THE NEOGENE MARINE BIOTA OF TROPICAL AMERICA ("NMITA") DATABASE: ACCOUNTING FOR BIODIVERSITY IN PALEONTOLOGY

ANN F. BUDD,1 CHARLES T. FOSTER, JR.,1 JOHN P. DAWSON,1 AND KENNETH G. JOHNSON2

1Department of Geoscience, University of Iowa, Iowa City, Iowa 52242, <ann-budd@uiowa.edu>, <tom-foster@uiowa.edu>, <john-dawson@uiowa.edu>, and 2Geosciences Research Division, Scripps Institution of Oceanography, University of California at San Diego, La Jolla CA 92093-0244, <kenjohnson@ucsd.edu>

ABSTRACT—The reliability of any survey of biodiversity through geologic time depends on the rigor and consistency by which taxa are recognized and samples are identified. The main goal of the Neogene Marine Biota of Tropical America (‘NMITA’) project is to create an online biotic database (http://nmita.geology.uiowa.edu) containing images and synoptic taxonomic information that are essential to collecting and disseminating high-quality taxic data. The database consists of an inventory of taxa collected as part of several large multi-taxa fossil sampling programs designed to assess marine biodiversity in tropical America over the past 25 m.y. In the first phase of the project, data for ~1,300 taxa and ~3,800 images are currently being entered into a relational database management system on an IBM RS6000 at the University of Iowa. Eleven taxonomic groups are represented: bivalves, gastropods (murexids, marginellids, strembinds), bryozoans (cheilostome, cyclostome), corals (azaoxygenellate, zooxanthellate), benthic foraminifers, ostracodes, fish. The lowest taxonomic rank is species (genera/subgenera in mollusks) and the highest is family. Data that are collected and displayed on taxon pages include: (1) taxonomic authority, synonyms, type specimens, and diagnostic morphologic characters; (2) images of representative specimens and associated museum catalog and measurement data; (3) distributional information including geologic ages, stratigraphic units, and spatial locations; and (4) higher level classification (genera and families) and bibliographic information. Illustrated glossaries of morphologic terms, character matrices, and identification tools are being developed for corals and mollusks. Interactive geographic maps and stratigraphic columns have been designed to provide information about taxa collected at different locations.

INTRODUCTION

Over the past decade, quantitative analyses of origination and extinction have become mainstream in paleontology, branching out from the original global assessments of Phanerozoic marine diversity (Raup, 1976; Bambach, 1977; Sepkoski 1978, 1984, 1988; Sepkoski et al., 1981; Sepkoski and Miller, 1985) to include focused studies of selected taxonomic groups within limited geographic areas and intervals of geologic time (e.g., Alroy, 1996; Behrensmeyer et al., 1997; Budd and Johnson, 1999; Jackson and Johnson, 2000; Rosenberg, 1993; Roy et al., 1995, 1996; Wing et al., 1995). These latter, more refined studies provide the resolution essential for assessing biases, and thus for correlating biotic and environmental signals and ultimately interpreting the causes of biotic events, both small and large in scale. As methods of assessing biases in fossil biodiversity estimates have become more complex, so have the databases on which the analyses are based. Some databases are specifically designed to examine sampling questions [e.g., C. Marshall and J. Alroy’s National Center for Ecological Analysis and Synthesis (‘NCEAS’) Paleobiology Database at http://www.nceas.ucsb.edu/public/pmpd/], whereas others focus on the concepts used to distinguish taxa. Here, as part of the current special issue on the evolution of the Late Cenozoic marine biota of tropical America, we take the opportunity to describe an online biotic database designed to do the latter; i.e., to summarize the taxa used in ongoing analyses of fossil biodiversity, the criteria used in their identification, and their occurrences in space and time. This database, named the “Neogene Marine Biota of Tropical America” (NMITA), contains high-quality images and synoptic information on taxa collected as part of several associated multidisciplinary fossil-collecting projects in the Caribbean, Central America, and northern South America (Fig. 1). Many of the contributors to the current special issue are also contributors to the NMITA database.

As databases have increased in complexity, so has the need for collaboration and sharing of high quality primary data among researchers. The NMITA database accomplishes this goal by obtaining data and images from an international group of researchers at various institutions, and making these data and images available on the World-Wide Web (“WWW”, http://nmita.geology.uiowa.edu). It focuses on morphology and recognition of taxa (systematics), and information is derived from specimens collected using standardized sampling procedures. The NMITA database provides a modern electronic alternative to more conventional systematic monographic treatments (e.g., the Paleontological Society Memoir series), and serves as a model for the future dissemination of taxonomic data in paleontology.

