DATA PROCESSING METHODOLOGIES TO INVESTIGATE THE ASSOCIATION BETWEEN DEPOSITIONAL ENVIRONMENTS AND TRACE FOSSIL OCCURRENCE

A Thesis Submitted to the College of

Graduate and Postdoctoral Studies

In Partial Fulfillment of the Requirements

For the Degree of Master of Science

In the Department of Geological Sciences

University of Saskatchewan

Saskatoon, Saskatchewan

By

Dean M. Meek

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other uses of materials in this thesis in whole or part should be addressed to:

Head of the Department of Geological Sciences
University of Saskatchewan
114 Science Place
Saskatoon, Saskatchewan S7N 5E2
Canada

OR

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9
Canada

Abstract

The transition from late Ediacaran to early Cambrian records important paleobiological and paleoecological changes. These are observed in the Fortunian diversification event and the Agronomic Revolution, which describe significant body plan diversification, increased behavioral complexity in trace fossils, and a shift from matgrounds to mixgrounds ecosystems. To provide a more thorough understanding of this dramatic transition, data mining techniques (i.e. visual and statistical analysis) are used to investigate the relationship between depositional environments and trace fossil occurrence. To facilitate analysis, an ichnological database has been designed and implemented using Microsoft Access. The creation of this database is important in that it provides a platform for data digitization and subsequent data mining, while also accounting for fundamental differences between trace fossils and body fossils. Current paleontology databases do not recognize this distinction, which stems from the fact that trace fossils represent organism behavior, while body fossils record the phylogenetic affinities of an organism. Analysis of the ichnologic data compiled is supported with additional datasets, with a large focus on utilizing detrital zircon to infer geodynamic settings and to provide validation of paleogeographic reconstruction models via visual provenance analysis. A more quantified version of detrital zircon provenance analysis by way of Multidimensional Scaling (MDS) was conducted; however, this study has shown that MDS is best utilized at a regional scale. In combining all supplementary datasets, paleogeographic reconstructions for the Ediacaran, Terreneuvian, and Cambrian Epoch 2 have been constructed. With an appropriate spatial and temporal context, visual analysis of ichnologic data displays a global distribution of trace fossils through this transition, implying the utilization of available ecospace and a lack of paleoclimatic restrictions. Statistical analysis in the form of Correspondence Analysis (CA) displays a clear lack of relationships between ichnogenera and depositional environments during the Ediacaran, suggesting trace fossils were facies-crossing prior to Phanerozoic-style ecosystems. CA produces markedly different results during the early Cambrian, displaying ichnogenera differentiation between depositional environments (i.e. increasing beta ichnodiversity) in the relationship between *Oldhamia* and deep marine depositional environments. These results lend support to the Agronomic Revolution, as microbial matgrounds were forced into increasingly stressful paleoenvironments (i.e., deep marine settings) during this paleoecological revolution.

Acknowledgements

I would like to express my gratitude towards both of my supervisors for their guidance and support throughout my time as a graduate student. Thank you Dr. Bruce Eglington for sharing your seemingly endless knowledge regarding statistics, databases, and geology. Your encouragement was essential in providing motivation and determination to complete a graduate degree. Thank you Dr. Luis Buatois for sharing your knowledge and affection for ichnology, sedimentology, and sequence stratigraphy. The creation of an ichnology database would not have been possible without your support.

I would also like to thank Dr. Gabriela Mángano, your enthusiasm and work ethic was, and continues to be a source of inspiration. To the ichnology group, I thank you for many interesting and enlightening discussions during our seminars and social gatherings. Your companionship, moral support, and encouragement was invaluable. I would specifically like to thank Brittany Laing for sharing valuable ichnology related insights and the testing of IchnoDB, in which your feedback was essential.

To Michael McConnell, for frequently providing a much-needed respite from the stresses of research and writing at Leopolds, or wherever we could find a table. To Lavie Nyguen, thank you for being continuously optimistic and providing endless moral support. To Susan James, for providing suggestions and insights, resulting in improvements and clarity.

This project was supported financially from the Department of Geological Sciences in the form of a Graduate Teaching Fellowship, and from the Murray Pyke endowment and Chair at the University of Saskatchewan.

Dedication

To my family and friends,

without your support this would not have been possible.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BI..... Bioturbation Index

BPBI..... Bedding-Plane Bioturbation Index

CA..... Correspondence Analysis

CRT..... Crustal Residence Time

DBMS .....Database Management Systems

DMTS .....Deep Marine Turbidite System

EDA .....Exploratory Data Analysis

GBDB .....Geobiodiversity Database

GIS ..... Geographic Information Systems

GPS .....Global Positioning System

GSSP..... Global Boundary Stratotype Section and Point

KDD..... Knowledge Discovery from Data

KDE .....Kernel Density Estimate

LA-ICPMS..... Laser Ablation – Inductively Coupled Plasma Mass Spectrometer

MDS.....Multidimensional Scaling

MSWD..... Mean Sum of the Weighted Deviates

MVA .....Multivariate Analysis

NAMG ..... North American Magmatic Gap

nMDS..... non-metric Multidimensional Scaling

PBDB ..... Paleobiology Database

PDP ..... Probability Density Plot

REE ..... Rare Earth Element

SIMS ..... Secondary Ion Mass Spectrometer

SQL ..... Structured Query Language

T_{2DM} ..... Two-stage Depleted Mantle model age

T_{CRes} ..... Crustal Residence model age

UPGMA ..... Unweighted Pair-Group Method using arithmetic Averages

1. Introduction

Exploratory data analysis (EDA) is the investigation of a dataset in an attempt to find connections and peculiarities between pieces of information (Tukey, 1977; Morgenthaler, 2009). The EDA process is not limited in the techniques employed to investigate the data, assuming the methods selected are logically sound and statistically valid, rather it is limited by the creativity of the analyst (Morgenthaler, 2009). With the increasing ease in which data can be collected and used, EDA can utilize analytical techniques that were previously prohibitive as a result of dataset size.

EDA is a broad term, under which many analytical processes can be grouped, including data mining. Data mining is recognized as a process, whereby large volumes of data are gathered and various methods are applied to the datasets in an attempt to extract meaning from patterns or trends exposed (Han et al., 2012). As the purpose of data mining is ultimately to gain a better understanding of that data being investigated, it is often described as knowledge discovery from data. Both EDA and data mining processes typically employ visualization, statistics, and machine learning techniques in the analysis of data (Han et al., 2012).

Over the past decade there has been an increase in the volume of paleontology data captured and stored within open accessible databases (e.g., the Paleobiology Database and the Geobiodiversity Database). Access to large datasets presents many more opportunities to exploit data mining techniques within paleontology. Unfortunately, ichnology has been largely bypassed in the paleontology databases mentioned. Proposals for the collection and storage of both published and unpublished ichnology data have been presented (e.g., Goldstein et al., 2010; Belvedere et al., 2011); however, these efforts have not produced an open access database. Despite the lack of a database, several authors have taken the time to create their own compilation of ichnology data for analysis (e.g., Orr, 2001; Uchman, 2004; Jensen et al., 2013; Mángano & Buatois, 2014; Buatois & Mángano, 2016, 2018; Buatois et al., 2016). These individual compilations are undoubtedly useful, but they are not always provided in the supplementary information. Further, these datasets are typically stored in a format that does not support sorting, restriction, extraction, or alteration of the data.

The absence of a distinct ichnology database is a problem if the solution is to be the digitization of ichnology data within a body fossil database because these two types of fossils have a contrasting nature, with trace fossils being representative of organism behavior, whereas body fossils reflect phylogenies (Pemberton & Frey, 1982). As a result of these fundamental differences, this thesis argues against the inclusion of trace fossil data into paleontology databases which are designed specifically for body fossils. Although metadata associated with body fossils and trace fossils is broadly the same, important differences are accounted for with comparable concepts (e.g., guild and ichnoguild). With the identification of this issue, it is recommended that paleontology databases include a new field to distinguish between body fossils and trace fossils for individual records. This would emphasize the unique dissimilarities and provide a distinction between similar but distinct concepts recorded as metadata. Further, the inclusion of an additional field is a basic alteration, producing no foreseeable negative impacts on existing database design.

The creation of an open access ichnology database or the alteration of paleontology databases described above would provide a solid foundation upon which data mining techniques could be used. This would provide many benefits to the wider geology and paleontology communities. The database structure supports rapid manipulation and extraction of data (Harrington, 2016), which in turn reduces the time prohibitive nature of data collection during the compilation of large datasets. Such a database would support different facets of ichnological studies (e.g., macroevolutionary, paleogeographic, paleoecology). Further, a database could influence standards for data presentation and nomenclature used within publications. Standards in data presentation would provide a consistent set of metadata, with a wide range of information applicable to many areas of research. Finally, consistent nomenclature supports the comparison of published results which may not have originally used similar terms.

Having emphasized the advantages of a database, the structure of an ichnology database is outlined in this thesis. The process of creating this database structure followed that described by Mannila & Räihä (1992), including requirement analysis, conceptual design, logical design, and physical design. The result of these efforts is an ichnology database containing published data from around the globe, specifically focused on trace fossils of Ediacaran to early Cambrian age. With this compiled data, both visual and statistical data mining techniques are employed to

look for any type of relationship that may exist between trace fossils and their depositional environments.

To visually analyze paleontological datasets for patterns or trends, GIS software can provide spatial and temporal context (Markwick & Lupia, 2002). This method enables visualization of data at their current spatial location and does not provide any context regarding the distribution of continental plates at the time of deposition. To address this issue, plate-tectonic reconstruction software (i.e. GPlates) was used for visual analysis of the ichnology data. Integrating the spatial and temporal aspects of the data with a plate-tectonic reconstruction model provides a paleogeographic context appropriate for the time period of interest. The paleogeographic context provides valuable information for the interpretation of processes controlling the distribution of trace fossils in both time and space. As a result, analysis of latitudinal and depositional environment distribution can be conducted.

Although the ichnology data compiled and the subsequent metadata indicating depositional environments can be informative when plotted on its own, it is likely that supplementary datasets can provide additional details. It has been observed that the preservation of fossils in a temporal and spatial context is non-uniform (Holland, 2016). This is explained as a result of the stratigraphic record, in both its extent and distribution. Therefore, additional information regarding basin locality and type can provide indications towards regions where fossils are more, or less likely to occur.

Detrital zircon data can provide multiple benefits in association with the visual analysis of paleontological patterns, through the inference of geodynamic setting and evolution, as well as the refinement of paleogeography. With the recognition that stratigraphy has an effect on fossil preservation, the ability to infer a tectonic setting from detrital zircon samples (Cawood et al., 2012) will aid visual analysis of paleontological distributions. Additionally, the use of Multidimensional Scaling (MDS) with detrital zircon analysis can highlight the evolution of geodynamic settings through time (e.g., Spencer & Kirkland, 2016), providing detail to the constantly evolving distribution of continental plates. The final benefit derived from detrital zircon provenance analysis is the potential for improvements to paleogeographic reconstructions used for visualizing datasets of any kind through Earth history.

In addition to detrital zircon data, other datasets have been compiled to create supplementary paleogeographic reconstructions for visual analysis. Included in these additional datasets were multiple geological maps, and data points in the form of geochronology, thermochronology, paleontology, paleoclimate, and paleocurrent data. This data provided the basis for the interpretation of four distinct paleogeographic settings including deep marine, shallow marine, terrestrial, and orogenic. Three paleogeographic reconstructions have been created as part of this thesis to support visual analysis, beginning with the Ediacaran and extending through to Cambrian Epoch 2.

