1 Introduction

A number of studies have emphasized potential links between research experience and recruitment and retention of undergraduate science majors (e.g., Nagda et al., 1998; Lopatto, 2009; Singer et al., 2012). However, implementation of such approaches is often hampered by lack of lab space and curricular materials. While many two-year and four-year colleges lack extensive fossil, mineral, and rock collections, all have access to big data science initiatives, such as the Paleobiology Database (PBDB, paleobiodb.org). Although this database is used extensively by researchers (over 1000 scientists from 31 countries), its incredible potential as a learning tool – for undergraduates in particular – has yet to be fully realized.

Early engagement in authentic research opportunities in the classroom has been linked to increased recruitment, more positive scientific attitudes and values, greater retention and persistence, and increased motivation of undergraduate students in Science, Technology, Engineering, and Mathematics (STEM) disciplines (e.g., Nagda et al., 1998; Lopatto, 2009; Singer et al., 2012). The evidence in support of research-based approaches has compelled the President's Council of Advisors on Science, Technology, Engineering, and Mathematics (PCAST) to advocate for the elimination of standard laboratory courses in favor of discovery-based courses (Olson and Riordan, 2012). As the cyberinfrastructural framework of the geosciences evolves (for example EarthChem, GEOROC, PETDB, NAVDAT; see also EarthCube for an overview of NSF [National Science Foundation]-funded cyberinfrastructure projects in geosciences; earthcube.org), the extent to which these resources are used in the undergraduate classroom lags far behind (Lehnert et al., 2006; Ratajeski, 2006). Research databases may provide an effective research tool that is inexpensive, accessible, and easily disseminated to two- (2YC) and four-year college (4YC) introductory geology courses, including distance learning initiatives (Teaching with Data, Simulations, and Models, 2014).

1.1 What Is the Paleobiology Database?

The PBDB is a large open access database that seeks to catalogue all fossil collections and occurrences, through geologic time, and across the tree of life to enable data-driven projects that ask the big questions in paleontology. PBDB data types consist of taxonomic names (including classifications, synonymies, and re-identifications); geographic information (including county, state, province, country, latitude, longitude, and paleo-coordinates calculated to take into account plate tectonic movement); stratigraphic information (including stratigraphic group, formation, member); bibliographic information; and geologic time scale (Uhen et al., 2013). These data are linked together as occurrences—in which a taxon (e.g., species or genus) is recorded from a particular place (geographic information) and a particular time (stratigraphic information, geologic time scale) by a particular author (bibliographic information). These basic pieces of information are sometimes supplemented by paleoecological (e.g., abundance, life habit, feeding mode), geologic (e.g., lithology), paleoenvironmental, sedimentological, morphological (e.g., measurements of size and shape), sampling (e.g., mode of collection), and preservational (e.g., mode of preservation, taphonomic damage) information. Primary literature sources represent the majority of the data, but legacy databases and unpublished data can be contributed for active research projects. As of May 2018, the PBDB contains data for over 65,000 references, 370,000 taxa, and 1,369,000 occurrences in 193,000 collections. These data were entered by 410 users from 130 institutions in 33 countries.

Scientific output from the PBDB has been truly remarkable. As of May 2018, the list of official PBDB publications numbered 311 (PBDB Publications, 2018) with over 872 publications total using PBDB data in some fashion (Google Scholar, 2018). Citations to these papers exceed 27,000, giving the PDBD a Google Scholar h-index of 74. These data have been used for the original purpose of the database, which was to quantify patterns of paleobiodiversity on Earth (Alroy et al., 2001, 2008; Tennant et al., 2018). Applications have extended well beyond this to include conservation paleobiology (Barnosky et al., 2011; Harnik et al., 2012; Blois et al., 2013; Finnegan et al., 2015), biases in the fossil record or lack thereof (Foote, 2001; Uhen and Pyenson, 2007), extinction (Payne and Finnegan, 2007; Button et al., 2017), taphonomy (Behrensmeyer et al., 2005; Kosnik et al., 2011), taxonomy (Soul and Friedman, 2015), the history of paleontological discovery (Uhen, 2010; Uhen et al., 2013), and more recently, the use of the PBDB as an educational tool (George et al., 2016; Lockwood et al., 2016; Bentley et al., 2017; Lukes et al., 2017; Ryker et al., 2017).