Specifically, the NMITA database contains high-quality images and synoptic taxonomic information that are essential to the consistent identification of taxa in collections of tropical American marine fossils. The database focuses on material collected as part of two large team projects (see below) that are unique in their rigorous sampling and age-determination protocols. In combination with the high-quality taxic data provided by NMITA, these collections permit unprecedented opportunities for quantitatively assessing patterns of biodiversity through geologic time. The NMITA database provides an illustrated inventory of taxa in the collections made by these team projects, and it documents taxonomic concepts used in their identification and classification. At minimum, NMITA contains annotated lists of all names that have been used in the identifications, images of representative specimens of each taxon, and a taxonomic bibliography. Wherever possible, it provides illustrated summaries of morphologic characters used in identifying the material, common synonyms, and summary information on the stratigraphic and geographic distribution of each taxon. Distributional data can also be obtained using interactive maps and stratigraphic charts. Online search tools including identification keys and character matrices are available for selected taxonomic groups.

The NMITA database was developed for use by professional paleontologists, students (graduate, undergraduate, K-12), and amateurs. For professional paleontologists, it summarizes the diversity of taxonomic groups at various levels, and the taxonomic concepts used in determining that diversity. For graduate students, it offers a starting point for reconstructing the phylogenies of selected clades, and is important in pointing out diverse and influential clades in need of future systematic study. For all users, it provides tools that can be used to identify specimens in new collections, and familiarizes them with the biota.
that minimize bias; and 3) geologic age dates are obtained for the samples by integrating data on planktonic foraminifera (W. A. Berggren), nannofossils (M. P. Aubry), paleomagnetics (D. F. McNeill), and strontium isotopes. The resulting sampling and locality data have been entered into the PPP database, which is managed by L. S. Collins at Florida International University. At present, these data include >500 bulk samples and >500 bags of individual specimens, containing over a million specimens. After collection, samples have been sent from STRI to designated PPP systematics and research teams for preparation and identification (Table 1). More detailed studies of the systematics and evolution are currently in progress for selected taxa. After study, all micro- and macrofossils, except mollusks and fish, are deposited at the U.S. National Museum of Natural History (‘’NMNH’’); mollusks and most fish specimens are deposited at the NMB. Further details concerning the current status of collections and specimens are provided on the NMITA homepage for each taxonomic group. The PPP collections have served as the basis for two volumes: 1) Jackson et al. (1996); and 2) Collins and Coates (1999). Other references are listed on the PPP web site (http://www.fi.edu/~collinsl/).

The Dominican Republic Project.—The DR project was one of the first large multidisciplinary projects collecting fossils through a continuous 5–10 m.y. Neogene Caribbean sequence ranging from early Miocene to early Pliocene in age. During three field seasons in 1978–1980, a small field party, led by J. B. Saunders and P. Jung of the NMB, measured sections and collected large samples of micro- and macrofossils at closely spaced intervals along nine river sections in the Cibao Valley of the northern Dominican Republic. Geologic age dates were determined using planktonic foraminifera and nannofossils. The collections were accessioned by the NMB, and distributed to specialists for identification and preparation of systematic monographs. To date, 17 monographs have been published in the Bulletins of American Paleontology series “Neogene Paleontology of the Northern Dominican Republic.” The DR collections have served as the basis for classic empirical studies in evolutionary paleontology, including Cheetham’s (1986) demonstration of punctuated equilibrium in bryozoans and Jung’s (1989) revision of the Strombina group. Collecting in the DR project has been recently resumed by C. M. Tang and R. H. Nehm as part of an ongoing study assessing community stability within the Río Gurabo section.

Other related field sampling projects.—Taxa collected as part of associated projects using equally rigorous sampling and age-dating protocols [e.g., collections made in Jamaica, Curacao, and Bahamas by A. F. Budd (Budd et al., 1998; Budd and McNeill, 1998), and in Venezuela and Trinidad by J. B. C. Jackson] have also been included.

Overview of the six main taxonomic groups.—During the first stage of the project (through January 2001), data are being entered for ~1,300 taxa (~3,800 images) within a total of six major taxonomic groups: scleractinian coral species (zoanthellate and azooxanthellate), bryozoan species (cheilostome and cyclostome), mollusks (bivalve genera and subgenera; cumbellid, muricid, marginellid gastropod species), benthic foraminifera species, ostracode species, and teleost and elasmobranch fish species (Table 1). These taxonomic groups include two kingdoms (Protista, Animalia), one of which (Animalia) contains five phyla. They represent all abundant phyla in the collections, except calcareous nanoplankton.

