means that we expect to sample 31% of species having the median species duration. Unless otherwise stated, we use 100 binary characters with 150 changes among those characters. We use two different branching (speciation) models that have been documented in real clades (see, e.g., Wagner and Erwin 1995; Huang et al. 2015): a bifurcation model in which ancestral morphospecies become pseudo-extinct at speciation (e.g., Slowinski and Guyer 1989), and a budding model in which any morphospecies persist after cladogenesis and can give rise to daughter species at any time over their duration (e.g., Raup and Gould 1974). We describe the “test” simulations (i.e., Monte Carlo analyses) in more detail below. A C program used to conduct these analyses is available in the Supplementary Material available on Dryad at http://dx.doi.org/10.5061/dryad.h5971).

Simulations (Fig. 2a) corroborate the first principle predictions that compatibility should be common among characters with little or no homoplasy (Le Quesne 1969; Estabrook 1983). However, the proportion of compatible character-pairs that are stratigraphically compatible remains very high (over 90%) regardless of the overall frequencies of compatibility (Fig. 2a). Thus, although homoplasy reduces expected compatibility, it does not affect expected stratigraphic compatibility when character-pairs are compatible.

Simulations also show sampling intensity does not affect expected stratigraphic compatibility (Fig. 2b): over a wide range of sampling intensities, we still expect over 90% of compatible character-pairs to be stratigraphically compatible (see also Angielczyk 2002; Wagner and Estabrook 2014). This might seem surprising. However, prior simulation studies show that compatibility among individual characters is highest for those that change infrequently among the sampled taxa (O’Keefe and Wagner 2001; Wagner 2012). From this, it follows that per-taxon rates of change for highly compatible characters are much lower than per-taxon rates of sampling. The absolute rates of sampling or character change are not important: if 30 sampled species represent 45 or 300 total species, then our expectations concern characters that are changing infrequently among the 30 sampled species. Therefore, it is not just that we expect many species retaining an original 00 state-pair to evolve before the first species with a 10 or 01 state-pair evolves (e.g., Estabrook 1977), we also expect to sample multiple species diagnosed by a derived state-pair.

Sampling is not uniform over time and space, or among closely related taxa (Smith 1988; Foote 2001; Wagner and Marcot 2013). However, this has little effect on expected stratigraphic compatibility in simulations. For example, simulations with lognormal variation in sampling rates over time so that 32% of intervals have sampling either less than one half or greater than twice the median sampling rates yield essentially identical results (Fig. 3a). We examine the effects of one particular type of variable sampling model, low sampling early in clade history, below.

How finely we can resolve stratigraphic data varies among fossil groups. For example, in some cases we might resolve first and last appearances only to stages (e.g., the Campanian or Maastrichtian) whereas in other cases we might resolve first and last appearances to
particular zones within stages (e.g., the *Pachydiscus neubergicus* or *Menuites fresvillensis* ammonite zones within the Maastrichtian). Conversely, sometimes we are forced to use coarser stratigraphic resolution (e.g., lumping the Maastrichtian, Campanian, etc. as the Late Cretaceous). Simulations using different levels of stratigraphic resolution with the same origination, extinction, and sampling rates over “one” stratigraphic unit yield similar expectations until sampling is so coarse that fewer than 10% of taxa are expected to survive one interval (Fig. 3b). The logical extreme is using stratigraphic bins so coarse that all first and last are in the same interval (e.g., the “Cretaceous”); in such cases, we cannot detect stratigraphic incompatibility.

A different aspect of sampling is the proportion of characters that fossils yield (Mannion and Upchurch 2010). As the proportion of missing character data increases, the expected stratigraphic compatibility decreases (Fig. 3c). This reflects: (1) missing data causing us to miss “fourth” combinations demonstrating homoplasy and (2) additional opportunities to miss intermediate state-pairs when we cannot observe the
FIGURE 4. Effects of additional evolutionary parameters on expected stratigraphic compatibility. Simulations here assume 150 steps and per-taxon sampling rates = 0.312 × extinction rate. a) Diversity dependent cladogenesis, where rates of diversification start at R but decrease as richness approaches some maximum K. As R/K increases, expected stratigraphic compatibility increases slightly. Exponential diversification is the case where R/K = 0. b) Lognormal variation in rates of change among characters, where m gives the relative rate increase/decrease 1 standard deviation away from the mean. c) Rate shifts contrasting rates during the first third of clade history with that of the last two-thirds. At 1.0, rates are continuous through time; at < 1.0, there are delayed bursts; at > 1.0, there are early bursts. d) Correlated character evolution, in which the probability of change is >3 greater than the probability of independent change. Each character set includes five characters, and each character in these sets has three states. e) Evolvability introducing new character space. We again use sets of five characters with three states each. Here, all five characters shift upon introduction, with independent change afterwards.