The final data mining technique used for the analysis of trace fossil occurrence and depositional environments is correspondence analysis (CA). This statistical technique is commonly employed with data mining and provides a more quantified approach when compared with the visual techniques. In the analysis between ichnology and depositional environments, CA provides an advantage in that it measures the dependence between variables (e.g., trace fossils), while also comparing objects (e.g., depositional environments) simultaneously (Davis, 2002; Borcard et al., 2011). The results of this technique are then open to interpretation using scientific phenomena to explain the significance of observed patterns or trends.

The data mining techniques (e.g., visual and statistical analysis) employed throughout this thesis will aid in the analysis of possible relationships between trace fossils and depositional environments. It is the intention of this thesis to provide a more thorough understanding of trace fossil geographic distribution (i.e. gamma ichnodiversity), and the adaption to paleoecological conditions (i.e. beta ichnodiversity) present during the transition from the Ediacaran to Cambrian Epoch 2. Further, in conducting these different analyses, the techniques used will be shown to be advantageous to paleontology and the utility of an ichnology database will be clear.

1.1 A Review of Databases

Databases are collections of data that are stored within an organized structure which allows for the retrieval of subsets of information (Batini et al., 1992; Mannila & Räihä, 1992; Uhen et al., 2013; Harrington, 2016). Relational databases were originally proposed by E.F. Codd in the 1970's (Codd, 1970, 1979), where the structure of the data and how data manipulation could take place was logically described. The fundamentals for relational

databases as described by Codd have been carried forward to the present day, where software packages known as database management systems (DBMS) conduct the job of storing, manipulating, and retrieving data (Batini et al., 1992; Mannila & Räihä, 1992; Harrington, 2016).

There are many different DBMS platforms available, from those that are designed for a single user, to those that are intended for simultaneous input by many users (Harrington, 2016). All DBMS software provides a general set of functions, including: functions for creating the structure of the database; a way to enter, modify, and delete data; a way to retrieve data; and methods for restricting access to data (Mannila & Räihä, 1992; Harrington, 2016). Creating the structure of the database through relationships will be the backbone of the entire database, later ensuring the proper retrieval of information, as well as consistent data entry as defined by the constraints specified upon initial design (Mannila & Räihä, 1992).

The ability to retrieve, enter, delete, and modify data in a database is done through the use of SQL, the international standard for such processes (Batini et al., 1992; Mannila & Räihä, 1992; Harrington, 2016). The seminal publication describing SQL was written by Chamberlin & Boyce (1974), in which they referred to it as the Structured English Query Language (SEQUEL). Their intent was to create a sublanguage that would be equally applicable to computer scientists as it would to other professionals due to the growing influence of databases. The process of defining queries using SQL may be simplified even further for users through the use of forms. These forms can be present as a standard feature within a DBMS (e.g., Microsoft Access), or they have been constructed by the creators of the database to enter and withdraw information without knowledge of SQL (e.g., DateView).

1.2 A Review of Database Use and Creation in Paleontology

One of the ways databases serve the scientific community is through data preservation, by acting as data repositories (Uhen et al., 2013). These authors observed that even during the "Information Age," data behind many paleontological publications is not readily available, rather it is listed as available upon request. This creates the potential for data loss if the authors leave academia, retire, or die. Using established databases to store supplementary data not only negates this risk, it may be required by public funding bodies (Uhen et al., 2013). Data management plans are currently required by the U.S. NSF (Grant Proposal Guide, Chapter

II.C.2.j) and are expected from the Canadian NSERC (Tri-Agency Statement of Principles on Digital Data Management), as both organizations strive for preservation and access to data generated or collected.

Geological databases are often created to store an assortment of information such that a 'core' table is produced with additional supplementary and 'lookup' tables (Eglington, 2004; Carrasco et al., 2007). The core table is described as the principle data component of the database, the remaining data is meant to support and supplement this component. Often these core tables consist of site localities, with related tables supplementing these geographic localities with details regarding geochronology, geochemistry, paleontology, or sedimentology. Beyond their function to support a core table, lookup tables also serve to establish an acceptable nomenclature and prevent typographical errors (Eglington, 2004). Additional details to be considered within both core and lookup tables is the assignment of a primary key, which is the definition of one or more columns that contain a value that uniquely describes each row (Harrington, 2016). With a unique primary key duplicate records cannot be entered into the database (Eglington, 2004; Harrington, 2016).

Compiling paleontology data is often troublesome, as publications span many decades and the nature of the fossil record is geographically widespread, and contained within unique stratigraphic units (Carrasco et al., 2007). This time consuming and difficult task has been addressed in part, through the creation of several different vertebrate and invertebrate fossil databases. The Paleobiology Database (PBDB, https://paleobiodb.org/) and the Geobiodiversity Database (GBDB, http://www.geobiodiversity.com/) serve as excellent examples. Unfortunately, to this point ichnology data has been largely bypassed in the entry of data in these paleontology databases. Suggestions for the creation of a separate database specifically created and designed for ichnology data have been proposed (e.g., Goldstein et al., 2010; Belvedere et al., 2011); however, to this point a lack of funding has been the primary obstacle of these proposals (Belvedere, personal comm. 2018; Goldstein, personal comm. 2018).

As previously stated, many geological and paleontological databases store information with its related spatial location. With this data, geologists are able to take advantage of geographic information systems (GIS) software to conduct spatial analysis (Carrasco et al., 2007). Implementation of this described process has been used extensively in

paleobiogeographical studies over the past decade (e.g., Stigall & Lieberman, 2006; Hendricks et al., 2008; Stigall et al., 2014; Pohl et al., 2016). It should be observed that using databases is not a requirement for analysis of geological datasets with GIS software (e.g., Rode & Lieberman, 2004; Boucot et al., 2013); however, this process is undoubtedly more time consuming than using databases to extract datasets. Examples of research emerging from the use of databases includes marine taxonomic patterns through the Phanerozoic (Peters, 2005; Alroy, 2008), Paleozoic ocean circulation patterns (Pohl et al., 2016), Ordovician speciation patterns (Lam et al., 2018), and the revision of paleogeographic coastlines using paleontology (Cao et al., 2017). Furthermore, the integration of multiple datasets may result in new discoveries (Uhen et al., 2013), this process is better known as exploratory data analysis (EDA) and is discussed further in section 1.5.

With access to GIS software, relational databases are the most practical method for analyzing paleontological patterns and the processes responsible for their creation (Markwick & Lupia, 2002). Designing a relational database is a complex process, displayed in Figure 1.1 with the visualization of the four major steps (Batini et al., 1992; Mannila & Räihä, 1992). Requirement analysis is the first step described by Mannila & Räihä (1992), where the overall purpose of the database is defined along with what data will need to be collected and stored to achieve this purpose. The second step in this process is the formation of a conceptual design, which produces a detailed description of the database and the information that will be stored. Logical design is the third step, which outlines the structures that will be used to store the data. Data normalization is often identified with logical design, a process in which tables are constructed to eliminate redundant data storage (Mannila & Räihä, 1992). The final step of database creation is physical design, which is the implementation of the logical design into a specific DBMS (Batini et al., 1992; Mannila & Räihä, 1992). Mannila & Räihä (1992) suggested that this methodology for database creation is most effective when feedback and review of the designs are done at each successive step to identify potential issues at any stage of the process. Markwick & Lupia (2002) provided a general guide, specifically for the creation of a paleontology database that can be incorporated during the database design process. This included a basic logical design, the incorporation of audit trails for both sources and alterations to the original data, and the collection of data at the finest resolution possible.

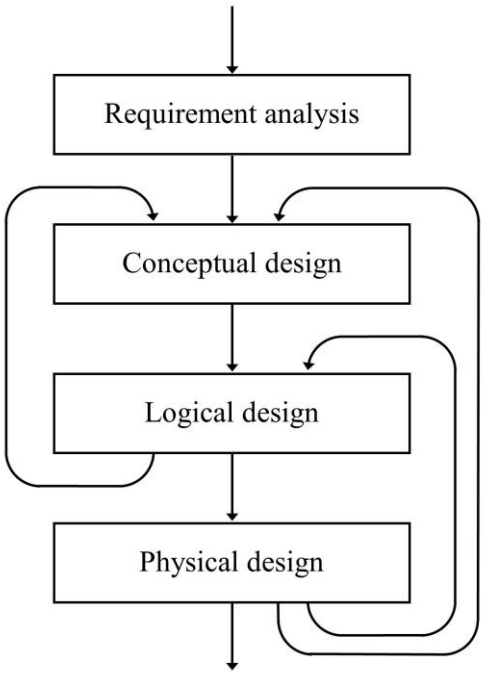


Figure 1.1 The work flow of database design, displaying the four major steps with repeated design reviews (after Mannila & Räihä, 1992)

1.3 A Review of Ichnology

Ichnology is the study of trace fossils produced by both animals and plants, on or within a substrate, including bioturbation, bioerosion, and biodeposition (Bromley, 1990, 1996; Pemberton et al., 1992a; Buatois & Mángano, 2011; Minter et al., 2016a). Ichnology has experienced explosive development since the seminal publications of Dolf Seilacher more than half a century ago (e.g., Seilacher, 1955a, 1955b, 1964, 1967a). Ichnology is a valuable asset to paleontology, sedimentology, and stratigraphy as a result of the unique characteristics of trace fossils, namely a narrow facies range, limited secondary displacement, a wide temporal range, production by non-preservable soft bodied biota, and occurrence in rocks lacking body fossils (Frey, 1975; Pemberton et al., 1992a).

One of the unique aspects of ichnology is that one organism may produce more than one ichnotaxa, and one ichnotaxa may be produced by more than one type of organism (Frey & Seilacher, 1980; Ekdale et al., 1984; Buatois & Mángano, 2011). This fact precludes the use of

the taxonomic scheme used by biologists which bases classification on phylogeny, rather ichnotaxonomy bases classification on behavior (Bromley, 1990, 1996). Trace fossil classification is not restricted to ichnotaxonomy, classification via ethology or architectural design is also possible. Ethological classification is based on fundamental behavioral patterns, controlled by evolutionary adaptations and environmental conditions present (Pemberton et al., 1992b). Ethological categories have been modified several times since Seilacher (1953, 1964) proposed this type of classification, with a current list of 14 categories in Figure 1.2 (Buatois & Mángano, 2011; see also Vallon et al., 2016). The other trace fossil classification method mentioned is architectural design, with a comprehensive list of 79 categories provided by Buatois et al. (2017). These authors make the distinction that categories of architectural design are both practically and conceptually not ichnotaxonomic categories, although this type of classification relies on accurate use of ichnotaxobases as defined by Bromley (1996). Through the use of ichnotaxobases these categories try to capture easily recognizable morphological changes in trace fossils.

A well-known and important characteristic of ichnology is that trace fossils display a narrow facies range, a characteristic caused by the organism's response to paleoecological factors (Frey, 1975; Pemberton et al., 1992b). Some of the paleoecological factors effecting trace fossil distribution and morphology include bathymetry, salinity, substrate coherence and stability, sedimentation rate, turbidity, temperature, hydrodynamic energy, oxygen content and food supply (Frey & Seilacher, 1980; Pemberton, et al., 1992a; Brett, 1998; Buatois & Mángano, 2011). Buatois & Mángano (2011) noted that the tolerance of benthic organisms to any of the factors listed is typically defined by a combination rather than individual isolated factors. These authors suggested that ichnological studies should focus on the paleoecological indications that trace fossils provide as part of a multidisciplinary approach to depositional environment determination.