1.2 Big Data's Role in Undergraduate Education

The recognition that teaching the practice of science is more important than memorizing the product of science has been reinforced by policy statements across STEM disciplines. The Discipline-Based Education Report (DBER) report from the National Research Council (Singer et al., 2012) clearly states that research-based instructional strategies are more effective in developing conceptual knowledge and improving attitudes about learning than traditional lecture formats. In the geosciences, the critical importance of field and research-based experiences to geoscience major recruitment and retention has long been recognized (Houlton, 2010), but has proven difficult to implement, due to the expense and inaccessibility of research tools. These problems are particularly acute at two-year institutions, which may lack the research infrastructure and facilities to promote "doing" in entry-level or distance-learning courses.

The PBDB, like other research databases, can provide an accessible and inexpensive research tool applicable to entry-level courses at both two- and four-year colleges. Countless STEM databases exist, and visualizations from a limited number of these databases have been leveraged as conceptual demonstrations or parts of inquiry labs (see for example Science on a Sphere by National Oceanic and Atmospheric Administration [NOAA]). How the use of these databases has increased student access to research experiences has received remarkably little attention in geoscience or other STEM disciplines (Singer et al., 2012). This is despite the fact that, during the 2014 Future of Undergraduate Geoscience Education Summit, the geoscience community identified student skills working with big data and datasets as a critical, often missing, component of an undergraduate geoscience education (Mosher et al., 2014). Similarly, the DBER report from National Research Council (Singer et al., 2012) emphasized that inquiry-based student activities using online datasets have been found to improve student learning (e.g., Rissing and Cogan, 2009), increase student competency with science practices (e.g., Brickman et al., 2009), and increase student confidence in their ability to do science (e.g., Seymour et al., 2004). As far back as 2003, geoscience educators explicitly called for the development of database-specific research skills (Manduca and Mogk, 2003), but progress has been limited. As geoscientists transition from traditional lab to authentic research courses, the need for easily accessible sources of big data will only increase.

There are numerous benefits to teaching earth science in the context of cyberinfrastructure, including:

1. Facilitation of effective teaching pedagogies and student skills such as inquiry and evidence-based decision making and analysis (Teaching with Cyberinfrastructure, 2014).
2. Production of customized datasets relating to geologic time, plate tectonics, climate change, mass extinction, and geospatial data.
3. Highlighting the scale and diversity of materials and processes in the Earth system (Teaching with Cyberinfrastructure, 2014).

4. Development of quantitative skills including data transformation, univariate and multivariate statistics, calculation of rates, graph building and interpretation.
5. Introduction to research skills including finding and compiling primary literature, critical thinking, developing hypotheses, working independently, problem solving, data manipulation, etc. (Kardash, 2000; Ishiyama, 2002).
6. Building of technological familiarity literacy for students and instructors.
7. Socialization of students into scientific culture (Hunter et al., 2006).
8. Appreciation of the community and team-based nature of modern science (Understanding Science, 2014). Pedagogical research emphasizes the utility of group learning, but many undergraduates view this approach with skepticism, having few examples of collaborative science to build on.
9. Facilitation of communication and collaboration between two- and four-year college geology departments without significant travel between them.

2 How to Use the Paleobiology Database

2.1 How to Access the PBDB

The PBDB is organized as a relational database with multiple web forms for data entry, visualization, and retrieval. In addition, an Application Programming Interface (API) has been developed, which simplifies data retrieval tasks and makes it much easier for scientific and/or educational users to develop their own web or mobile-based applications (apps) to search, download, visualize, and analyze data. A data input API is also under development that will allow the creation of customized data input applications as well. A number of PBDB apps currently exist and are accessible from the PBDB Resources webpage (paleobiodb.org/#/resources). The two most commonly used applications for searching, visualizing, and analyzing PBDB data are PBDB Navigator and Fossilworks. PBDB Navigator (paleobiodb.org/navigator/) was developed by Shanan Peters, Michael McClennen, and John Czaplewski and launched in 2013. Navigator allows users to explore when and where fossils occur, in addition to mapping their geographic distribution, graphing their diversity through time, and examining their taxonomy. Information on how to use PBDB Navigator is provided below. Fossilworks (fossilworks.org/), which was launched by John Alroy in 2013 based on the original Paleobiology Database platform, allows users to search all aspects of the database and map occurrences. Downloading PBDB data, for users to analyze on their own, can be accomplished using the PBDB Application

Programming interface (API; paleobiodb.org/data1.2/), PBDB website (paleo biodb.org/classic/displayDownloadGenerator), or Fossilworks.