The simulations summarized above show a consistent relationship between stratigraphic compatibility and speciation models: bifurcation slightly elevates expected stratigraphic compatibility relative to budding (Figs. 2–4). This reflects the anagenetic component of bifurcation making it impossible to sample ancestral characters on taxa truly possessing the intermediate combinations.
species in younger strata than we sample descendant species. That in turn removes one possible source of stratigraphic incompatibility.

The simulations above assume exponential diversification, which is the simplest diversification model (e.g., Raup 1985). However, fossil data provide considerable support for diversity-dependent models (e.g., logistic diversification; Sepkoski 1978; Alroy 2010). Simulations using logistic diversification, in which net cladogenesis begins at some initial rate of intrinsic diversification (R) and subsequently decreases as richness (S) approaches some threshold (K), generate slightly more stratigraphic compatibility than do simulations using exponential diversification (Fig. 4a; note that exponential diversification is essentially a special case of logistic in which K = ∞ and R/K effectively equals zero). Moreover, expected stratigraphic compatibility rises slightly as R/K increases (and thus as S more rapidly approaches K). This is akin to the effect of ancestral pseudo-extinction in the bifurcation model: when S ≥ K, most speciation will happen only after extinction; this in turn reduces the possibility of sampling taxa (and thus state-pairs) in an incorrect sequence.

The simulations above assume consistent rates of change among characters even though basic rates of change likely vary among characters (Clarke and Middleton 2008; Wagner 2012; Harrison and Larsson 2015). However, rate variation (here using lognormally distributed rates yields expectations essentially identical to uniform rates (Fig. 4b)). Similarly, the simulations above assume consistent rates of change over time although numerous studies provide contrary examples (e.g., Hopkins and Smith 2015). Hughes et al. (2013) suggest that one such pattern, “early bursts” (Gavrilets and Losos 2009), is common in fossil clades. Increasing the relative rate of early change decreases the expected stratigraphic compatibility considerably (Fig. 4c). Note that “early burst” effectively doubles as a “low early sampling model”: if we sample (say) only half the branches among early taxa that we typically sample among later taxa, then we expect approximately twice as much change on these early branches even if true rates were constant through time (e.g., Smith 1988; Wagner 1995a). Thus, both secular decreases in rates of change and secular increases in preservation rates should decrease stratigraphic compatibility.

Finally, the simulations above assume that characters change independently from each other. As discussed in the introduction, many studies provide contrary examples. Simulating character evolution under two models in which sets of characters (and thus probabilities of change) are linked to each other also diminishes expected stratigraphic compatibility. Correlated sets of characters, in which the probability of independent change is much lower than that of mutual change, make it more probable that the first change for Character A accompanies a change for linked Character B instead of happening alone. As the proportion of characters in correlated sets increases, expected stratigraphic compatibility decreases (Fig. 4d). (Simulated characters in correlated suites now have three states, so there is only a 50% chance of reversing to 0 after an initial change to 1.) Similarly, change among linked characters followed by alteration of linkages patterns among those characters (akin to increasing evolvability by altering modules of integrated characters; e.g., Goswami et al. 2014) also increase the probability of introduced Character C changing at the same time as introduced Character D followed by independent reversals. Correspondingly, as the proportion of characters in introduced sets increases, expected stratigraphic compatibility decreases (Fig. 4e).

Monte Carlo Tests

The simulations above set up the basis for Monte Carlo tests for assessing whether we see too little stratigraphic compatibility in data sets given the most “liberal” expectations of independent change (see also Wagner and Estabrook 2014). For any one data set, we can estimate expected stratigraphic compatibility by simulating morphological evolution over phylogenies as follows:

1) Diversification is exponential with budding cladogenesis (both of which minimize expected stratigraphic compatibility relative to alternative models).