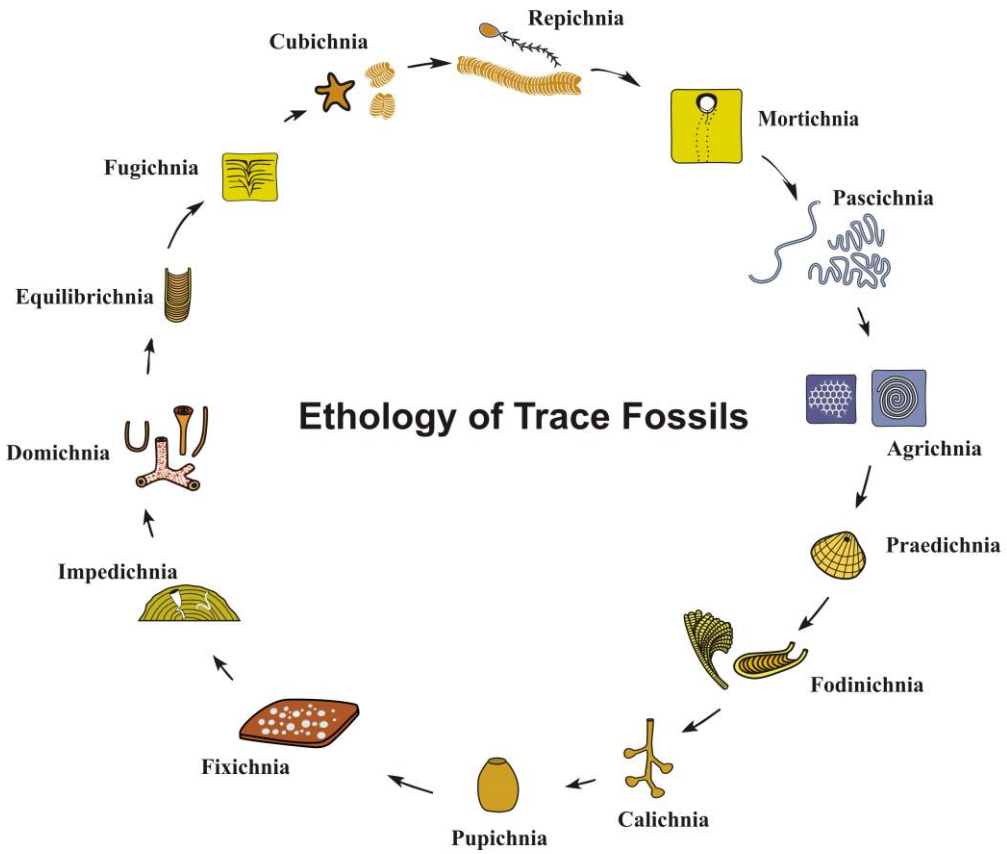


Figure 1.2 Ethological categories of trace fossils in common use today (after Buatois & Mángano, 2011).

Bioturbation is described as biogenic sedimentary structures that display the disruption of sedimentary fabrics or stratification features through the activity of an organism (Frey & Wheatcroft, 1989). Measuring the amount of bioturbation present in outcrop or core has been described and defined by Taylor & Goldring (1993), in which there are seven categories (Table 1.1) defining the bioturbation index (BI). This index was based off of categories originally defined by Reineck (1963), which adequately describe the variations of bioturbation observed; however, this is a semi-quantitative approach. The BI presented by Taylor & Goldring (1993) favors a descriptive approach, where each class is defined in terms of burrow intensity, the amount of overlap, and the clarity of the original sedimentary fabric. In situations where vertical sections are not sufficiently exposed, a bioturbation index can be assigned to the bedding-plane, as described by Miller & Smail (1997). Their proposed bedding-plane bioturbation index (BPBI) is a semi-quantitative approach, where five categories are assigned based on the percentage of bioturbation observed.

Table 1.1 Bioturbation Index (BI) as defined by Taylor & Goldring (1993).

Grade	Percent bioturbated	Classification
0	0	No bioturbation
1	1-4	Sparse bioturbation, bedding distinct, few discrete traces and/or escape structures
2	5-30	Low bioturbation, bedding distinct, low trace density, escape structures often common
3	31-60	Moderate bioturbation, bedding boundaries sharp, traces discrete, overlap rare
4	61-90	High bioturbation, bedding boundaries indistinct, high trace density with overlap common
5	91-99	Intense bioturbation, bedding completely bioturbated (just visible), limited reworking, later burrows discrete
6	100	Complete bioturbation, sediment reworking due to repeated overprinting

Beyond the analysis of depositional environments and bioturbation, ichnology can also be used to characterize the change in ichnodiversity through Earth history (Buatois & Mángano, 2016b, 2018). In ecological and macroevolutionary studies, the term global diversity has three clear divisions: alpha, beta, and gamma (Whittaker, 1972). Buatois & Mángano (2011) adapted the divisions that were defined by Whittaker (1972) and applied them in an ichnological sense to ensure a clear meaning within such studies. In this adaption, alpha ichnodiversity is the assessment of diversity within individual facies or environmental zones (Buatois & Mángano, 2011, 2013). It is observed that alpha diversity is the most commonly used division of ichnodiversity, typically employed in the identification of environmental changes in both space and time. Beta ichnodiversity is observed as useful in determining the degree of ichnospecies habitat differentiation, through the identification of ichnofaunas along a depositional profile. The final division of ichnodiversity is gamma, which is used to analyze potential trace fossil provincialism.

Trace fossils can provide insight into biotic changes through time, as they represent the morphology and behavior of the organisms that produced them (Crimes, 1994). These changes are summarized by the concept of alpha ichnodiversity, which addresses the trends of ichnotaxonomic richness through time (Buatois & Mángano, 2011, 2013; Minter et al., 2016a).

An example of ichnodiversity relevant to this study is the diversification of trace fossils through the Ediacaran-Cambrian transition leading up to the Cambrian Explosion (Mángano & Buatois, 2014). There are two contrasting hypotheses regarding the evolution of ichnodiversity during this transition (Buatois & Mángano, 2016a). The more classic approach implies a gradual diversification model, based on the work of Crimes (1994). This hypothesis focuses on the significant diversity increase recognized during the Vendian, a time period approximately synonymous with the Ediacaran (McCall, 2006), and lower Cambrian leading up to the Cambrian Explosion. Crimes (1994) displayed a steady increase of the recorded ichnodiversity through the Ediacaran to the lower Cambrian, with 36 ichnotaxa observed during the Ediacaran, followed by 66 ichnotaxa observed during the Terreneuvian, and finally 86 ichnotaxa observed during Epoch 2. The second hypothesis proposes a rapid diversification, which is based on revisions of Ediacaran ichnotaxa (Seilacher et al., 2005; Jensen et al., 2006; Mángano & Buatois, 2014). Comprehensive reviews of Ediacaran ichnotaxa show that ichnodiversity within the Ediacaran is significantly lower than previously published literature would suggest, and it is also observed that behavioral complexity is very limited at this time. Mángano & Buatois (2014) showed revised ichnodiversity values through the Ediacaran to the lower Cambrian, with 9 ichnotaxa observed during the Ediacaran, followed by 43 ichnotaxa observed during the Terreneuvian, and 55 ichnotaxa observed during Epoch 2.

While alpha ichnodiversity is a common area of focus, beta and gamma diversity studies are often overlooked (Buatois & Mángano, 2013). The application of beta diversity is of particular importance to this study, as this is the reflection of species differentiation along an environmental gradient (Whittaker, 1972; Sepkoski, 1988). Adapting this description to ichnology, Buatois & Mángano (2013) observed that beta ichnodiversity can identify the degree of differentiation between offshore and nearshore depositional environments (e.g., Mángano et al., 2013). In contrast, gamma ichnodiversity has the potential to suggest spatial trace fossil restriction, known as provincialism (Buatois & Mángano, 2011). These authors observe that trace fossil provincialism is not common, since trace fossils can be produced by more than one organism. This characteristic of trace fossils often results in paleogeographic distributions that are widespread (i.e. cosmopolitan), with rare cases of provincialism (e.g., Systra & Jensen, 2006; Jensen et al., 2013). In the investigation of a global trace fossil dataset, gamma ichnodiversity trends may become apparent.

1.4 A Review of GIS Software for Spatial and Temporal Analysis

Geographic Information Systems (GIS) are described as computer systems that allow a user to manage spatial data (Bonham-Carter, 1994). It is with these GIS systems that geologists can evaluate large volumes of geological data and the associated spatial data with ease. Included with most geological data is an approximation of time, whether that is based on radiometric ages, geological correlations, or biostratigraphy (Markwick & Lupia, 2002). The inclusion of time as an additional dimension is essential for understanding and modeling space, leading to a more thorough understanding of the distribution and evolution of spatial data (Claramunt & Thériault, 1995).

In a geological sense, GIS software is most commonly used as a tool to assist data visualization, spatial queries, merging datasets, and analysis (Bonham-Carter, 1994). For the effective application of GIS software it is important to understand the nature of how geological data is utilized. GIS uses spatial data as the primary component for organizing and displaying datasets, while the secondary components are the attributes, or non-spatial data (Volta & Egenhofer, 1993). Attributes, often referred to as metadata, are the associated descriptive or statistical component of the dataset. It is emphasized that both components of the datasets are equally important, and the effective application of GIS is centered around the intimate link between the two. Using GIS to visualize data offers an advantage over viewing tabular data, as visual patterns are much easier to observe when attributes are assigned logical symbology (Bonham-Carter, 1994). Data visualization can be further refined using spatial queries, where user defined criteria are imposed on the dataset and only locations that meet these restrictions are displayed. It is suggested that this technique is helpful during analysis, often assisting in the determination of what caused the spatial pattern (Bonham-Carter, 1994).

Dataset manipulation may be required to simplify visual analysis. Categorical coverage is a common mapping technique where attribute categories are defined first, followed by determining the spatial zones that will correspond to these attributes (Volta & Egenhofer, 1993). Volta & Egenhofer (1993) noted that less detailed coverages may be defined from the original categorical coverage, in which the categories are reduced into simplified attributes. This process, described as coarsening, ultimately merges initially unique attributes with their new generalized

attributes. Coarsening will result in fewer and less descriptive attributes and in turn fewer, but larger spatial zones (Volta & Egenhofer, 1993). GIS software also supports manipulation in the form of data integration, where a user may combine different datasets as layers on the same map, resulting in a new map from two or more previously distinct maps (Bonham-Carter, 1994). This process is particularly important, as the addition of separate datasets can assist with analysis and interpretation, leading to the recognition of patterns that would not have been possible with the analysis of individual datasets in isolation.

GIS software facilitates the visual analysis of spatial data and possible extrapolation of patterns observed (Bonham-Carter, 1994). This analysis is not exclusive to visual observations, as additional analysis may be carried out using statistics. Statistical analysis can serve three purposes to further supplement visual analysis, including: hypothesis testing, hypothesis generation, and model building (Birks, 1987). This author describes the use of multivariate data analysis as simply rotating or summarizing datasets and assisting with pattern recognition. The advantage of viewing statistical results in a spatial context was observed before modern GIS applications existed, with David et al. (1974) recognizing that the results from correspondence analysis could be plotted on maps to produce "striking results."

