A sampling-standardized analysis of Phanerozoic marine diversification and extinction

A Working Group proposed by

John Alroy
Department of Paleobiology
Smithsonian Institution
MRC 121
Washington, DC 20560

Charles R. Marshall
Department of Earth and Space Sciences, Molecular Biology Institute, and Institute for Geophysics and Planetary Physics
University of California
Los Angeles, CA 90095-1567

Summary

Our picture of global diversification and extinction on long time scales is mostly based on generalized data for Phanerozoic marine macroinvertebrates. While every effort was made to guarantee the comprehensiveness of this data set, the community has been aware that sampling artifacts may contribute to the observed trends. Until now, we have been unable to remove these effects. Several robust methods for doing this are now available, but these methods use locality-specific data that are not a part of the existing, more generalized compilations. In order to confirm the reality of the major observed patterns, a collaborative data compilation project needs to be initiated. We wish to form a working group to do this. As a first step, we propose a workshop this August involving workers who have specialized in analyzing paleontological diversity data. This workshop will determine the scope, goals, structure, and time table of a database project. Immediately after the workshop, a post-doc who will serve as project coordinator will begin a two-year residency at NCEAS. Over the following two years, experts specializing on particular parts of the fossil record will meet at NCEAS to guide the data collection process. A final meeting will focus on preparing collaborative publications showing how these data influence our picture of marine diversification and extinction.

Problem Statement

Perhaps the most important development in paleoecology over the past two
decades has been the compilation of a global database documenting the
diversification and extinction of marine invertebrates throughout the Phanerozoic
(Flessa and Imbrie, 1973; Sepkoski, 1978, 1993). This database has proven
invaluable in: identifying periods of rapid diversification (Sepkoski, 1978, 1984) and
mass extinction (Raup and Sepkoski, 1984); documenting the existence of three
major "evolutionary faunas" (Sepkoski, 1984); testing models of diversification and
extinction (Sepkoski, 1978; Gilinsky and Good, 1991); and contrasting taxonomic
diversity trajectories with morphological innovations (Foote, 1993), onshore-
offshore trends (Sepkoski and Miller, 1985), and changes in community structure
(Sepkoski et al., 1981; Sepkoski, 1988; Jablonski and Sepkoski, 1996). Indeed, the
master diversity curve for Phanerozoic marine animals may now be the most widely-
reproduced image in the field of paleobiology.

As with any empirical study, it is important to establish that major signals are
not artifactual. Some types of bias have been dealt with. For example, the degree
of sampling per se (Sepkoski, 1993) and the use of higher-level taxa as proxies for
species (Sepkoski and Kendrick, 1993; Wagner, 1995) do not seem to have strongly
distorted the Phanerozoic diversity data. Because of such studies, there is a general
consensus that the major Phanerozoic trends do accurately reflect at least coarse-
scale patterns (Sepkoski et al., 1981).

The purpose of our working group is to grapple with another problem that has
long been suspected but not properly dealt with: the fact that diversification and
extinction trends may be strongly influenced by sampling biases (Raup, 1976;
Sheehan, 1977). A number of paleobiological studies have now shown that
sampling effects are often quite important (Raymond and Metz, 1995; Wing and
DiMichele, 1995; Alroy, 1996; Miller and Foote, 1996). These effects include
variable temporal, geographic, environmental and taxonomic sampling intensities, as
well as biases introduced by geological controls such as sea level changes and
sequence architectures (Holland, 1995). Addressing these problems has been made
feasible by improvements in statistical methods for removing sampling biases (Paul,
1982; Holman, 1985; Marshall, 1990, 1994; Foote and Raup, 1996; Alroy, 1996, in
press; Miller and Foote, 1996; Westrop and Adrain, 1998). Developing a protocol
for dealing with these factors is the primary goal of the workshop that will initiate
this working group.

**Goals and Methods**

Given the importance of the Phanerozoic marine diversity pattern, it is crucial
to establish which of its primary features are real and which are sampling artifacts.
How fast were the Cambrian, Ordovician, and Triassic diversification episodes?
Was there really a diversity plateau throughout most of the Paleozoic? Has diversity
increased exponentially since the early Mesozoic? Are there three dynamically
distinct "evolutionary faunas"? Has local species richness increased in tandem with
global diversification? How did marine communities recover from the five major
mass extinctions? How quickly have extinction rates declined through the
Phanerozoic, and why? Does the selectivity of extinction vary with extinction
intensity? Are extinction and origination rates periodic, and do they show self-
similar or non-linear behavior? These questions about long-term trends only can be
addressed firmly by standardizing sampling across the data set.