2) Per-taxon origination, extinction, and sampling rates follow either empirical estimates given observed stratigraphic ranges (see Foote and Raup 1996) or separate Monte Carlo analyses that find the combinations of relative rates maximizing the probability of the same fossil record.

3) The number and distribution of missing characters matches that of the original data.

4) The number of states per character matches that of the original data.

Morphological evolution continues until the overall compatibility of the simulated matrix matches that of the original data. We repeat these 1000 times to generate a distribution of expected stratigraphic compatibility proportions using C programs written by the senior author.

The analyses outlined above test a null hypothesis of independent change with no major rate shifts. Because rate shifts might be common (see Hughes et al. 2013), and because there might be lingering concerns that sampling is chronically low early in clade histories, we ran a second set of analyses with separate rates for “early” (the first-third of species) and “late” (the last two-thirds of species) taxa. We obtain the relative “early” and “late” rates as follows.

1) We identify the oldest S/3 taxa from a clade of S species, rounding up if S is not evenly divisible
by three. If Species 6–9 in a clade of 20 species all appear in the same interval, then we designate the two species with the lowest average phenetic distances from the initial five species as Species 6 and 7 (Wagner and Estabrook 2014).

2) We run 2500 Monte Carlo simulations to estimate the number of steps maximizing the probability of the observed compatibility among the “early” S/3 taxa (X) and among all S taxa (Y).

3) We approximate the relative “early” rate as \( \frac{2X}{(X-Y)} \). The 2 in the numerator reflects that X steps happen on approximately one half the branches that the subsequent Y – X steps happen.

4) We then re-conduct the first set of Monte Carlo analyses among all S taxa, but with the probability of change on “early” branches being \( \frac{2X}{(Y-X)} \) times greater than that on “later” branches.

These analyses also are conducted in C programs written by the senior author (see Supplementary Material, available on Dryad at http://dx.doi.org/10.5061/dryad.h5971).

**DATA**

We analyze 259 published character matrices originally assembled for phylogenetic reconstruction (Supplementary Material, available on Dryad at http://dx.doi.org/10.5061/dryad.h5971; see also http://www.paleobiodb.org).

We focus on species-level and genus-level analyses. Analyses of suprageneric taxa frequently code taxa based on the type genus of the taxon rather than the earliest species in that taxon. As such, first appearance of the taxon might be much earlier than the first appearances of species in that taxon. We make exceptions in cases where (say) a taxon might be much earlier than the first appearances of species in the type genus of the taxon rather than the earliest species. We do not use suprageneric analyses: outgroups often are token representations of clades and thus might badly misrepresent the appearances of character states. We cull extant species or genera unless they have fossil representatives. In such cases, we use stratigraphic ranges matching their known fossil records rather than assuming that they range through to the present. In order to have a sample size of at least four taxa for the early rate estimates, we analyze only studies with 12+ ingroup taxa after these culls.

We treat multistate characters as unordered (Fig. 1f,g), which maximizes the compatibility of multistate characters (e.g., Meacham 1984; Salisbury 1999). This also allows us to treat each binary breakdown with three state-pairs as a compatible pair, which means that there are theoretically more compatible pairs than there are character pairs. (Note that stratigraphic incompatibility requires three state-pairs regardless of whether binary or multistate characters are involved.) However, we get essentially identical results if we treat these as fractions (e.g., one stratigraphically compatible and one stratigraphically incompatible breakdown = half a stratigraphically compatible pair). In all cases, character pairs are considered stratigraphically incompatible only if the intermediate state-pair appears last: if the intermediate state-pair appears at the same time as another state-pair, then we consider this to be stratigraphically compatible.

Because of the effects of missing data, we cull characters scored for fewer than 10 taxa. For the remaining characters with missing data, we retain missing entries in the same distributions. Thus, if observed taxon 16 is scored as unknown for characters 22–25, then simulated taxon 16 is always scored as unknown for characters 22–25. This maintains not just the proportion of missing data, but also the fact that which types of characters are missing often is very non-random: for example, we typically lose post-crinal elements more often than we lose skull features (Kearny 2002; Sansom et al. 2010; Smith et al. 2014). We retain autapomorphies in the original data sets; however, if no autapomorphies were present in a data set, then the Monte Carlo analyses excluded simulated autapomorphies. Finally, we fixed polymorphic characters to the state that maximized the character’s overall compatibility. Because the probability of character compatibility usually increases as greater proportions of taxa are coded with the same state (Estabrook et al. 1976), this usually meant choosing the state seen in the most taxa. If two or more states left the same overall compatibility for a character, then we chose the oldest appearing state, which maximizes the stratigraphic compatibility of the character.