A limitation of GIS software is that geological datasets are confined to present-day spatial locations (Qin et al., 2012). Specialized software, such as GPlates, can overcome this problem by allowing users to calculate and view plate-tectonic reconstructions (Boyden et al., 2011; Qin et al., 2012). Plate-tectonic reconstructions are described as "the calculation of position and orientation of tectonic plates at an instant in the history of the earth" (Boyden et al., 2011). The visualization of these reconstructions is a valuable tool for understanding the progression of systems and processes at the surface of the Earth. This software allows the user to visualize and analyze deep time geological datasets, as data is attached to a plate tectonic reference frame (Boyden et al., 2011; Qin et al., 2012). The data is then reconstructed with the plates, revealing the suggested configuration of the tectonic plates at a specified time, according to the rotation model being used (Qin et al., 2012).

1.5 A Review of Exploratory Data Analysis and Data Mining

Exploratory data analysis (EDA) is the process of looking at data with different techniques to infer meaning (Tukey, 1977; Morgenthaler, 2009). Morgenthaler (2009) explained that there are no limitations to what tools are used by a data analyst, rather the limitations that exist are time and creativity. The purpose of EDA is described as discovering interesting connections between datasets, and identifying peculiarities. Morgenthaler (2009) noted that computational statistics, data visualization, and data mining are modern data analysis techniques that have been strongly influenced by exploratory data analysis.

Data mining is described as a procedure where large volumes of data are processed, followed by the use of selected methods to extract patterns that are then evaluated, and knowledge is gained (Lloyd-Williams, 1999; Han et al., 2012). Han et al. (2012) explained that the term data mining is often used as a synonym for an equally popular term, knowledge discovery from data (KDD). A general summary of the KDD or data mining process follows four steps: selection of data; transformation of data; data mining; and the interpretation and evaluation of the results (Figure 1.3) (Lloyd-Williams, 1999; Han et al., 2012). The process of data selection and transformation is the gathering of information from a database, at which point data transformation through the process of consolidation and summarization can take place (Han et al., 2012). The data mining step is described as the application of methods to derive patterns from within the data. Data mining adopts techniques from many fields, including pattern recognition, visualization, statistics, and machine learning to discover new information. Han et al. (2012) observed that the world is data rich and information poor, which makes data mining such a useful and powerful tool.

Statistics can offer analytical assistance in two distinct forms, one is in the summarization of data, and the other is to test hypotheses based on calculations of probability (Shaw, 2003). Of particular interest are techniques that provide descriptive statistics, resulting in a numerical description of patterns within or between variables (Shaw, 2003). Further, there are two types of tests recognized in statistics, those that are parametric and conversely those that are nonparametric (Davis, 2002). Parametric tests require datasets follow a normal or Gaussian distribution, the shape of which is known, while nonparametric tests make no assumptions about the dataset distribution. Nonparametric techniques can be used with data that follow a Gaussian

distribution; however, these tests are considered less powerful than their parametric counterparts in such a situation (Davis, 2002; Quinn & Keough, 2002; Motulsky, 2014).

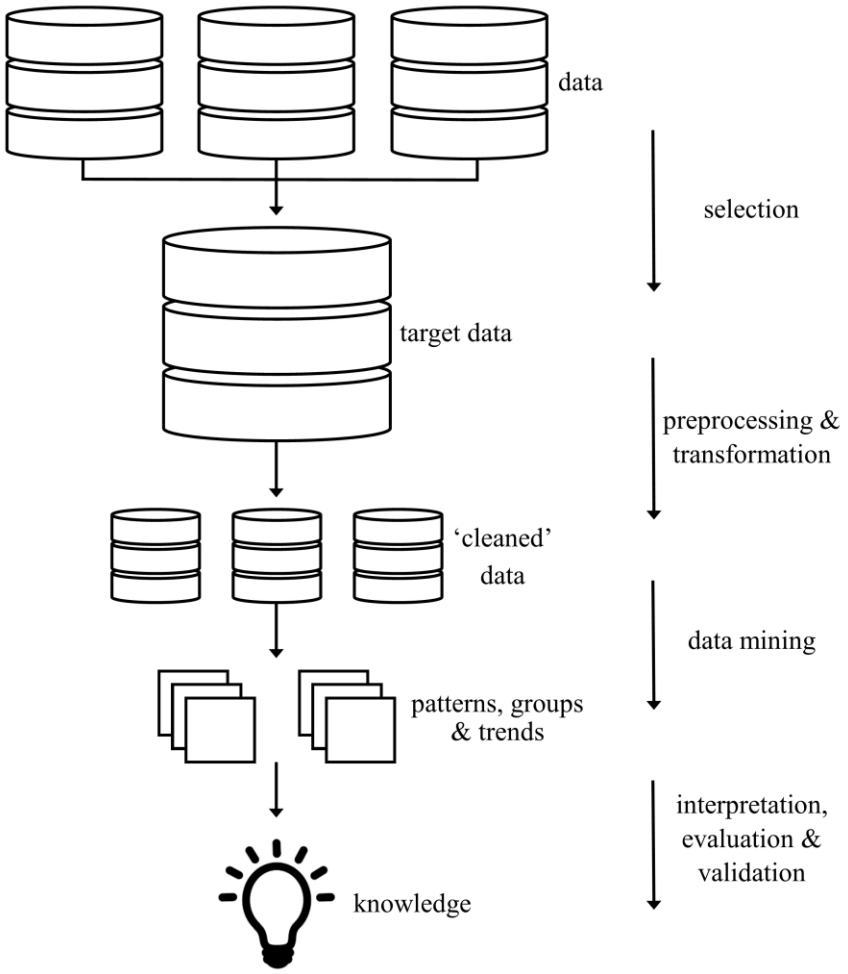


Figure 1.3 Diagram of the Knowledge Discovery from Data (KDD) process (after Lloyd-Williams 1999).

Before selecting a statistical test of any kind, the type of data being analyzed must be thoroughly understood as not all data types are suitable for particular statistical tests (Davis, 2002; Shaw, 2003). The four data types described by these authors are nominal, ordinal, interval, and ratio. Nominal data involves descriptions of categories, any numerical values assigned only serve as labels and are meaningless (e.g., colours, rock types, trace fossil ichnotaxa). Ordinal data involves categories that can be given an order, but a valid numerical difference between the ranked data cannot be assigned (e.g., Moh's hardness scale). Interval

data involves data that can be ordered and a difference between the two values is relevant (e.g., temperature). The final data type is ratio data in which rank order, differences and ratios between the values are all relevant (e.g., measurements of physical properties such as length, mass, volume).

Of interest to this project is the use of ordination for statistical analysis. Ordination is a term applied to mathematical techniques used to summarize and describe the underlying structure and patterns within multivariate data (Davis, 2002; Shaw, 2003; Legendre & Legendre, 2012). Ordination techniques plot results in a multidimensional diagram with the number of axes equivalent to the number of descriptors in the study (Legendre & Legendre, 2012). For ease of analysis, these authors observe that a reduced number of axis are selected which represent a large portion of the variability of the multidimensional data, thereby producing ordination in reduced space. Ordination techniques can be divided into two general classes, labelled as R-mode and Q-mode techniques (Davis, 2002). R-mode techniques are described as those that are concerned with the interrelations between variables, and they function by extracting eigenvalues and eigenvectors from a covariance or correlation matrix. Q-mode techniques are different in the fact that they are concerned with the relationships between objects, with the purpose of identifying patterns or groupings within multivariate space (Davis, 2002). Examples of R-mode techniques relevant to the geosciences include principal component analysis and correspondence analysis, while multidimensional scaling is an example of a Q-mode technique.

Principal component analysis (PCA) is a multivariate ordination technique that essentially treats a dataset as a geometrical object and determines the most informative orientation of axes to view the shape (Shaw, 2003). Legendre & Legendre (2012) explained that PCA is conducted through matrix algebra on quantitative data, which produces principal components within principal-component axes. The principal components describe the position of the objects in the new system of coordinates produced by a rigid rotation of the original system of coordinates. The rigid rotation means that Euclidean distances among the objects is preserved through this process, which is an important property that makes PCA unique among other ordination techniques (Legendre & Legendre, 2012). The results left to be interpreted by the user after a PCA ordination has been conducted are eigenvectors and eigenvalues (Davis, 2002; Shaw, 2003). The first new axis of a PCA ordination, referred to as principal component

one, represents the greatest variation within the original data (Quinn & Keough, 2002). The second new axis, principal component two, is orthogonal to the first axis, and represents the second greatest variation. There will be an equivalent number of principal components to the number of original variables, all of which are summarized by eigenvectors.

For ordinal or nominal data, correspondence analysis is another ordination technique that produces eigenvalues and eigenvectors (Davis, 2002). Correspondence analysis uses contingency tables between two variables, typically consisting of frequency counts of species found at specific sites (Legendre & Legendre, 2012). This ordination technique then calculates the chi-squared (χ^2) distance to quantify the relationships that exist between the rows and columns of the contingency table. These authors explain that the ordination is similar to PCA, but rather than preserving the Euclidean distance between objects, CA preserves the chi-squared distances between objects. Although CA produces eigenvalues and eigenvectors similar to PCA, they are not interpreted in the exact same way (Shaw, 2003). It is explained that eigenvalues for CA do not add to 1, as is the case with PCA, rather they add to a value known as the inertia of the dataset. To find the resulting eigenvectors of CA, each eigenvalue must be divided by the inertia of that dataset to show what percentage of the variation is explained by the ordination.

Both PCA and CA analysis produce descriptor-axes and object-vectors that can be plotted in fewer dimensions than the original data (typically two), which if plotted together result in the formation of a biplot (Quinn & Keough, 2002; Legendre & Legendre, 2012). Biplots reveal structure in the data that may not have otherwise been evident by displaying patterns of correlations between variables or relationships between observations (Greenacre, 2010). Of particular interest to this study is the production and interpretation of CA biplots, which Quinn & Keough (2002) described as point-point plots and Greenacre (2010) described as an asymmetric map. The interpretation of these CA biplots is relatively straightforward in that variables with large contributions to the position of the objects will be close to the object of interest on the plot (Quinn & Keough, 2002). In CA the chi-square distance is particularly susceptible to variables with small frequency counts, which plot on the periphery of the biplot and often lead to the incorrect interpretation of a highly significant occurrence (Legendre & Legendre, 2012). These authors suggest the outright elimination of the small frequency count variables, as they contribute very little to the first few ordination axes. As a guide to the effective elimination of

deceptive variables in CA, Legendre & Legendre (2012) suggested conducting CA in a stepwise fashion with the progressive elimination of rarely occurring variables. Recording and plotting the results of successive CA ordinations will display a significant drop in both total inertia and eigenvalues after the removal of too many variables (Figure 1.4). This technique visualizes the unimportance of small frequency count variables up to the point where they begin to have a significant influence on the components of CA.