The PBDB seeks to catalogue all fossil collections and occurrences – through geologic time, around the globe, and across the tree of life. The primary record for most fossils in the PBDB is the occurrence, which is defined as where an individual sample or set of samples of a fossil species occurs in space (geographically) and time (stratigraphically). For example, 10 specimens of a species of trilobite found in a rock unit dated to the late Cambrian would be an occurrence of that taxon. A PBDB collection is a set of occurrences from the same site (i.e., locality). For example, a collection might be a group of eight different fossil taxa occurrences (which themselves could constitute one or more actual specimens of the same taxon) found in one outcrop. One of the easiest ways to access the PBDB is using a web application called PBDB Navigator (paleobiodb.org/navigator/).

2.2 Delving into PBDB Navigator

PBDB Navigator consists of five parts (Fig. 1):

1. Map (center of screen, Fig. 1) showing continents with dots representing fossil collections. The color of these dots represents their geologic age. If you zoom in and click on the dots, you can see the information on each site and the taxa that occur there. The map changes as you zoom in or out, search for specific taxonomic names, and/or select particular time intervals (see below).
2. Geologic Time Scale (bottom of screen, Fig. 1) showing the major eras, periods, and stages. If you click on the geologic time scale, the map will

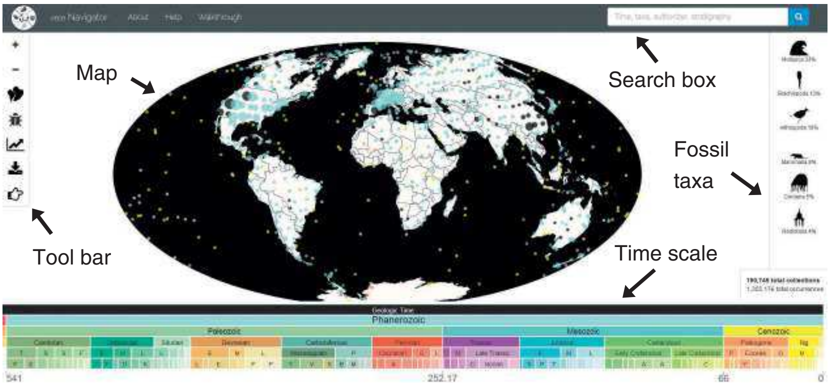

Figure 1. PBDB Navigator "landing page" labeling the map, toolbar, geologic time scale, search box, and fossil taxa bar

show you the location of all fossil collections from that time interval. For broader time resolution, you can click on the upper rows of the geologic time scale, including eon (Phanerozoic) or era (e.g., Paleozoic, Mesozoic, Cenozoic). For more specific time resolution, you can select the lower rows of the geologic time scale, including periods (e.g., Jurassic, Cretaceous) and stages (e.g., Aptian, Albian). Double clicking on a time period will zoom the time scale to that particular time period.

3. Tool Bar (left side of screen, Fig. 1) showing the tools you can use to explore the database (Table 1). If you position the mouse cursor over each tool, the tool name will pop up. The *Zoom in* and *Zoom out* tools allow you to zoom in and zoom out of the map, so that you can focus on a particular geographic area. The *Toggle Paleogeography* button allows you to switch between modern and paleo tectonic plate configurations, but note that you must select a period or lower time interval first. The *Toggle Taxa* button allows you to open a taxonomic browser, which you can use to search for particular taxonomic names at any taxonomic level. As you type in this browser, the taxonomic name will be auto-filled with the name of organisms contained in the PBDB. The *Toggle Stats* button makes it possible to quickly and easily create diversity, extinction, and origination curves. It is important to note that the data for calculating diversity, extinction, and origination

Table 1 Toolbar for PBDB Navigator

Symbol	Description
	*Zoom in/out* on the map
	*Toggle Paleogeography*: Reconstructs plate tectonic configurations for time interval (era or smaller) you are exploring
	*Toggle Taxa* Browser: Narrow down which taxonomic group is plotted on map
	*Toggle Stats*: Create a diversity curve for the collections currently plotted on map
	*Save Map Data*: Download the data (lat/long, geologic age, etc.) for the occurrences plotted on map
	*See Examples*: Start using PBDB Navigator with three examples

curves is determined by the geographic region that is currently displayed on the map. Finally, the *Save Map Data* button allows you to download the data for the occurrences plotted on the map.

4. Search Box (upper right corner of screen, Fig. 1) allowing you to search for particular time intervals, taxa, or stratigraphic unit. As you type in the search box, the words will be auto-filled with the time intervals, taxonomic names, and stratigraphic units contained in the PBDB.
5. Fossil Taxa Bar (right side of screen, Fig. 1) showing the faunal composition (in percent) of the collections currently represented on the map, plus the total number of collections and occurrences. This information changes as you zoom in or out, search for specific taxonomic names, and/or select particular time intervals.