The reason that the available data do not lend themselves to straightforward
sampling bias correction is that they consist only of first and last appearances of
taxa on a global basis. More robust sampling correction methods require either
knowledge of gaps in age ranges (Lazarus taxa: Paul, 1982; Holman, 1985; Alroy,
in press), biostratigraphic ranges through measured sections (confidence intervals:
Paul, 1982; Strauss and Sadler, 1989; Marshall, 1990, 1994), or faunal inventories
representing individual fossil localities (rarefaction: Raup, 1975; Alroy, 1996; Miller
and Foote, 1996). This makes building a new, more detailed compilation of the
literature imperative.

Compiling biostratigraphic sections for every part of the Phanerozoic is not a
realistic goal. Merely recording global gaps in age-ranges would be of limited
utility because this would not address the more complex sampling biases of
differential geographic coverage, temporal changes in alpha diversity and
biogeographic provincialism, and differential sampling along paleoenvironmental
gradients. The most realistic approach is rarefaction, which requires compiling
locality-specific faunal inventories and recording geographic, environmental, and
geological data for each of these localities, and ideally, abundance data. Because
rarefaction attempts to standardize the number of taxonomic records or inventories
in each time interval, data collection need only reach a predetermined sampling
target in each of these intervals.

Obviously, compiling all published faunal data for the Phanerozoic marine
record during a two-year project is not feasible. Our goal is much more focused:
because the rarefaction results will only be reliable if the data themselves meet the
method's assumptions, we will put a premium on collecting high-quality faunal
inventories. Thus, the working group's first major goal will be to define criteria for
accepting faunal data, which may include restriction to narrow stratigraphic and
geographic contexts, comprehensive taxonomic coverage, and reporting of specimen
counts. We will prioritize data compilation by focusing on well-studied and well-
preserved taxa, and on geographic regions and environments that are well-sampled
in every, or most, major time intervals.

Because useable results must be obtained quickly, data collection will begin
with the most promising "module." This initial module might represent a synthesis
of faunal lists from just one or two environments on one or two continents, would certainly focus only on marine macroinvertebrates, and would largely rely upon synthesizing pre-existing compilations. In fact, there have been similar, but more limited efforts by others (e.g., Sepkoski and Miller, 1985; Wilson and Palmer, 1990; and especially Bambach, 1977). A quick survey of the literature shows that there are at least 3563 lists in these pre-existing databases (Table 1). Many of these lists could be integrated into the new database, which would minimize the need for primary data compilation.

Nonetheless, there are several reasons why new lists will have to be compiled: 1) certain time intervals, continents, environments, or taxonomic groups have not been surveyed thoroughly; 2) some of the older databases may no longer be accessible; and, 3) some of the databases may not have sufficient information about localities, or might be otherwise incommensurate with our database format.

Based on previous experience, a realistic goal would be to compile about 1000 new faunal lists in two years. Assuming very conservatively that only 500 of the 3563 previously compiled lists prove to be useable, we would have a total of 1500 lists. Following Sepkoski (1984), these might be distributed evenly across the approximately 80 marine stages of the Phanerozoic. This would give us to 15 - 20 faunal lists for each interval, very similar to the rarefaction level used in an analogous study of fossil mammals (Alroy, 1996).
Table 1. Previously compiled data sets that will be useful in this study. Most are restricted to North America, but several are global. Most of the lists pertain to individual faunal horizons in particular outcrops, but some of them pertain to more stratigraphically or geographically generalized collecting localities.