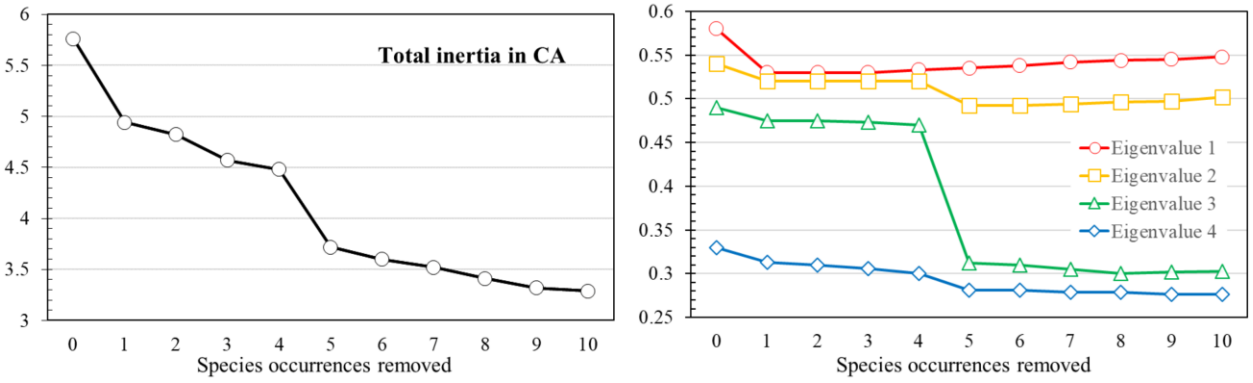


Figure 1.4 An example of removing species with very few occurrences from the correspondence analysis while recording the total inertia of the ordination and the first four eigenvalues (after Legendre & Legendre 2012). These plots show a distinct drop in the sixth iteration of CA of both total inertia and the third eigenvalue. As described by Legendre & Legendre (2012), it is at this point that too many species have been removed. Therefore, the removal of any species with four or fewer occurrences from the CA ordination would produce similar results without deceptive variables on the periphery of the biplot.

Another important characteristic of the chi-squared distance used in CA, is that it is not influenced by double zeros (Borcard et al., 2011). These authors observed that this characteristic makes CA particularly suitable for the analysis of species abundance data, as zero values may have several ecological meanings which make them difficult to interpret. Ecologically, the presence of a species (nonzero value) suggests a site met the minimal conditions for species survival, whereas the absence (zero value) may be caused by a variety of circumstances (Borcard et al., 2011). Examples described by these authors include the occupation of an ecological niche by a replacement species, the species has simply not reached the site despite favorable ecological conditions, the species may not show a regular distribution of the sites studied, the site does not contain favorable ecological conditions for the species, or finally the species is present but not

observed. To summarize, the presence of double zeros cannot be interpreted as resemblance because the absence may be caused by different reasons.

An additional multivariate ordination technique of interest to this study is multidimensional scaling (MDS), where calculated differences between samples are compared and an algorithm is used to find the best solution between dissimilarities and inter-object distances (Quinn & Keough, 2002). The result of this ordination is a plot with similar samples grouped close together, while dissimilar samples are far apart (Vermeesch & Garzanti, 2015; Spencer & Kirkland, 2016). Vermeesch (2013) describes two distinct types of multidimensional scaling that can be conducted, metric MDS, and non-metric MDS (nMDS). The distinction made between the two is in the type of regression that is utilized in determining the best fit of the ordination (Logan, 2011). Metric MDS takes advantage of metric regression, while nMDS takes advantage of monotonic regression to determine the stress value of fitting the ordination into the dimensions defined by the user.

The process of conducting MDS is computationally demanding, as a data matrix composed of *n* rows of samples and *p* columns of variables must be compiled (Holland, 2008). From this original data, a symmetrical *n* x *n* matrix of pairwise (dis)similarity must be calculated. Common measures of dissimilarity include the Bray-Curtis distance, Manhattan distance, Euclidean distance, and Jaccard distance for ecological data (Shaw, 2003), while measures such as the Chi-Square distance, Kolmogorov-Smirnov dissimilarity, Kuiper statistic, and the Cramér-von-Mises statistic are used to compare detrital zircon samples (Vermeesch, 2018).

This technique is distinctly different from previously described ordination techniques (i.e. PCA or CA) as the user defines the number of axes to which the data are then fit (Holland, 2008). Typically, two dimensions are chosen as the number of axes (e.g., Vermeesch & Garzanti, 2015) due to the ease of plotting and interpreting this number of dimensions. The quality of the nMDS results are described by the "goodness of fit" metric, often referred to as the stress value (Kruskal, 1964; Vermeesch, 2013). Kruskal (1964) described the interpretation of stress values ranging from a poor fit to a perfect fit (Table 1.2).

Table 1.2 Guide to interpreting stress values from Kruskal (1964) to determine the goodness of fit between dissimilarities and the distances of nMDS.

Stress (%)	Goodness of fit (g.o.f.)
20	poor
10	fair
5	good
2.5	excellent
0	"perfect"

Holland (2008) noted that MDS does not produce a unique solution, and subsequent analysis may result in similar but different ordinations. The primary sources of nonunique solutions are the starting configuration of your data and the algorithm that is used to determine the optimal solution. In subsequent sections R, and R studio have been utilized as the software for conducting multivariate analysis (MVA). Within this software there are two algorithms that are commonly used for nMDS. The first algorithm is isoMDS, which is a function provided by the MASS package (Venables & Ripley, 2002) within R. The isoMDS function conducts nMDS with an initial configuration defined by PCA followed by running the iterative algorithm until stress values for successive iterations does not change considerably. The second algorithm is the monoMDS function provided by the vegan package (Oksanen et al., 2018) within R. The monoMDS function also conductions nMDS; however, unlike isoMDS the initial configuration is randomly generated for the algorithm to run iteratively. The distinct difference of starting configurations between the two algorithms produces different ordinations. Holland (2008) observed that nMDS algorithms may become stuck on a local minima which is the best solution compared to all nearby solutions, but it is not the true best solution. To counter this issue, Oksanen et al. (2018) recommended running several random starts to compare the resulting stress values and determine if you have reached a local minima or a true "global" optimum.

1.6 A Review of Detrital Zircon

Detrital zircon has many applications, including the determination of the maximum age of stratigraphic successions, identification of unconformities in the absence of biostratigraphic information, determining provenance characteristics, providing detail to mantle and crustal evolution, suggesting the tectonic setting in which deposition occurred, and lend support for plate-tectonic reconstruction models (Fedo et al., 2003; Cawood et al., 2012). The utility of zircon as a detrital mineral in sedimentary systems originates from its high physiochemical resilience, it is found in a wide range of lithologies, it is able to resist metamorphic resetting, and it concentrates important trace elements (Cherniak & Watson, 2000; Cawood et al., 2012). Fedo et al. (2003) added a cautionary note that although detrital zircon analysis assists in unravelling the complexity of sedimentary systems, these results are most effective when used in conjunction with additional analytical methods.

Detrital zircon analysis can treat grain populations differently through qualitative or quantitative methods (Fedo et al., 2003). Qualitative analysis describes the selection of zircon grains based on optical characteristics such as transparency, colour, shape, and degree of rounding (Gehrels, 2000). The advantage of qualitative analysis is described as an increased chance of selecting and analyzing zircon grains from all age populations within a sample, whereas random selection can miss age populations that exist in minuscule quantities. Quantitative analysis is quite different, this method aims to gather a sample that is representative of the overall detrital zircon population within a stratigraphic unit (Fedo et al., 2003). A statistical probability describes the number of zircon grains that must be analyzed to ensure zircon grain populations of a specified size are sampled with a defined level of confidence (Dodson et al., 1988; Fedo et al., 2003). While Dodson et al. (1988) described the statistical equation to determine the probability of missing a zircon grain population, Fedo et al. (2003) presented a graphical representation of this equation (Figure 1.5).

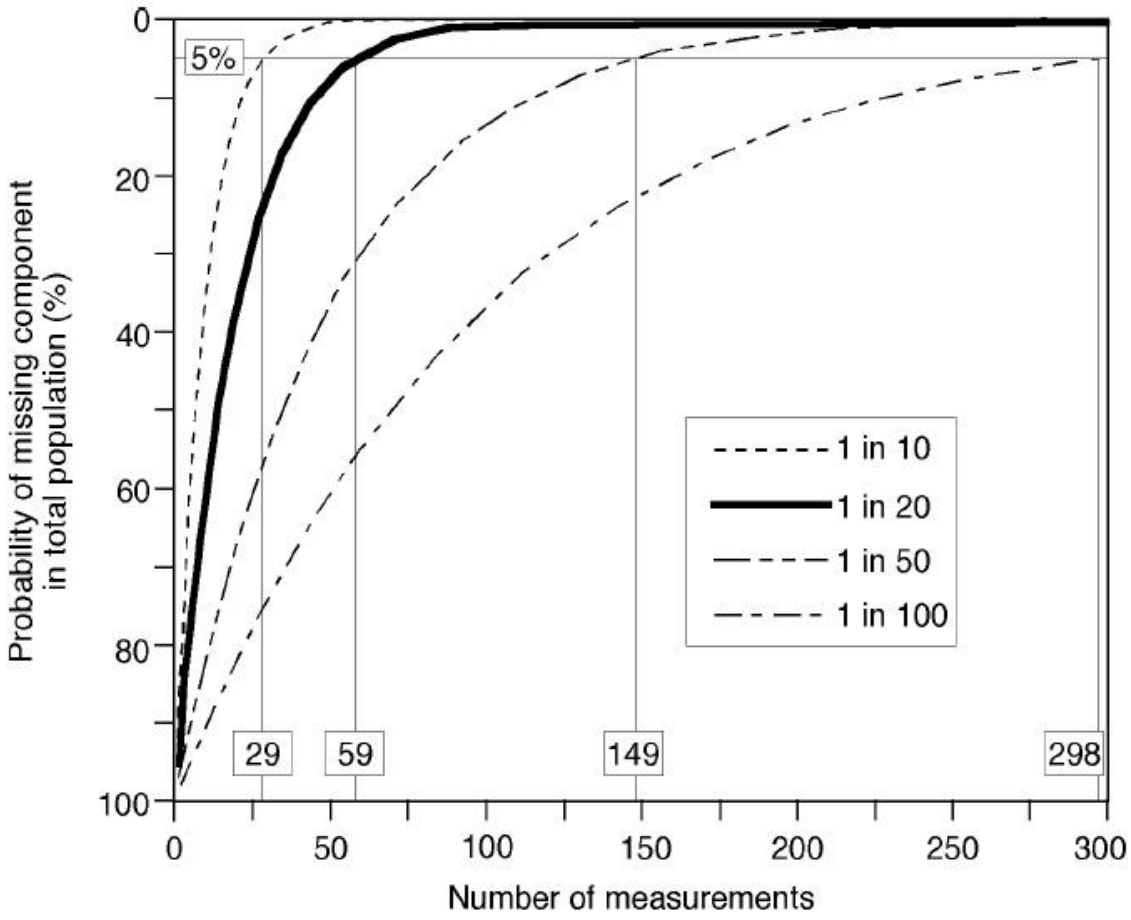


Figure 1.5 Graphical representation of the probability of missing a grain age population based on the number of measurements (after Fedo et al., 2003). The bold line indicates that there will be a 5% chance you will miss a grain age component comprising 5% of the total sample if you analyze 59 detrital zircon grains. The remaining curves represent different levels of confidence that a component comprising 5% of the total sample will be missed.

Detrital zircon grains can be used to constrain the age of siliciclastic stratigraphic successions, although age limitations established through this method typically lack precision and are usually only applied to Precambrian successions where biostratigraphy cannot be applied (Fedo et al., 2003). Generally, only a crude maximum age of deposition can be determined, based on the principle of inclusions where the youngest zircon grain is used to infer that the deposition age must be younger. In special cases where detrital zircon grains contain metamorphic rims, the contrast in geochronological ages between the core and rim can be used to define limitations on the age of deposition (Schiøtte et al., 1988). With this method, the maximum age of deposition is defined by the age of the detrital zircon core and the minimum

age of deposition is defined by the age of the metamorphic rim. Depositional age bracketing can also be applied in situations where a stratigraphic succession is crosscut by igneous rocks (Fedo et al., 2003). This method does not always lack precision, radiometric dating of detrital zircon can be very reliable in determining the deposition age of the stratigraphic unit when the zircon grains are extracted from volcaniclastics that have been interstratified with sediments (e.g., Bowring et al., 1993).