**RESULTS**

**General Patterns among Major Groups**

For ease of comparison among major taxonomic groups, we divide the data sets into four groups: brachiopods + molluscs, arthropods, echinoderms, and chordates. Although brachiopods and molluscs do not share shells from a common ancestor, both represent “simple” fossils of one or two basic homologies (one or two shells) with many semi-autonomous
regions on those shells. Arthropods, echinoderms, and chordates all possess multi-element skeletons, although serial repetition of hard-parts is common in all groups. We might, therefore, expect different levels of integration among fossilizable parts in these groups. Finally, we also separate Cambrian from post-Cambrian trilobites in order to assess whether the previously documented conservation of integration patterns among post-Cambrian trilobites relative to Cambrian trilobites correctly predict lower stratigraphic compatibility among post-Cambrian trilobites than among Cambrian trilobites.

It is very common for metazoan clades to have less stratigraphic compatibility than predicted by independent change and constant rates (Fig. 5; Table 1). Moreover, it is very common for clades to show major deviations: 119 of 259 clades show deviations that only 65 clades should show (binomial \( P = 2 \times 10^{-13} \)), and 57 of 259 clades show deviations that only 13 clades should show (binomial \( P < 10^{-16} \); Table 1; also, insets of Fig. 7).

There is a strong correlation between the magnitude of best-fit early bursts and excess stratigraphic incompatibility (i.e., negative deviations from expected stratigraphic compatibility) assuming constant rates of
change over time (Fig. 6). This correlation is significant within each taxonomic partition save brachiopods + molluscs. However, Monte Carlo tests assuming these early rates still find too little stratigraphic compatibility among the preponderance of clades, particularly among arthropods, echinoderms, and chordates (Fig. 7). Still, 89 of 259 data sets show deviations that we expect to see from 65 clades (binomial $P = 5 \times 10^{-6}$), and 31 of 259 show deviations expected from 13 clades (Table 1; insets of Fig. 7). Among our partitions, brachiopods + molluscs now provide an exception: the distributions of deviations now fit expectations assuming elevated early rates and independent change.

Cambrian versus Post-Cambrian Trilobites

Within arthropods, Ordovician–Carboniferous trilobites show significantly more deviations at the $P = 0.50$, $P = 0.25$, and $P = 0.05$ levels than expected even given possible early bursts (Table 2; Fig. 8). Conversely, the distribution of deviations for Cambrian trilobites is completely within expectations not just given early bursts, but also under time-homogeneous rates.

**DISCUSSION**

**Alternative Explanations for Excess Stratigraphic Incompatibility**

As we note above, low sampling intensities early in clade histories can mimic “early burst” patterns (Smith 1988; Wagner 1995a, 1997; Ruta et al. 2006). Empirical studies suggest that sampling rates within species are low early in species histories in some clades (Low and Stenseth 2007), although there are exceptions to this (Wagner 2000b). However, this only changes the question to whether peak sampling intensities of early species are lower than that of later species. If not, then this will not create trends in per-taxon sampling over time and thus not mimic early bursts. It is important to distinguish between “per-taxon sampling” and “clade sampling” here. If per-taxon sampling remains constant, then the probability of missing all members of a diversifying clade decreases exponentially as richness increases. A corollary of this is that expected stratigraphic compatibility given our tests assumes lower sampling of clades early in their histories by assuming continuous exponential diversification even while assuming no trends in per-taxon sampling (see, e.g., Foote 1996a). Finally, even if we do attribute apparent early bursts to relatively poor per-taxon sampling, then we still need to explain excess stratigraphic incompatibility given the assumptions of independent change and separate early rates.