<table>
<thead>
<tr>
<th>Interval</th>
<th>Group/Environment</th>
<th>Lists</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phanerozoic all</td>
<td>all</td>
<td>386</td>
<td>Bambach (1977)</td>
</tr>
<tr>
<td>Phanerozoic hardground</td>
<td>&gt;63</td>
<td>Wilson and Palmer (1990)</td>
<td></td>
</tr>
<tr>
<td>Paleozoic level-bottom</td>
<td>540</td>
<td>Sepkoski and Miller (1985)</td>
<td></td>
</tr>
<tr>
<td>Late Cambrian - Ordovician</td>
<td>trilobites</td>
<td>175</td>
<td>Westrop and Adrain (1998)</td>
</tr>
<tr>
<td>Ordovician all</td>
<td>&gt;400</td>
<td>Miller and Mao (1995)</td>
<td></td>
</tr>
<tr>
<td>Ordovician brachiopods</td>
<td>116</td>
<td>Patzkowsky and Holland (1993)</td>
<td></td>
</tr>
<tr>
<td>Middle Mississippian crinoids</td>
<td>69</td>
<td>Ausich et al. (1994)</td>
<td></td>
</tr>
<tr>
<td>Late Pennsylvanian crinoids</td>
<td>&gt;34</td>
<td>Holterhoff (1996)</td>
<td></td>
</tr>
<tr>
<td>Mesozoic level-bottom</td>
<td>334</td>
<td>Aberhan (1994)</td>
<td></td>
</tr>
<tr>
<td>Jurassic all</td>
<td>331</td>
<td>Tang and Bottjer (1996)</td>
<td></td>
</tr>
<tr>
<td>mid-Cretaceous molluscs</td>
<td>&gt;167</td>
<td>Elder (1989)</td>
<td></td>
</tr>
<tr>
<td>Jurassic - Cenozoic bryozoans</td>
<td>728</td>
<td>Lidgard et al. (1993)</td>
<td></td>
</tr>
<tr>
<td>Maastrichtian bivalves</td>
<td>105</td>
<td>Jablonski and Raup (1995)</td>
<td></td>
</tr>
<tr>
<td>Cretaceous - Paleogene molluscs</td>
<td>&gt;56</td>
<td>Hansen et al. (1993)</td>
<td></td>
</tr>
<tr>
<td>Neogene corals</td>
<td>59</td>
<td>Johnson et al. (1995)</td>
<td></td>
</tr>
</tbody>
</table>

Participants

All of the researchers in the following list have been informally contacted and have expressed strong interest in attending the working group's initial workshop. All of them are experts on analytical methods, relevant field areas, and large paleobiological databases.

- John Alroy, Smithsonian Institution
- Richard Bambach, Virginia Tech
- Karl Flessa, University of Arizona
- Mike Foote, University of Chicago
- Steven Holland, University of Chicago
- Scott Lidgard, Field Museum of Natural History
- Charles Marshall, University of California, Los Angeles
- Michael McKinney, University of Tennessee
- Arnold Miller, University of Cincinnati
- Mark Patzkowsky, Pennsylvania State University
- David Raup, University of Chicago
- Kaustuv Roy, University of California, San Diego
Rationale for NCEAS Support

NCEAS provides three crucial elements that ensure the project's viability.

1) Interaction with ecologists. NCEAS has initiated a working group run by Russell and McKinney that will focus on sampling curves (including rarefaction curves), with the participants all being leading researchers in community ecology. This working group's activities will be largely concurrent with, and complementary to, ours. Dr. McKinney also will be a member of our working group, and both he and the on-site program coordinator will serve as liaisons. These and similar interactions will keep the project's goals focused and in tune with current developments in ecological theory and analysis.

2) Infrastructure for long-term support of the data set. The data need to be made available to the larger scientific community, not just during the working group's lifetime but afterwards. Development of software to analyze the data may be computationally demanding, which would require the use of high-performance computers. The Center's staff and computer facilities will prove invaluable for these purposes.

3) Support for meetings of the working group, and for the group coordinator. Members of the group need a forum for meeting face-to-face to make concrete decisions about implementing the project, and for evaluating its results once the data have been compiled. An on-site coordinator will be crucial to the project's ability to collect data in a timely, focused, and consistent manner.

Proposed Activities and Timetable

Project coordinator.--An efficient compilation procedure would have to be overseen by a coordinator at NCEAS. This would best be a postdoctoral researcher with expertise in analytical paleobiology. The coordinator would presumably be available to begin work in late August or September of this year, and would have to continue for two full years in order to compile a complete, useable data module. The coordinator's responsibilities would include converting contributed data into a uniform electronic format, checking contributed data for formatting and other problems, keeping contributors informed about current priorities for data collection, interacting with specialists brought in to NCEAS to participate in the working
group's activities, and interacting with ecological researchers at NCEAS.