STRUCTURE OF THE NMITA PROJECT

The NMITA database currently runs on an integrated WWW and enterprise (=UNIX-based) relational database management system (“RDBMS”) server at the University of Iowa. The main hardware consists of IBM RS6000 43P-140 computer; the web
<table>
<thead>
<tr>
<th>Taxonomic group</th>
<th>Primary contributors</th>
<th>Estimated number of neogene and quaternary tropical American taxa</th>
<th>Estimated number of images</th>
<th>Collecting projects</th>
<th>Institution housing the collections</th>
<th>Dynamic search</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benthic foraminiferastrans</td>
<td>L.S. Collins, Florida International Univ.</td>
<td>100 species</td>
<td>250 (all SEM)</td>
<td>PPP</td>
<td>NMNH (Paleobiology)</td>
<td><a href="http://nmita.geology.uiowa.edu:8001/ows-bin/owa/...">http://nmita.geology.uiowa.edu:8001/ows-bin/owa/...</a></td>
</tr>
<tr>
<td>Cheilostome bryozaans</td>
<td>A. H. Cheetham, J. Sanner, NMNH</td>
<td>300 species</td>
<td>900 (all SEM)</td>
<td>DR, PPP; Venezuela &amp; Trinidad</td>
<td>NMNH (Paleobiology); NMB</td>
<td>currently static pages only</td>
</tr>
<tr>
<td>Cyclostome bryozaans</td>
<td>P.D. Taylor, BMNH</td>
<td>30 species</td>
<td>150 (all SEM)</td>
<td>PPP</td>
<td>NMNH (Paleobiology)</td>
<td>cyclo_species</td>
</tr>
<tr>
<td>Corals (zooxanthellate)</td>
<td>A.F. Budd, Univ. Iowa</td>
<td>175 species</td>
<td>600 (400 light, 200 SEM)</td>
<td>DR, PPP; Jamaica, Curacao, Bahamas</td>
<td>NMNH (Inv. Zoology); NMB</td>
<td>type_genus; coral_species</td>
</tr>
<tr>
<td>Corals (azoxanthellate)</td>
<td>S.D. Cairns, NMNH</td>
<td>50 species</td>
<td>200 (100 light, 100 SEM)</td>
<td>DR, PPP; Jamaica, Curacoe</td>
<td>NMNH (Inv. Zoology); NMB</td>
<td>currently static pages only</td>
</tr>
<tr>
<td>Fish (elasmobranch teeth; teleostean otoliths &amp; teeth)</td>
<td>O. Aguilera, UNEFM</td>
<td>230 species</td>
<td>405 (369 light, 36 SEM)</td>
<td>DR, PPP; Venezuela &amp; Trinidad</td>
<td>NMB, UNEFM</td>
<td>otolith_family; otolith_species</td>
</tr>
<tr>
<td>Bivalve mollusks</td>
<td>J. Todd, BMNH</td>
<td>299 genera &amp; subgenera</td>
<td>900 (600 light, 300 SEM)</td>
<td>PPP &amp; STRI Reference collections</td>
<td>NMB</td>
<td>biv_idform</td>
</tr>
<tr>
<td>Gastropod mollusks (marginellids)</td>
<td>R. Nehm, NMB</td>
<td>120 species</td>
<td>480 (360 light, 120 SEM)</td>
<td>DR, PPP</td>
<td>NMB</td>
<td>marg_genera</td>
</tr>
<tr>
<td>Gastropod mollusks (murencids)</td>
<td>D. Miller, NMB</td>
<td>120 species</td>
<td>480 (360 light, 120 SEM)</td>
<td>DR, PPP</td>
<td>NMB</td>
<td>murencid_genera</td>
</tr>
<tr>
<td>Gastropod mollusks (colmbellids)</td>
<td>H. Fortunato STRI; P. Jung, NMB</td>
<td>120 species</td>
<td>480 (360 light, 120 SEM)</td>
<td>DR, PPP</td>
<td>NMB</td>
<td>hstrombiniids</td>
</tr>
<tr>
<td>Ostracodes</td>
<td>T.M. Cronin, USGS; P. Borne, Louisiana State Univ.</td>
<td>300 species</td>
<td>600 (all SEM)</td>
<td>DR, PPP</td>
<td>NMNH (Paleobiology); NMB</td>
<td>currently static pages only</td>
</tr>
</tbody>
</table>
server is Oracle Web Application Server 2.1, and the database software is Oracle 7.3. A team of students at the University of Iowa is responsible for designing web pages and the database, processing images, and creating Procedural Language/Standard Query Language ("PL/SQL") code to query the Oracle database.