The origins of sedimentary rocks, referred to as sedimentary provenance, can be investigated by analyzing the petrographic, geochemical, and geochronological characteristics of detrital zircon grains (Fedo et al., 2003). Geochronology is the most common characteristic analyzed in provenance studies, where hundreds or thousands of individual zircon grains are analyzed through the use of laser ablation inductively coupled plasma mass spectrometry (LA-ICPMS) or secondary ion mass spectrometry (SIMS) techniques and radiogenic isotopic data is collected (Fedo et al., 2003; Ireland & Williams, 2003; Košler & Sylvester, 2003). The U/Pb ratios measured with either of these techniques can then be used to calculate isotopic ages, with the two most common calculated ages resulting from $^{206}Pb/^{238}U$ and $^{207}Pb/^{206}Pb$ ratios (Spencer et al., 2016). After parameters are defined to extract only high quality detrital zircon grain analyses (e.g., Spencer et al., 2016), the remaining data is typically plotted in probability density plots (PDPs) or kernel density estimation (KDE) plots to visualize the grain age distribution for each sample (Sircombe, 2004; Vermeesch, 2012). From these plots grain age peaks of interest can be identified and provenance sources may be suggested.

Fedo et al. (2003) noted that this type of geochronological analysis does have a distinct disadvantage in the fact that it can only provide information regarding the initial source of the zircon grains and it does not provide insight into possible sediment recycling. Recycling is a recognized and observed part of the sedimentary cycle, with zircon being especially susceptible to such processes due to its resistance to physical and chemical weathering (Fedo et al., 2003; Hadlari et al., 2015; Andersen et al., 2016). To identify possible sediment recycling Hadlari et al. (2015) suggested the use of temporal statistical analysis between samples to look for significant correlations among age spectra. Alternatively, these authors observed that detrital biostratigraphy (e.g., recycled palynomorphs) has also been shown to be an effective method in observing sediment recycling when available. Andersen et al. (2016) provided a more

conservative approach and suggests that prior geological knowledge such as stratigraphy and basin geometry should take precedent over theories constructed to explain the provenance of detrital zircon.

Petrographic techniques can assist provenance analysis through joint statistical analysis with U/Pb geochronology data (e.g., Stevens et al., 2013; Vermeesch & Garzanti, 2015). Stevens et al. (2013) showed the utility of adding heavy mineral data to geochronology results through combined PCA and MDS, displaying clear spatial heterogeneity of sand sources. Vermeesch & Garzanti (2015) combined geochronological results with heavy mineral data, bulk petrography, major element composition, and trace element composition into comprehensive statistical techniques for analysis. The two methods investigated by these authors are higher order versions of MDS known as Generalised Procrustes Analysis and Individual Differences Scaling, with both methods extracting geologically meaningful differences between datasets. Although sampling issues such as hydraulic sorting (Schuiling et al., 1985), source rock fertility (Malusà et al., 2016) and sampling bias (Malusà et al., 2013) should be acknowledged, Vermeesch & Garzanti (2015) observed that adding additional analyses to geochronology is beneficial to provenance analysis.

Detrital zircon samples can be used to infer the general tectonic setting into which the sediment was deposited by analyzing the grain age spectra produced by individual samples (Cawood et al., 2012). These authors observe that a cumulative proportion plot of crystallization age minus deposition age provides the basis to which a tectonic setting can be inferred (Figure 1.6). Cawood et al. (2012) indicated that the characteristics of sedimentary basins are a response to tectonic settings, of which convergent, collisional, and extensional settings can be identified using this method. Convergent settings (e.g., forearc and backarc basins) are recognized by a high proportion of detrital zircon grains with ages similar to the time of sedimentation, due largely to synsedimentary igneous activity. Collisional settings (e.g., foreland basins) are observed to contain a large portion of zircon grains showing an age within 150 Ma of the time of sedimentation, due in part to syn-collisional magmatism and magmatic arcs formed during ocean closure. Extensional settings (e.g., passive margins, rift basins, and intracratonic basins) display

detrital zircon grain ages that are significantly older than the time of sedimentation, due to the stability and large hinterland of these basins. This technique is useful when the original nature of the basin is no longer preserved; however, Cawood et al. (2012) cautioned that uncertain deposition ages may influence the classification of tectonic setting, especially convergent and collisional settings.

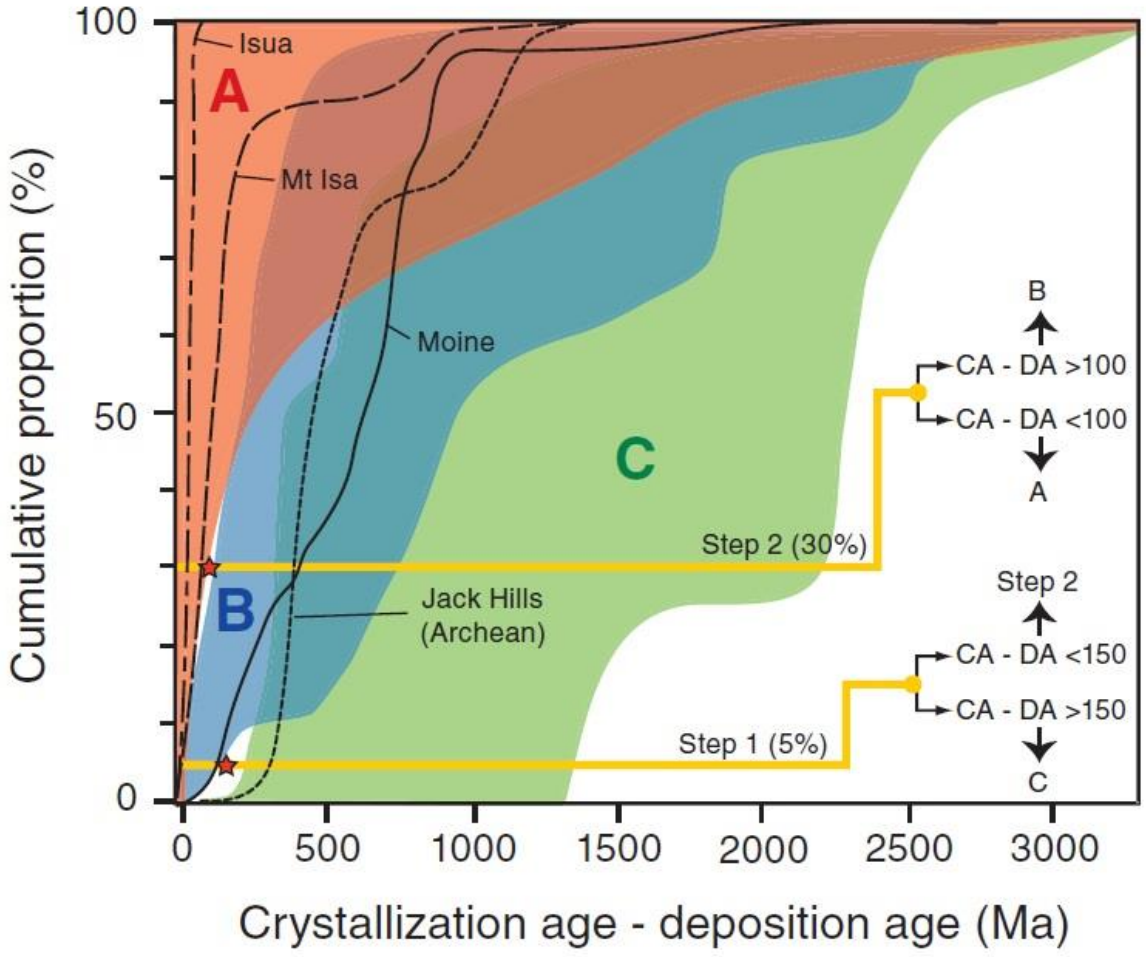


Figure 1.6 Summary of tectonic classification from a detrital zircon cumulative proportion plot (after Cawood et al., 2012). The green field, C, representing extensional tectonic settings is defined by a cumulative proportion curve that shows crystallization age (CA) minus deposition age (DA) greater than 150 Ma for the youngest 5% of the detrital zircon grains. If this criterion is not met, the second determining factor is if the CA – DA is less than 100 Ma in the youngest 30% of zircon grains. Where this is true, the sample will be classified as a convergent tectonic setting, represented by the red field, A. If the samples cumulative proportion curve falls between these two criteria, it is classified as a collisional tectonic setting, represented by the blue field, B.

In addition to providing an inferred tectonic setting, MDS can be used to plot detrital zircon samples in an attempt to observe the orogenic cycle (e.g., Spencer & Kirkland, 2016). This method has been shown to provide mixed results, producing patterns that show a logical geodynamic progression in some situations, and no pattern in others. Spencer & Kirkland (2016) provided two examples in which sedimentary successions display parts of the idealized orogenic cycle. Foreland basins of the Grenville orogeny display a transition from subduction and accretion to rifting and the formation of a passive margin. The opposite geodynamic evolution can be observed in the cratonic basin sediments of the Albany-Fraser orogeny, transitioning from rifting to subduction, followed by continental collision. These authors also describe situations where sedimentary successions display a more complex relationship (e.g., sediment recycling), which leads to the production of an MDS ordination that does not identify an intelligible pattern.

Finally, provenance analysis can provide assistance in refining or validating paleogeographic and tectonic reconstructions (e.g., Ross et al., 1992; Gehrels, 2000). This process uses tectonic reconstructions to test provenance theories, at which point they may agree or suggest alterations to accommodate results from their analysis. For example, Ross et al. (1992) used geochronology and geochemistry to analyze the provenance of sediments within the Belt-Purcell Supergroup of western North America. With the data obtained, these authors were able to show that the provenance for a particular problematic group of zircons was likely from Australia. When this idea was initially proposed it did not agree with the tectonic reconstruction models at the time, although their conclusions ultimately lead to model revisions.

2. IchnoDB: Structure and Importance of an Ichnology Database

Meek, D.M., Eglington, B.M., Buatois, L.A., Mángano, M.G. (2019). IchnoDB: Structure and Importance of an Ichnology Database. Submitted to Ichnos.

2.1 Abstract

The design of a relational database for ichnology data is presented herein, as an attempt to fill a gap in present-day paleontology databases. Currently, paleontology databases employ the architectural design and terminology used for body fossils interchangeably with trace fossils. This study suggests that fundamental differences between body and trace fossils make this practice inappropriate. These differences stem from the fact that trace fossils represent the behavior of the tracemaker, and not the phylogenetic affinities of an organism. This database, referred to as IchnoDB, has been tested by the authors throughout the design process to ensure functionality and to provide data for research purposes. At publication this database contains over 700 distinct trace fossil localities, describing 363 lithostratigraphic units in which 3758 individual trace fossil occurrences have been described and recorded. Previous ichnology databases have only produced a minor impact on the paleontology community, as they have not been able to make the leap from concept and design to public accessibility. In publishing this design, it is our desire to see established paleontology databases adopt this concept to encourage future research.

2.2 Introduction

2.2.1 Use of Databases

Large datasets come in many types of electronic files, including PDFs, spreadsheets, word documents, and most importantly, databases. Databases are the emphasis of this paper, as their organized structure promotes a much more efficient method for data sorting and gathering when compared to other electronic files (Mannila & Räihä, 1992; Harrington, 2016). Of particular interest are relational databases, originally proposed by E.F. Codd in the 1970's, which utilize the Structured Query Language, more widely known as SQL (Codd, 1970, 1979; Chamberlin & Boyce, 1974). SQL has become the standard language used within relational

databases, allowing users to enter, delete, modify, and retrieve data easily (Chamberlin & Boyce, 1974; Harrington, 2016).