**Initial workshop.**--Major logistical questions need to be resolved before data collection can begin. These decisions need to be made by the researchers who will help to compile the data, and who will eventually analyze them. Therefore, an initial workshop strictly concerned with database issues should begin the working group's activities. This workshop should be held shortly before the coordinator begins residence at NCEAS in August, 1998.

**Additional meetings.**--At later stages, the working group will be extended to include taxonomic and paleoecological experts specializing in particular parts of the fossil record. We envision bringing together three small groups of experts at six-month intervals to focus on particular time intervals (e.g., Paleozoic, Mesozoic, and Cenozoic), taxonomic groups, or other subsets of the database. These meetings would be held in early 1999, the middle of 1999, and early 2000. Following completion of the data collection phase, we would hold a final meeting in the middle of 2000 to begin work on one or several joint publications presenting the major results to the scientific community.

**Anticipated Results and Beneficiaries**

We anticipate that the working group will generate both immediate and long-term results. In the short term, we will generate a data module that will show whether the general features of the Phanerozoic diversity and extinction curves are still seen after standardizing for sampling intensity. These high-quality data will be amenable to paleoecological and macroevolutionary analyses at local, regional, and global scales. In the long term, we hope that the database will serve as a nucleus for an ongoing data compilation effort that will involve much of the paleontological community.

**References**


Ausich, W. I., T. W. Kammer, and T. K. Baumiller. Demise of the middle


Wagner, P. J. 1995. Diversity patterns among early Paleozoic gastropods:


Budget

<table>
<thead>
<tr>
<th>ITEM</th>
<th>YR.1</th>
<th>YR.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Travel</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U.S. (# of participants)</td>
<td>30</td>
<td>16</td>
</tr>
<tr>
<td>(total travel expenses)</td>
<td>$15,000</td>
<td>$8,300</td>
</tr>
<tr>
<td>Foreign</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(# of participants)</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>(total travel expenses)</td>
<td>$5,000</td>
<td>$5,000</td>
</tr>
<tr>
<td>Expenses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Visitor days x $140.00)</td>
<td>$26,640</td>
<td>$16,800</td>
</tr>
<tr>
<td>Salary ***</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(for Center Fellows)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other expenses</td>
<td>None</td>
<td>None</td>
</tr>
</tbody>
</table>

Keywords Worksheet

1. Organizational Focus
   (pick 1)
   -----------------------
   _X Global
   __ Ecosystem
   __ Community
   __ Meta-population
   __ Population
   __ Organismal
   __ Cellular
   __ Molecular

2. Regional Focus
   (pick all that apply)
   -----------------------
   __ California
   __ United States
   __ Southwest
   __ Northwest

4. Taxonomic Group
   (pick all that apply)
   -----------------------
   _X Plants
   __ Invertebrates
     __ insects
     _X other terr. inverts
     __ marine inverts
     __ aquatic inverts
   __ Vertebrates
     __ mammals
     __ birds
     __ reptiles/amphibians
     __ fish
     __ Microbes
     __ Fungi

5. Methods
__ Southcentral                  (pick all that apply)
__ Northcentral
__ Southeast
__ Northeast
__ Africa
__ Antarctic
__ Arctic
__ Asia
Mapping
__ Australia/NZ
__ Canada
__ Central America
__ Europe
__ South America
_X Global

3. Ecological Theme
   (rank up to 3)
   -----------------------
   _X amensalism
   _1 biodiversity
   __ biogeography
   __ commensalism
   __ community dynamics
   __ competition
   __ complex systems
   __ dispersal
   __ disturbance
   __ ecological economics
   __ evolution
   __ genetics
   __ global change
Change/Extinction
   _2 methodological innovation
   __ microclimate
   __ mutualism
   __ nutrient cycling
   _3 paleoecology
   __ plant-animal interactions
   __ population dynamics
   __ predator-prey interactions
   __ primary production
   __ soil processes
   __ succession
   __ parasitism
   __ symbiosis
   __ Other ___________________

6. Research Application
   (rank up to 3)
   -----------------------
   _X acid rain
   __ agriculture
   __ aquaculture
   __ coastal resources
   __ conservation
   __ ecosystem management
   __ energy
   __ environmental policy
   __ fisheries
   __ forestry
   __ global warming
   __ human population
   __ land management
   __ ozone
   __ pollution
   __ reserve siting/design
   __ restoration
   __ toxicology
   _X Other_Global