As described above, data are contributed to NMITA by specialists at different institutions, who are experts in the systematics of a given taxonomic group (Table 1). Contributors submit data and images in a print-oriented format (usually Tagged Image File format or "TIFF") format to NMITA staff in Iowa, who process the images and import the data into Oracle. Contributors typically submit images in 8-bit grayscale mode (black and white photos) or in 24-bit true color mode (color photos). The resolution depends on the features that are illustrated in the image, but are usually ~2,000 × 2,000 pixels. After final processing, all images are converted into an on-screen graphics format (usually Graphics Interchange Format or "GIF") at three different sizes: 1) regular size or 450 × 450 pixels (file size ~100–300 K); 2) large size or 900 × 900 pixels (file size ~1 MB); and 3) thumbnail size or 150 × 150 pixels (file size <20 KB). The original TIFF files are saved as an archive on CD-ROM.

Contributors also provide at least the following textual information: 1) a hierarchical list of taxa (including at least the following ranks: Class, Order, Family, Genus); 2) a bibliography including all systematics citations (in Journal of Paleontology format); 3) spreadsheet tables containing information about the image and the specimen in the image, type specimens and authorship.
of each lowest-level taxon, and localities at which each lowest-level taxon occurs.

THE DATA

Taxon pages.—At the heart of NMITA are taxon pages representing the lowest taxonomic rank that is currently being used in analyzing biodiversity within a given taxonomic group. Also available are pages listing and comparing lowest-rank taxa within an appropriate higher rank. For most taxonomic groups, these pages are at the species and genus level. One exception is mollusks where the lowest rank is genera and subgenera, and the higher rank is families.
FIGURE 5—Reduced database model (10 tables, 80 fields) currently being used to generate web pages for the zooxanthellate coral family Faviidae.

Lowest-rank taxon or species pages (Fig. 2) summarize nomenclatural, descriptive, and distributional criteria that are useful in identification. They contain:

- **Name**—author—date information: As used in analyses of biodiversity based on the data.
- **Images of representative specimens**: In a range of standard orientations and prepared using various techniques. In zooxanthellate corals, for example, whole colony, calical surface, thin-sections, field photos are included. Each image is linked to information about the image and illustrated specimen (the museum catalog number, type, and locality information, scale and orientation).
- **Synonyms**: Different names applied by other authors, with an emphasis on the last 50 yr.
- **Type specimen information**: Museum catalog number, type locality information.
- **Morphologic information**: A list of characters and their states. The characters are linked to a glossary, which provides a complete list of possible states for each character.
- **Distribution (Occurrences)**: Geologic Ages, Formations, and general geographic locations.

In many respects, the information provided on these lowest-rank pages is similar to that of a traditional species description in a printed monograph. The main difference is that information in NMITA is contained in a database, and queries of the database are used to generate static web pages and to perform dynamic online queries. Unlike a printed monograph, NMITA is regularly updated to include recent discoveries and developments as new material is identified. At present, NMITA is still in the initial stages of data entry, and has not yet been subjected to the peer-review process. In future years, we plan to overcome this problem by developing a system for regular peer review, which would be similar to those conducted by professional scientific journals.

In addition, in NMITA, three components of a traditional species description involve links to other databases:

1. **Morphologic Description**: In NMITA, character information is highly structured and formatted as a character matrix. The characters and their states are linked to a glossary, which provides a complete list of possible states for each character.
2. **Material**: Instead of a list of specimens, NMITA provides links to specimen databases, such as the “Cenozoic Coral Database” in Microsoft Access, which is available for downloading from NMITA. Future plans involve links to museum catalogs.
3. **Individual occurrences and interpreted distributions**: NMITA provides links to datasets with occurrence information that is used in larger-scale assessments of biodiversity through geologic time. One such dataset is Cenozoic coral compilation of species occurrences “cc97” (based on Budd et al., 1994, and used in Budd and Johnson, 1999), which is available for downloading from NMITA.

Higher-rank or genus taxon pages (Fig. 3) contain:

- **Name**—author—date information, synonyms, other nomenclatural information.
- **Thumbnail photos of each lowest-rank taxon within the given higher-rank taxon**.
- **Character matrices summarizing morphologic differences among lowest-rank taxa**.

**Database model and organization.**—Taxon pages are generated by a relational database linking tables containing taxon, morphologic, specimen, and occurrence data. The full database model consists of 71 tables and is patterned after the 1993 Association of Systematics Collections (“ASC”) Information Model for Biological Collections (http://biodiversity.uno.edu/). Detailed information concerning the NMITA model, including table and field definitions, is available on the NMITA web site. The NMITA database model consists of six major subject areas: taxon, specimen, locality, morphology, illustrations, and literature (Fig. 4). It expands on the ASC model by adding fields for morphologic characters, measurements, and illustrations (e.g., photographs, drawings, and diagrams).