Databases associated with geological information are typically designed to store data such that a 'core' table is produced with additional supporting 'lookup' tables containing supplementary data (Eglington, 2004). The core table is described as the principle data component that all other data will support and supplement. In a geological context, a core table could consist of many features, including geochronological data, geochemical data, fossil occurrences, etc. Lookup tables serve as supplementary data in addition to establishing acceptable nomenclature and the prevention of typographical errors (Eglington, 2004). Further design aspects to be considered within the aforementioned tables is the assignment of primary keys, which are defined as "a column or combination of columns with a value that uniquely identifies each row" (Harrington, 2016). A primary key must be singular and cannot contain null values, this helps to ensure that duplicate records are not entered into the database (Eglington, 2004; Harrington, 2016).

Databases provide several important functions for the scientific community, including a means to archive data, which is often required by public funding bodies (Uhen et al., 2013). Beyond their archiving potential, online databases provide access to information that would otherwise be restrictive in terms of the time required to personally compile. Further, it is with easy access to these large datasets that new types of analysis can take place through the examination of data from many perspectives in an effort to discover new relationships or oddities (Morgenthaler, 2009). It is observed that theoretically the same functions could be provided by paper archiving; however, databases offer a significant improvement through the increased efficiency in which you find, sort and analyze large volumes of data (Uhen et al., 2013).

2.2.2 Paleontology Databases and their Application

Compiling vertebrate and invertebrate data in the past has been difficult due to the nature of the fossil record, where publications span many decades and discoveries are geographically distinct and within unique stratigraphic units (Carrasco et al., 2007). This issue has been addressed through the creation of many different databases, with the Paleobiology Database (PBDB, https://paleobiodb.org), and the Geobiodiversity Database (GBDB,

http://www.geobiodiversity.com) being popular examples. The open access to massive amounts of data has had a positive impact on scientific research. This is exemplified on the PBDB website (accessed January 25, 2019), where they list 333 official publications that have used their data in some fashion. Many of these publications focus primarily on characteristics of vertebrates and invertebrates; however, a significant number use these datasets investigate larger questions. Several examples include the use of PBDB and Macrostrat (https://macrostrat.org) to quantify fossil distribution in hiatus-bound marine sedimentary stratigraphic sections (Peters & Heim, 2011), the use of PBDB and paleogeographic reconstruction models to analyze the association of continental distribution with Phanerozoic biodiversity (Zaffos et al., 2017), and the use of PBDB to improve paleogeographic reconstructions throughout much of the Phanerozoic (Cao et al., 2017).

Both paleontology database examples discussed (i.e. PBDB and GBDB) contain a wide array of vertebrate and invertebrate fossil data, along with limited amounts of trace fossil data. Ichnology is the study of traces produced by organisms (both animals and plants) on or within a substrate, including all issues related to bioturbation, bioerosion, and biodeposition (Bromley, 1990; Pemberton et al., 1992a; Bromley, 1996; Buatois & Mángano, 2011; Minter et al., 2016a). This branch of paleontology experienced rapid development following the seminal publications of Dolf Seilacher more than half a century ago (e.g., Seilacher 1955a, 1955b, 1964, 1967a). Ichnology serves as an important asset to paleontology, sedimentology, and stratigraphy owing to the special characteristics of trace fossils (Frey, 1975; Pemberton et al., 1992a).

Unfortunately, the lack of detail associated with the ichnology entries and limited data set size in PBDB and GBDB precludes effective use for many studies. Despite the lack of open access ichnology data, examples exist where large ichnology datasets have been compiled manually to study macroevolutionary trends and paleogeographic distributions (e.g., Orr, 2001; Uchman, 2004; Jensen et al., 2013; Mángano & Buatois, 2014; Buatois & Mángano, 2016, 2018; Buatois et al., 2016). Despite the time and effort to compile the datasets used in these studies, none have been digitized in a database format. The possible utility of such information is clear, with potential for use in future macroevolutionary and paleogeographic analyses, or as supplementary material to paleontology, sedimentology, and stratigraphic studies, as observed by Frey (1975) and Pemberton et al. (1992a).

2.2.3 Lack of Database Structure for Ichnology Data

As stated previously, compiling paleontology data is a difficult and time consuming task (i.e. Carrasco et al., 2007), with ichnology being no exception to these issues. The body fossil community has addressed this problem with the creation of multiple databases, while the same cannot be said for ichnology. There have been suggestions for the creation of ichnology-focused databases, such as IchnoBASE (Belvedere et al., 2011) and The Ichnology Database (Goldstein et al., 2010); however, no further publications exist regarding their design or potential for data storage and research. Both databases still exist in a dormant, publicly inaccessible state due to a general lack of funding (Belvedere, personal communication, 2018; Goldstein, personal communication, 2018). In the absence of an ichnology database, limited amounts of ichnology data have been digitized within body fossil databases. The inclusion of this data is an oversight, as it applies the conceptual framework and taxonomy used for body fossils on the ichnology data, which are fundamentally different.

This fundamental difference is observed at an ontological level, with trace fossils representing the behavior of an organism, whereas body fossils reflect the phylogenetic characteristics of an organism (Pemberton & Frey, 1982). That trace fossils represent evidence of behavior is suggested to be the core of ichnology (Seilacher, 1967b; Frey, 1975; Buatois & Mángano, 2011). In addition to this essential concept, trace fossils display several unique characteristics when compared to body fossils (Frey & Seilacher, 1980; Pemberton et al., 1992a). The most distinct among these characteristics include the possibility that a trace fossil can be produced by more than one animal species, and a single animal can create more than one type of trace fossil (Frey & Seilacher, 1980). Trace fossils also display a much longer temporal distribution when compared to body fossils, and they are typically limited to a narrow facies range which reflects an organisms response to paleoecological conditions (Frey, 1975; Pemberton et al., 1992b).

Differences in nomenclature between trace fossils and body fossils also exist, with ichnological taxonomy based upon ethology rather than phylogeny (Bertling et al., 2006). As a result, ichnological taxonomic hierarchy does not contain equivalent ranks to kingdom, phylum, class, or order used in zoological taxonomy. The basic units of ichnological taxonomy are the

ichnogenus and ichnospecies, where the ichnogenus describes morphological features resulting from highly significant behaviors and ichnospecies describes features resulting from behavior at less significant levels (Pemberton & Frey, 1982; Bromley, 1990; Bertling et al., 2006). The addition of ichnofamilies has been discussed, which would ultimately aid in hierarchical classification (Bertling et al., 2006).

A review of the general database designs for popular paleontology databases, PBDB (Alroy et al., 2001; Peters & McClennen, 2016) and GBDB (Fan et al., 2013), indicates that these databases capture much of the same data that would be required for an ichnological occurrences. To account for the fundamental differences between trace fossils and body fossils in current paleontology databases, very little needs to be changed. It is recommended that tables storing paleontology data include a field indicating the entry of body or trace fossil information. This accounts for the conceptual differences between trace fossils and body fossils and takes advantage of existing database structure. In this way, additional tables for ichnology data are not required and an efficient database design is maintained. Rather, with the recognition of fundamental differences, the entry of equivalent concepts in the existing fields of the tables is feasible (Table 2.1). The only exceptions are the ethology and architectural design associated with ichnotaxonomy, which do not have equivalent concepts.

Table 2.1 Concepts commonly associated with body and trace fossils are listed, highlighting the differences between these two branches of paleontology. Paleoecology displays equivalent concepts, whereas bioturbation and ichnotaxonomic classification details do not have body fossil equivalents.

	Body Fossils	Trace Fossils
Fossil Count		
Abundance	X	X
Bioturbation		X
Paleoecology		
Facies	X	
Ichnofacies		X
Guild	X	
Ichnoguild		X
Classification		
Ethology		X
Architectural Design		X

2.3 Proposed Ichnology Database Structure

The structure of a relational ichnology database, herein referred to as IchnoDB, is presented to illustrate a functional concept for data capture and subsequent use. This database is centered around three main tables, which derive data from multiple separate lookup tables. The main tables store geographic coordinates, locality details, and all trace fossils that have been recorded at the specific locality. Among these tables multiple other pieces of information are stored, including references, age, lithostratigraphic unit, depositional environment, maximum bioturbation intensity, ichnofacies, ichnoguilds, and any ichnotaxonomic revisions to the original classifications recorded at the localities.

2.3.1 Site Section

A database relationship diagram for the Site section is shown in Figure 2.1. The figure displays tables as individual boxes, with clearly defined names and a list of data fields recorded within each table.

This section includes seven distinct tables, with Site being the primary table of interest. Within this table there are nine stored properties that are required for each location captured. The first field in the table is the "IchSiteID", which is a numeric autonumber value generated by the database. This value serves as the primary key for the Site table. The "Latitude" and "Longitude" fields are captured as decimal degrees within the database using WGS 1984 as the geographic coordinate system. The typical signage for decimal degrees applies here, with positive values representing the northern hemisphere for latitude, and the eastern hemisphere for longitude, while the opposite cardinal directions are represented by negative values. The quality of locality data can vary within the literature, with older publications typically providing less detail than modern publications. To account for this issue, precision of the given geographic coordinates is captured within the "LatPrcn" and "LongPrcn" fields. Data for "Elevation" and "ElevPrcn" fields is captured in meters above sea level (AMSL), although this information is rarely provided with outcrop localities. Rather, these fields typically capture the depth of trace fossils within core samples. Finally, this table also records the user ID associated with the individual entering the data.

There are two lookup fields within the Site table, which are linked to the Countries, and User tables. The inclusion of the "UserID" field records the user entering data, which can serve multiple purposes. This ID can be used to trace erroneous data entries and search for further errors made during data entry by said user. The "UserID" can also be defined in a query, which may be desirable when a single user has entered data required for a specific project.