It is difficult to attribute our results to assumptions about other general evolutionary processes. We get essentially the same results if we model our “early burst” on the first half or first quarter of taxa as we do using the first third of taxa. Our final tests use parameters (budding speciation, exponential diversification, and separate early vs. late rates of character change) that minimize expected stratigraphic compatibility relative to alternative models (e.g., bifurcating speciation, logistic diversification, and continuous rates of character change). Indeed, this makes our results more conservative. For example, bifurcation and other speciation modes involving anagenetic change likely do occur in some clades examined here (e.g., Pardo et al. 2008). Logistic or other richness-dependent diversification patterns are common for the groups studied here (Miller and Sepkoski 1988; Maurer 1989; Wagner 1995b; Alroy 1996; Ebie 1998; Brayard et al. 2009). Thus, the data probably demand an explanation beyond early bursts (or poor early sampling) even more than our results suggest.

Because our tests retain the number and distribution of missing characters as found in the original data sets, we also cannot attribute the results to missing data. It also is difficult to ascribe these results to the way that paleosystematists collect comparative data. Traditionally, systematists are trained to exclude characters that are obviously correlated with one another (e.g., Wilkinson 1995; Rieppel and Kearney 2002; Harris et al. 2003), these protocols still should bias data sets against the results that we obtain.

**Integration as a Likely Culprit**

Our results suggest that non-independent (=correlated) character change has a strong effect on observed patterns of anatomical characters.
and character states. Mechanisms that could create correlated change include common selective pressures, functional linkages, and integration. Common selective pressures would drive sets of characters to similar derived conditions. Numerous plausible examples of this phenomenon come from the repeated evolution of ecomorphologically important character complexes among fossil (e.g., Cheetham 1987; Alroy 1998; Vermeij and Carlson 2000; Wagner and Erwin 2006; Slater 2015) and living (Losos 1992; Gillespie 2004) taxa. However, pervasive selection favoring particular states also should generate driven trends (i.e., biased change toward particular conditions; McShea 1994). Driven trends in turn should generate copious homoplasy. Alternatively, if multiple characters can combine in different ways to achieve similar performance, then we expect numerous different combinations of these characters (Cheetham 1987; Wagner and Erwin 2006). Both abundant homoplasy and an expectation of many combinations of states for sets of characters greatly decrease expected compatibility (Fig. 2). Because character pairs first must be compatible to stratigraphically incompatible, correlated selection is an unlikely explanation for stratigraphic incompatibility.

A conceptually related alternative explanation is functional linkage of characters (e.g., Wainwright 1988). Complex structures combine multiple homologies into a functional unit (e.g., an appendage), allowing an organism to properly well only with particular combinations of characters. An advantage of this
explanation over common selective pressures is that it predicts (or at least allows for) infrequent changes among linked characters, which would preserve high compatibility. We also might expect it to be more common among echinoderms, arthropods, and chordates than among molluscs and brachiopods because the former group of organisms have many more skeletal parts. However, there are two key failings to this model. One, although functional linkage predicts elevated $00 \rightarrow 11$ shifts, it does not predict the subsequent evolution of $10$ forms; if such forms evolve at all, then we expect them to be intermediate and thus to create stratigraphic compatibility. Two, this model does not offer a ready explanation for why Cambrian trilobites differ from post-Cambrian trilobites as Cambrian trilobites, as both groups should have had similar functional demands on sets of characters. Thus, this explanation also seems inadequate.

Integration models, particularly the developmental or genetic linkage of characters (Olson and Miller 1958; Cheverud 1982, 1984, 1996; Zelditch 1987; Klingenberg 2008), offer another means of non-independent character change. Unlike the common selective forces model, we expect integrated characters to change infrequently: if a favorable modification to Character A coincides with unfavorable modifications to Characters B, C, and D, then selection often will act against the overall change (Lande and Arnold 1983). Unlike functional linkage, integration also allows for reversals in some characters that would generate stratigraphic incompatibility. If change in Character A accompanies change in Characters B, C, and D but there is no functional need to combine the derived states for the four characters, then infrequent independent change for Characters B, C, and D would have high probabilities of persisting if they return to the original states: after all, taxa with those states had been successful in the past. The probability of this happening increases if integration is subsequently reduced among the altered sets (e.g., facial characters in sabre-tooth felids; Goswami 2006b).

In such cases, shifts in integration will simply increase the overall evolvability within a clade (Vermeij 1973; Goswami and Folly 2010; Goswami et al. 2015; see also Fig. 4e). Thus, integration models predict stratigraphic incompatibility much better than do either correlated selection or functional linkage.