Reduced, denormalized versions (i.e., using fewer tables, which contain redundant fields) of the full model are currently being used for prototype development. Although their basic structure is similar, different models and databases are being developed and used for each taxonomic group, because different contributors have provided slightly different data fields for each group. Detailed lists of tables and fields are available on the NMITA web site. For example, in zooxanthellate corals, the database used to generate taxon pages consists of ten tables and 80 fields, including four tables within the taxon subject area and three within the locality subject area (Fig. 5). Information for the illustration and specimen subject areas are combined into one table, and two additional tables summarize morphologic data for species and for genera.

**Data entry.**—During the first stage of the project (through January 2001), we plan to enter data for ~1,300 lowest-rank taxa, including ~3,800 images (Table 1). Data are entered by entering the information into a spreadsheet, and using conversion routines to load the data into Oracle. Data for PPP localities are obtained by downloading the current version of the PPP database from the PPP web site, and importing relevant fields into the Oracle database.

**ACCESSING THE DATABASE**

To navigate the database, users select the name of a taxon from dynamically generated lists of taxa that match specified taxonomic, morphologic, or occurrence criteria. These queries consist of PL/SQL programs designed and compiled using the Oracle Schema manager, and code for these programs is posted separately on the NMITA web site. One example using taxonomic data involves a query of zooxanthellate coral species, consisting of two procedures (http://nmita.geology.uiowa.edu/ds.zcoral.html). The first
FIGURE 6—Pollyclave identification key for the zooxanthellate coral family Faviidae. The user selects states for one or more characters from a list of 23 characters (top left panel), and obtains a list of faviid coral species meeting the criteria (bottom left panel). From this list of taxa, the user may select one or more taxa, and obtain images and a complete list of characters for that taxon (right panel).
procedure queries the Oracle database for all coral species and generates a web page with a pull-down list from which the user selects the name of a species. This name is submitted to the second procedure, which queries the database and creates an actual species Hypertext Markup Language or “HTML” page.

Three types of identification keys using morphologic data are available for zooxanthellate corals: a) polychotomous keys using hypertext links, b) online searches of morphologic data in the Oracle database, and c) online searches using morphologic data in DELTA format (Dallwitz, 1980; Dallwitz et al., 1993–1999). All three involve highly-structured morphologic data with links to glossaries containing morphologic character and character state definitions. In the Oracle search, users select one character state for each of a pre-defined set of morphologic characters, and obtain a list of genera having the selected states. Each genus is linked to its corresponding genus page. Online searches of morphologic data in DELTA format (Fig. 6) are implemented using a computer program named “Pollyclave” (http://prod.library.utoronto.ca/pollyclave/). The Pollyclave searches are more flexible than the Oracle searches, and offer numerous additional advantages, including the ability to: a) weight characters and specify character reliabilities; b) handle numeric values (including overlapping sets of numeric values), variable characters, and missing values; and c) list similarities and differences among taxa.

Users may also obtain interactive lists of taxa occurring within a selected geographic region or stratigraphic horizon (Fig. 7). Such lists can be generated by clicking on geographic maps or associated stratigraphic columns. These procedures have been developed for three river sections in the DR project by downloading road, river, railroad and city data contained in ESRI’s

**Locality** = NMB16881 Cana, 352 m, Gurabo Formation, nm12, e.Pliocene

**Locality Group** = C5, Cana5, 330 - 375 m, Rio Cana, fig 15&16, Saunders et al. 1986, 5.6 - 4.5 Ma

**Faunal List:**

- Agaricia lamarcki
- Dichocoenia tuberosa
- Diploria zambesis
- Leptoseris gardineri
- Leptoseris glabra
- Leptoseris sp.A
- Leptoseris sp.B
- Madracis cf. herrickii
- Madracis decactis
- Meandrina braziliensis
- Montastrea-I brevis
- Montastrea-I limbata-2
- Montastrea-I limbata-3
- Montastrea-II canalis
- Musa angulosa
- Musiomi lia aff. hartii
- Pavona sp.A
- Pavona sp.B
- Porites-I macdonaldi
- Porites-I portoricensis
- Porites-II baracoensis
- Scolymia cubensis
- Stephanocenia duncani
- Sylphora monticulosa
- Trachyphyllia bilobata
- Undaria agaricites
- Undaria crassa
- Undaria sp.A

**Figure 7**—Links from localities shown on maps and stratigraphic columns allow users to obtain faunal lists for those localities, which are in turn linked to species pages.