A video tutorial (five minutes long) on how to use PBDB Navigator is available on the PBDB Resources page (paleobiodb.org/#/resources), along with other useful tutorials.

3 Educational Integration: Examples and Ideas

Here we present a handful of data-focused lessons for undergraduate geoscience courses using the PBDB. These lessons are a small subset of the lessons and tutorials developed as part of an NSF-funded study on the use of research databases in undergraduate education (NSF IUSE grant DUE-1504588 and 1504718). These resources are available on both the PBDB Resources webpage (paleobiodb.org/#/resources) and the SERC (Science Education Resource Center, serc.carleton.edu) educational resource platform (search term PBDB). Each lesson has undergone peer-review and been piloted at a variety of two- and four-year institutions, including George Mason University (Virginia), William and Mary (Virginia), Eastern Michigan University (Michigan), High Point University (North Carolina), Northern Virginia Community College (Virginia), and Thomas Nelson Community College (Virginia).

These lessons have been designed to: 1) be modular so that they range in duration from five minutes to three hours and 2) be flexible for use in lecture, laboratory, and field settings, or as homework. The lessons feature essential skills for scientifically literate citizens (including critical thinking, hypothesis generation, and data analysis; see Next Generation Science Standards) applied to a range of societally important topics, including extinction, evolution, and climate change.

A note on access and accessibility: PBDB Navigator requires the use of a laptop or desktop computer and Internet access. While many institutions are able to provide access to computing facilities, some are not, and emphasizing economic disparities among students and/or institutions should be avoided.

3.1 Integration into Introductory-Level Classes

The ease with which undergraduate students can explore PBDB Navigator makes it ideal for use in introductory courses, including physical geology, physical geography, and historical geology. It is a quick, inexpensive, and easy way to introduce scientific data into undergraduate teaching, which does not require substantial lab or research facilities. This makes it particularly effective for use by two-year colleges and distance learning initiatives.

As a quick five-minute activity during lecture or lab, instructors could ask students to look up which fossils or fossil sites occur in their local area (e.g., university location, hometown, etc.). As a longer-term lab or homework activity (20–40 minutes), students could determine which fossils occur through time in a particular region (state, province), then hypothesize how paleoenvironment has changed through time based on these fossils. This is a particularly effective exercise if the instructor chooses a region that has experienced substantial changes in environment (e.g., from marine to terrestrial) through time.

3.1.1 Pangea Puzzle

One activity, which allows instructors to emphasize the general importance of paleontology at the introductory level, focuses on the distribution of fossils across the supercontinent of Pangea. The Pangea Puzzle activity is accessible via the PBDB Resource Page or the SERC Resource Platform (serc.carleton.edu/NAGTWorkshops/intro/activities/177872.html), and includes learning objectives, classroom handouts, teaching tips, and suggestions for assessment. This activity can be used in an introductory or intermediate undergraduate course, including (but not limited to) physical geology, historical geology, and paleontology. Students can work as individuals or in pairs, and class size can range from a small seminar (< 10 students) to a large lecture (> 100), as long as sufficient computer facilities are available. It has been successfully piloted with non-majors and majors from both two- and four-year institutions. Each student or student pair will need access to a laptop or desktop computer connected to the Internet, running an Internet browser. The activity can be used in its entirety as a lab or homework activity (1–1.5 hours) or pared down to be used as a lecture activity (20–30 minutes). It is most effectively presented as a supplement to coursework on Pangea, continental drift, and plate tectonics.

Cambridge University Press
978-1-108-71790-8 — Utilizing the Paleobiology Database to Provide Educational Opportunities for Undergraduates
Rowan Lockwood , Phoebe A. Cohen , Mark D. Uhen , Katherine Ryker
Excerpt
More Information

In this activity, students learn how to use the PBDB to map the geographic distribution of several fossil taxa, using both modern and past plate configurations. They then use these maps to explore how the geographic distribution of fossils provides evidence for the supercontinent Pangea and plate tectonic movement. Concept goals for this activity include: 1) constructing a map of fossil occurrences on the present-day Earth's surface, 2) constructing a map of fossil occurrences on the Earth's surface at various times in Earth's past, 3) identifying the past distributions of fossils on ancient continents and supercontinents, and 4) explaining how the present-day distributions of fossil organisms are different from their distribution during the time of their deposition as fossils. A higher-order learning objective involves developing hypotheses regarding why the present-day distribution of fossil occurrences is dramatically different from their distribution in the past.