Even without first principle arguments, the distinct pattern among Cambrian trilobites suggests an association between how frequently integration patterns change and stratigraphic compatibility. We have no reason to think that correlated selection and/or functional linkages among Cambrian trilobites differed from those among post-Cambrian trilobites. On the contrary, morphometric studies of trilobite populations do suggest that integration patterns changed more often among Cambrian trilobites than among post-Cambrian species (Hughes 1991; Webster 2007; Webster and Zelditch 2008, 2011a, 2011b). Examples of strong phylogenetic conservation of integration patterns also exist for chordates (e.g., Goswami 2006a, 2006b; Sears et al. 2007, 2011; Finarelli and Goswami 2009) and echinoderms (e.g., McKinney 1986; Kjaer and Thomsen 1999). Thus, attributing stratigraphic incompatibility to infrequent shifts among integrated characters offers a simple explanation for the inverse association between phylogenetically conserved integration patterns and stratigraphic compatibility.

Macroevolutionary Implications

Workers have paid much attention to how integration and related concepts such as modularity might limit or encourage the range of character states that a clade can produce (i.e., evolvability; e.g., Riedl 1978; Plotnick and Baumiller 2000; Wagner and Müller 2002; Eble...
Cambrian genera and species a priority to make advanced systematic studies of non-trilobite study simply adds another reason for paleontologists arthropods, echinoderms, and early chordates. Our offer. Thus, it can be applied to groups such as other the dense sampling that Cambrian trilobites commonly compatibility analyses is that our analyses do not require the Cambrian. The one major benefit of stratigraphic taxa even if the phenomenon was relatively common in integration in the preserved systems of other Cambrian Thus, we do not necessarily expect to see “relaxed” conserved in some anatomical systems than in others. However, we would also expect elevated selection favoring change in the less-optimal characters, either by increasing the probability that less-common independent changes are fixed, or by favoring reduced integration among the characters (Monteiro and Nogueira 2010). This would lead to elevated disparity from characters moving on to still new states and stratigraphic incompatibility from characters reverting to the original state. If this model is correct, then future work should show a negative correlation between stratigraphic compatibility and how quickly clades fill character space (see, e.g., Wagner 2000a; Hughes et al. 2013).

Unfortunately, our Cambrian examples include only three non-trilobite clades. Thus, we cannot assess the next obvious question: is low phylogenetic conservation of integration patterns is a Cambrian phenomenon or a Cambrian trilobite phenomenon? Even among Cambrian trilobites, Webster (2015) suggests integration patterns were more tightly conserved in some anatomical systems than in others. Thus, we do not necessarily expect to see “relaxed” integration in the preserved systems of other Cambrian taxa even if the phenomenon was relatively common in the Cambrian. The one major benefit of stratigraphic compatibility analyses is that our analyses do not require the dense sampling that Cambrian trilobites commonly offer. Thus, it can be applied to groups such as other arthropods, echinoderms, and early chordates. Our study simply adds another reason for paleontologists to make advanced systematic studies of non-trilobite Cambrian genera and species a priority.

**Methodological Implications**

The macroevolutionary implications of stratigraphic compatibility patterns have corollary methodological implications. Pervasive stratigraphic incompatibility explains a common mismatch between stratigraphy and inferred phylogeny. The Stratigraphic Consistency Index (SCI; Huelsenbeck 1994) measures how frequently lower nodes on a tree include taxa with older fossil records than taxa within higher nodes. As random error in trees increases, expected SCI goes to 0.5 (Wagner and Sidor 2000). However, Benton et al. (2000; see also Wills 2001) document that over 50% of estimated phylogenies for 10+ species have SCI < 0.5. This suggests that portions of inferred phylogenies are “upside-down”: i.e., they reconstruct old species as derived and young species as primitive. This in turn is consistent with our suggestion that processes such as integration and subsequent evolvability generating numerous 00→11 transitions followed by 11→10 transitions as lineages “stagger” into more optimal character state combinations. If so, then future studies should find correlations between SCI and stratigraphic compatibility.