7. Biomes
   (pick all that apply)
   -----------------------
   _X Marine
   __ Benthic
   __ Pelagic
   __ Intertidal
   __ Terrestrial
   __ Forest
   __ Grassland
   __ Desert
   __ Wetlands
CURRICULUM VITAE

John Alroy

Personal information:
Born 3 July 1966, New York City

Address:
Department of Paleobiology
Smithsonian Institution
MRC 121
Washington, DC 20560
phone: (202) 357-4030
FAX: (202) 786-2832
e-mail: jack@homebrew.si.edu
http://home.dti.net/warr/jack.html

Education:
May, 1989   B.A., Biology, Reed College, Portland, Oregon.
March, 1993 M.S., Evolutionary Biology, University of Chicago, Chicago, Illinois.
Aug., 1994  Ph.D., Evolutionary Biology, University of Chicago, Chicago, Illinois. Title: "Quantitative mammalian biochronology, biogeography, and diversity history of North America."

Employment:
May, 1990   Research Associate, Research Training Group in the Analysis of Biological Diversification, University of Arizona, Tucson, Arizona.
Sept., 1994 - Postdoctoral Fellow, Department of Paleobiology,
Aug., 1996  Smithsonian Institution, Washington, D.C.

Grants, awards, and fellowships:
Feb., 1993   Hinds Fund Award, University of Chicago.
April, 1993  Research Grant, Geological Society of America.
April, 1993  Grant-in-Aid of Research, Sigma Xi.
Papers:
--------. In press. Methods for removing sampling biases from diversity curves. Paleobiology 24(3).
CURRICULUM VITAE

Name: Charles Richard Marshall
Date of Birth: January 13, 1961
Citizenship: Australia/ U.S.A. (dual)

EDUCATION

<table>
<thead>
<tr>
<th>Institution and Location</th>
<th>Degree</th>
<th>Year</th>
<th>Field of Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australian National University, Australia</td>
<td>B.Sc. (Hons)</td>
<td>1984</td>
<td>Math., Paleont., Zoology</td>
</tr>
<tr>
<td>University of Chicago, Chicago, IL USA</td>
<td>M.S.</td>
<td>1986</td>
<td>Evolutionary Biology</td>
</tr>
<tr>
<td>University of Chicago, Chicago, IL USA</td>
<td>Ph.D.</td>
<td>1989</td>
<td>Evolutionary Biology</td>
</tr>
<tr>
<td>Indiana University, IN USA</td>
<td>Postdoc.</td>
<td>1989-1991</td>
<td>Molecular systematics</td>
</tr>
</tbody>
</table>

Employment and Experience

<table>
<thead>
<tr>
<th>Institution and Location</th>
<th>Position</th>
<th>Dates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dept. of Earth and Space Sciences, UCLA, Los Angeles, CA</td>
<td>Full Professor</td>
<td>July 1997 - present</td>
</tr>
<tr>
<td></td>
<td>Vice Chair</td>
<td>April 1997 - present</td>
</tr>
<tr>
<td></td>
<td>Associate Professor</td>
<td>July 1993 - June 1997</td>
</tr>
<tr>
<td></td>
<td>Assistant Professor</td>
<td>Sept. 1991 - June 1993</td>
</tr>
</tbody>
</table>

Dept. of Biology, Indiana University, Bloomington, IN


Affiliations

Research Associate: Invertebrate Paleontology, Museum of Natural History, Los Angeles County (1991 - )
Molecular Biology Institute, UCLA (Member, 1992 - )
Institute for Geophysics and Planetary Physics, UCLA (Member, 1994 - )

Major Grants

- Guggenheim Foundation Fellowship
- Petroleum Research Fund Grant (Type AC)
- SGER Grant, Systematic Biology, NSF
- Petroleum Research Fund Grant (Type G)
- National Young Investigator Award, NSF

Analytical Paleontology 1997-1998
Quantitative Stratigraphy 1996-1998
Middle Repetitive DNA 1995-1997
Quantitative Stratigraphy 1993-1995
Molecular Paleobiology 1992-1997
Publications (only those most relevant to NCEAS Proposal)


Paleontology 8:383-394.