To begin the activity, students are asked to map the geographic distribution of a Permian fossil taxon, such as *Glossopteris* (seed fern), on a map with modern tectonic plate configurations. *Glossopteris* works particularly well for this activity, but students can use several taxa, including the original examples used by Alfred Wegener to support his ideas on continental drift, including *Lystrosaurus* (dicynodont therapsid, transitional fossil between reptiles and mammals) and *Mesosaurus* (aquatic reptile). Students can then be asked to figure out what each organism is (using the *Toggle Taxa* button on the left tool bar), when their fossils occur (using the geologic time scale at the bottom of the screen), and where they occur (using the map in the center). If students limit the distribution of mapped occurrences to the Permian (by clicking on the Permian interval on the geologic time scale at the bottom of the screen), they'll see that, for example, the distribution of *Glossopteris* is limited to the southern continents—South America, Africa, Australia, and southern Asia (Fig. 2A). The instructor can then reveal that these taxa cannot travel long distances over water, and ask students to explain why this is a problem, given the modern distribution of these taxa.

Next, the students can click on the *Toggle Paleogeography* button (on the left tool bar) to map the distribution of Permian *Glossopteris* on a map with Permian plate configurations (Fig. 2B). Students can then be asked to explain how this distribution provides evidence of the supercontinent Pangea and plate tectonics. Follow-up questions can focus on: 1) what other types of taxa could be used to reconstruct plate positions through time, 2) what direction each continent is moving through time, and 3) what the estimated rate of spreading is for the Atlantic Ocean (if instructor provides additional information regarding the current size of the Atlantic).

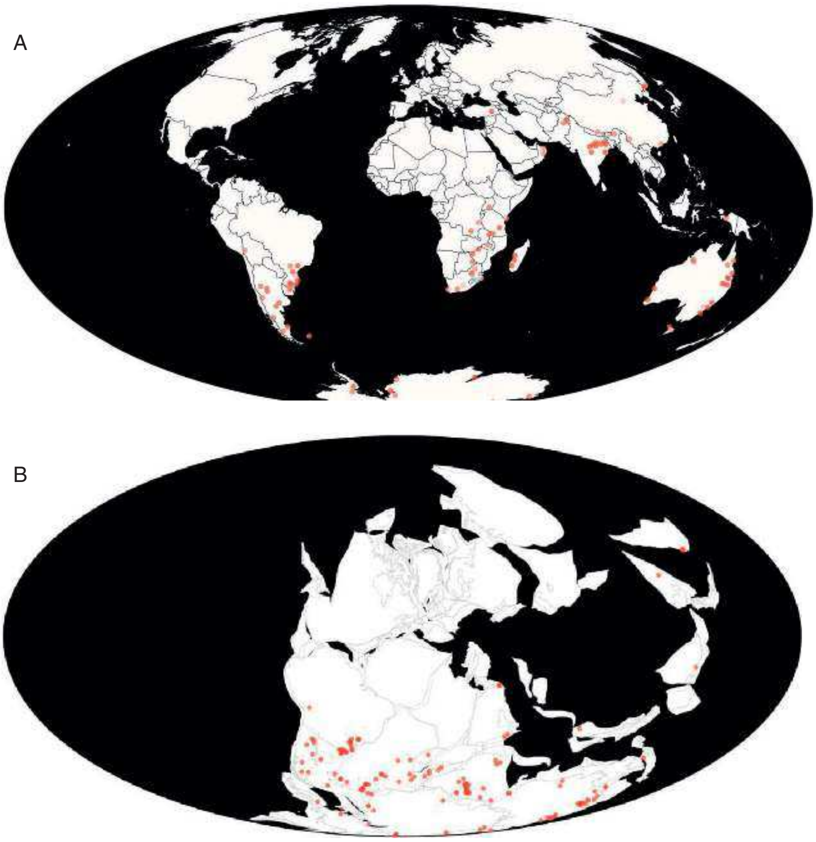

Figure 2. Example of output from Puzzling Pangaea activity using PBDB Navigator. (A) Students map the distribution of fossil occurrences of *Glossopteris* ferns during the Permian, superimposed on modern continental configuration. (B) After attempting to explain this distribution, students then map the same data, overlain on continental configuration for the continent of Pangaea during the Permian.

The web link for this activity (serc.carleton.edu/NAGTWorkshops/intro/activities/177872.html) provides a wealth of information including approaches for student assessment, background reading, and pre-requisite skills.

3.2 Integration into Upper-Level Classes

The wealth of fossil data contained in the PBDB makes it very useful for upper-level courses, including paleontology, sedimentology, and stratigraphy. Quick activities (10 minutes or less) could include mapping the distribution of a particular fossil group across both space and/or geologic time; looking up