Pervasive stratigraphic incompatibility re-raises the issue about how we should model morphological evolution. The differences between how we model morphological change in our different simulations point us in a general direction. Consider a model in which all character change reflects independent rates coming from a lognormal distribution (Fig. 9a; see Fig. 4b). If the median rates do not decrease over time, such models generate the very high stratigraphic compatibility we find in most of our simulations (Figs. 2–4). Distributed models often are better than single-rate models for predicting many aspects of character distributions (e.g., Clarke and Middleton 2008; Wagner 2012; Harrison and Larsson 2015). However, independent change models of any sort fail to predict stratigraphic compatibility patterns in fossil taxa other than Cambrian trilobites. Simulations that reduce expected stratigraphic compatibility require fundamentally different models from Figure 9a. In addition to parameters for independent change (the solid curves in Fig. 9b,c), these models require an additional parameter for joint change (the solid bar in Fig. 9b,c). In our correlated change models (Fig. 4d), this solid bar represents the probability of joint change; in our “evolvability” models,
Figure 9 lays out the challenge for morphological systematists: we not only need to identify distributions of independent change among characters, but: (1) sets of characters with some degree of linked change; and (2) how to model rates of joint change among (and even within) different sets of characters. Of course, our study is hardly the first to make this clear: copious “evo-devo” studies over the last decade have made this obvious. Instead, our point is that these models can make distinct predictions about stratigraphic compatibility as well as general compatibility (e.g., Wilkinson 1997, 1998; O’Keefe and Wagner 2001; Dávalos et al. 2014). Thus, stratigraphic distributions of character combinations can and should play an important role in identifying and testing appropriate models even when the fossil record consists of few or even single specimens per taxon.

Under simple models of morphological evolution, the different state-pair combinations of compatible characters should appear in the fossil record in the order in which they evolve in nearly 95% of cases. Thus, compatible character pairs should be stratigraphically compatible. Models of character change predicting that state-pair combinations will appear in orders not matching simple character state trees include elevated rates, correlated change, and increased evolvability. Monte Carlo analyses indicate that stratigraphic compatibility is much lower in brachiopod, mollusc, arthropod, echinoderm, and chordate data sets than expected given independent change and constant rates. Additional analyses allowing for early bursts still indicate that stratigraphic compatibility is too low in arthropod, echinoderm, and chordate data sets. This suggests that mechanisms such as integration that encourage correlated change and (under some circumstances) evolvability affect the distributions of morphological states in the fossil record. This conclusion is further corroborated by Cambrian trilobites being the one major exception to our general results; as is further corroborated by Cambrian trilobites being the one major exception to our general results: as the stratigraphic data can join developmental and morphometric analyses as a powerful tool for developing and justifying complex character evolution models, which in turn will improve phylogenetic and macroevolutionary hypothesis testing.

**CONCLUSIONS**

**SUPPLEMENTARY MATERIAL**

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.h5971.

**ACKNOWLEDGMENTS**

Although this project was completed after George Estabrook’s death, George helped design the basic tests and concurred with the general conclusions from the initial results. P.D. Polly, M. Wilkinson, and M. Hughes provided invaluable constructive reviews; we also thank an anonymous reviewer who suggested that we test the possible effects of early burst models in more detail. For general discussion and/or comments on earlier drafts of the manuscript, we thank J. Alroy, D.H. Erwin, F.R. McMorris, and J. Marcot. For discussion on how to best illustrate linked-change models, we thank S. Darroch, R. Racicot, C. Simpson, S. Tweedt, and R. Warnock. This is Paleobiology Database Publication 228.

**REFERENCES**


Downloaded from http://sysbio.oxfordjournals.org/ by guest on November 8, 2016
Goswami A., Polly P.D. 2010. The influence of modularity on cranial
Goswami A. 2006b. Cranial modularity shifts during mammalian
Gillespie R. 2004. Community assembly through adaptive radiation in
Gerber S. 2013. On the relationship between the macroevolutionary
Foote M., Raup D.M. 1996. Fossil preservation and the stratigraphic
Foote M. 1996b. Models of morphologic diversification. In: Jablonski D.,
Felsenstein J. 1981. A likelihood approach to character weighting and
Felsenstein J. 1978. The number of evolutionary trees. Syst. Zool. 27:
Foote M. 1996b. Models of morphologic diversification. In: Jablonski D.,
Foote M. 2001. Inferring temporal patterns of preservation, origination,
Foote M., Raup D.M. 1996. Fossil preservation and the stratigraphic
Foose M. 1996b. Models of morphologic diversification. In: Jablonski D.,
Foose M. 2001. Inferring temporal patterns of preservation, origination,
Lande R., Arnold S.J. 1983. The measurement of selection on correlated