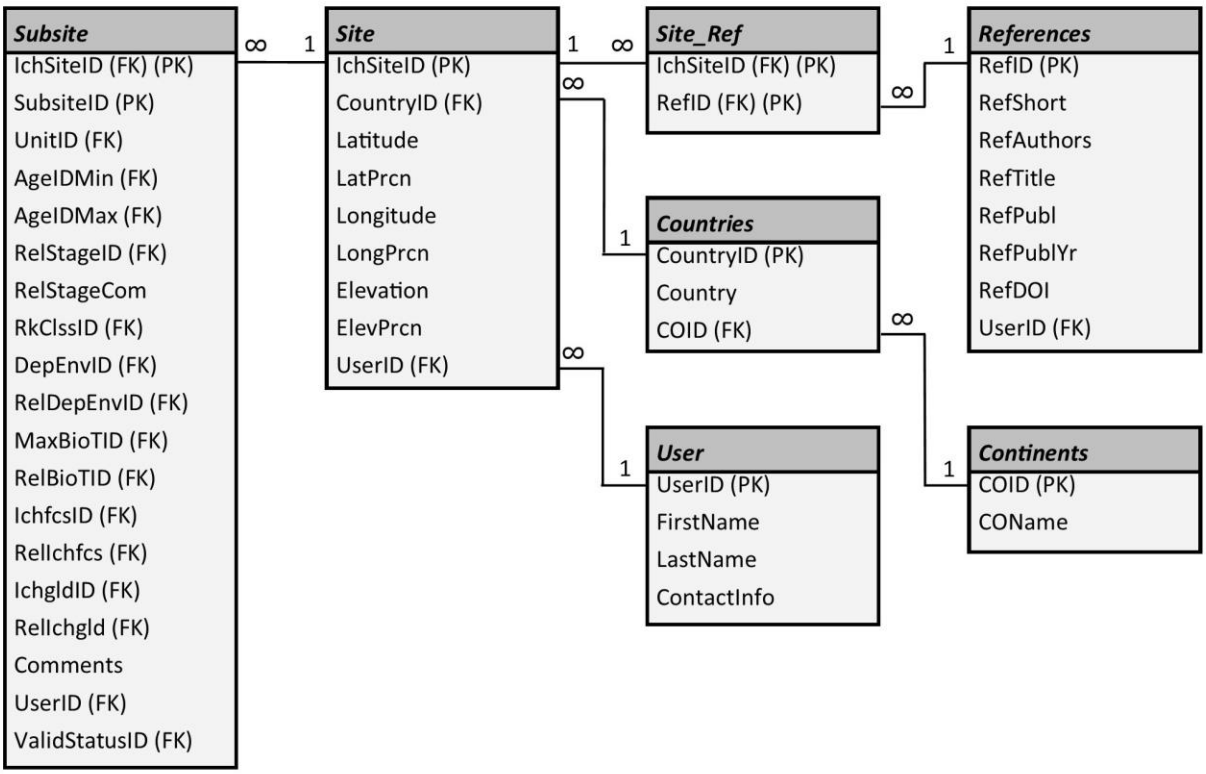


Figure 2.1 Relationship diagram of the Site section, which captures the geographic coordinates of a fossil locality. PK = Primary Keys; FK = Foreign Keys. All relationships are one (1) to many (∞).

Inclusion of the "CountryID" field may seem redundant; however, it serves two purposes. The first is a basic method to quality check the data entered. If the geographic coordinates entered do not fall within the political boundaries of the country recorded, this locality can be flagged for review. The second purpose is for easy attribute coarsening for use within GIS software (Volta & Egenhofer, 1993). All countries have been assigned to a specific continent,

which can then be extracted with ease, without constructing a quadrangle query based on latitude and longitude.

Tables Site_Ref, and References have been included to associate the reference from which each locality was extracted. These tables allow the database to capture data provenance, for easy data verification and review should the entry prove to be contentious (Markwick & Lupia, 2002). The Site_Ref table has a composite primary key, which requires the foreign keys of both "IchSiteID" and "RefID" to be a unique combination.

The last table in this section is Subsite, which records additional details regarding the locality being captured. This table is discussed in the following section.

2.3.2 Subsite Section

A database relationship diagram for the Subsite section is shown in Figure 2.2. This figure follows the same format outlined in the Site section, with tables displayed as boxes with their associated name and a list of data fields stored. Following the logic described by Markwick & Lupia (2002), the Subsite table records many aspects related to each subsite. These authors observed that data captured at a high resolution can be simplified, whereas refined data cannot be provided additional detail with such ease.

This section includes 15 distinct tables, with Subsite being the principal table of interest. The Subsite table is composed primarily of lookup fields, which source their data from the supplementary tables shown in Figure 2.2. Inclusion of the Subsite_Ichnotax table in this figure is meant to illustrate the connection to the final section of the database.

The Subsite table contains a composite primary key, created from the combination of the "IchSiteID" foreign key, and the "SubsiteID" primary key. The combination of these values must be unique, ensuring that duplicate entries are not stored within the database (Eglington, 2004). The "IchSiteID" field has been described in the previous section and is found as a lookup value in this table. The "SubsiteID" field is a numerical value that represents different stratigraphic positions at a location. This field contains a zero-based index, meaning that the initial value recorded is zero, rather than one. A zero-based index was chosen because it

supports logical additions below the initial recorded stratigraphic level should that be required. This field must increase (or decrease) with each new stratigraphic unit at a given location.

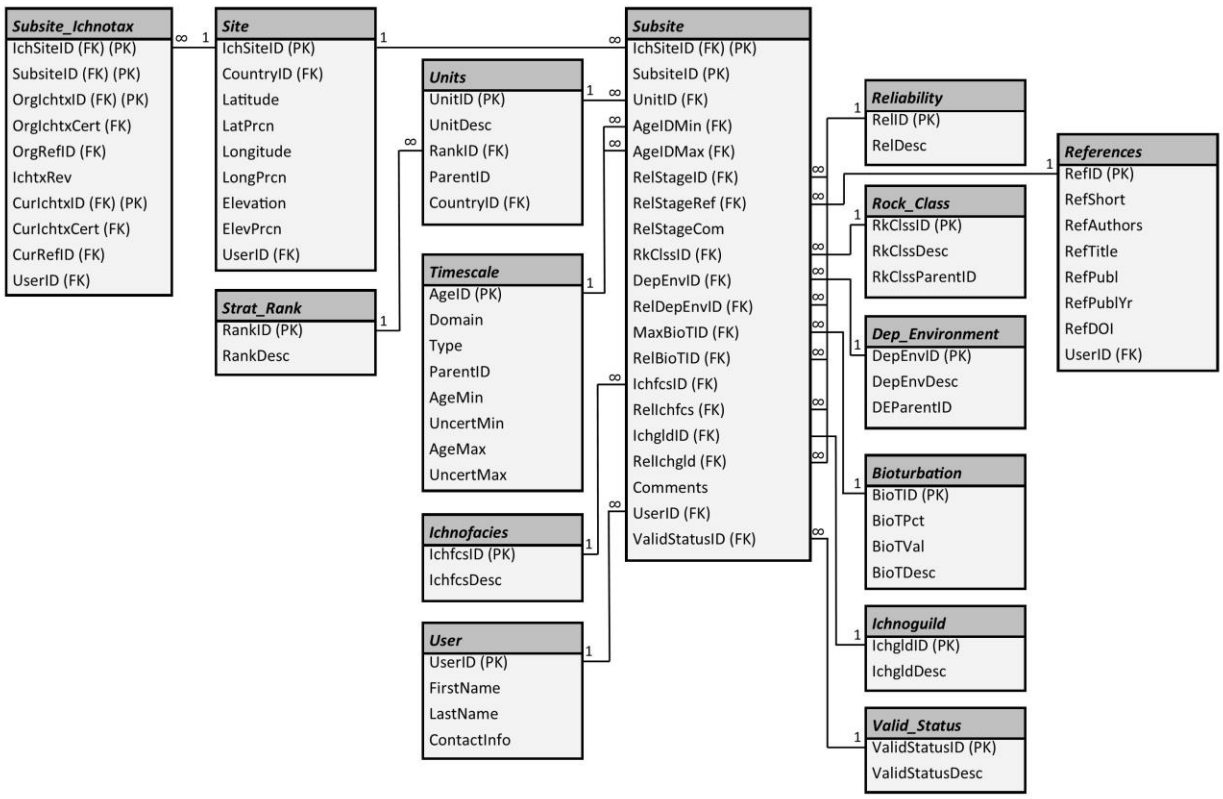


Figure 2.2 Relationship diagram of the Subsite section, which captures additional details (Stratigraphic, Sedimentary, and Ichnology information) of the previously recorded geographic site. PK = Primary Keys; FK = Foreign Keys. All relationships are one (1) to many (∞).

The "UnitID" field within the Subsite table is a numerical value that represents the stratigraphic unit being studied. This field is related to the Units table, which contains basic stratigraphic details including the name, the stratigraphic hierarchy, and the country in which it is present. Inclusion of the country of origin is to account for possible confusion with different units sharing the same name in separate countries. Users must query the existing units present in the Units table before creating new entries to ensure they are entering unique material. This also reduces the possibility of entering the same unit spelled differently, which could negatively impact search queries on data entered.

The “AgeIDMin” and “AgeIDMax” fields of the Subsite table are lookup fields, with values extracted from the Timescale table. The primary values within the Timescale table were extracted from the International Chronostratigraphic Chart (v2018/08) (Cohen et al., 2014, updated). All relevant Ages, Epochs, Periods, and Eras have been entered, along with their numerical values. This method of defining temporal range was selected based on the common reference to the International Chronostratigraphic Chart in publications, and the infrequent occurrence of temporal resolution finer than Ages within publications describing lithostratigraphic units. Several exceptions exist within this table, where geochronological data provides more detailed temporal information for a lithostratigraphic unit. This these circumstances the radiometric date and reference are recorded and used. A reliability field, “RelStageID”, is associated with the temporal range to signify the reliability of the designated minimum and maximum age. There is also a “RelStageRef” field, which records the reference from which the determination of the lithostratigraphic unit’s temporal range was derived, again recording data provenance (Markwick & Lupia, 2002). This is followed by a “RelStageCom” field for any comments related to the temporal determination, which may provide a brief justification for the reliability of the recorded age.

Lithology and depositional environment details are recorded in the “RkClssID”, “DepEnvID”, and “RelDepEnvID” fields. A list of lithologies in which trace fossils are preserved is compiled within the Rock_Class table, acting as the source for the “RkClssID” lookup field. The “DepEnvID” field captures the depositional environment associated with the lithostratigraphic unit being investigated at the site of interest. A list of 84 distinguishable depositional environments was compiled by the authors and stored in the Dep_Environment table to serve as the source for the “DepEnvID” lookup field. Due to the variability in reported depositional environments within publications, a reliability field, “RelDepEnvID”, was associated with the depositional environment to record any uncertainty in the determination of the recorded environment.

There are three ichnology specific fields within the Subsite table, the first of which is the “MaxBioTID” field. The values found within this lookup field are selected from the Bioturbation table. These values, ranging from 0 to 6, are based on definitions of the degree of bioturbation put forward by Taylor & Goldring (1993). The intensity of bioturbation can vary

widely within a formation or member, so this field is meant to capture the maximum value of bioturbation intensity observed within the lithostratigraphic unit of interest at the locality being referenced. An associated "RelBioTID" field is included to capture the reliability of the maximum bioturbation intensity values recorded. This is helpful when providing estimated values from published figures where no values are provided in the text.

The next ichnology specific field within the Subsite table is the "IchfcsID", which captures an ichnofacies interpretation for the specific locality. This lookup field draws acceptable entries from the Ichnofacies table. The concept of ichnofacies was established in several papers published by Dolf Seilacher first in German and later in English (e.g., Seilacher, 1964, 1967a). An ichnofacies is the description of a 'prototype' assemblage of trace fossils, which are characteristic of specific regimes representing environmental conditions recurrent through space and time (Frey & Seilacher, 1980; Pemberton et al., 1992b; Buatois & Mángano, 2011). The ichnofacies concept has been shown to be particularly useful in conjunction with facies analysis and sequence stratigraphy, through the refinement of paleoenvironmental interpretations and the identification of significant sequence stratigraphic surfaces (Pemberton et al., 1992b; Buatois & Mángano, 2011). As with other fields containing data that may be influenced by subjectivity, an associated "RelIchfcs" field captures a reliability of the ichnofacies assigned.

The last ichnology specific field in the Subsite table is "IchgldID", which captures the ichnoguild determined (or lack thereof) for the specific locality. This lookup field is based on the information within the Ichnoguild table. The ichnoguild concept was originally proposed by Bromley (1990, 1996), as an ichnology specific adaption to a concept applied to invertebrate paleontology (Bambach, 1983). This concept reflects three parameters in food source, bauplan, and use of space (Bromley, 1990, 1996). These guilds can then be used to understand trends in the utilization of ecospace through time (Bambach, 1983). Further, this concept has potential for paleoecological and evolutionary analysis of ichnofaunas (Buatois & Mángano, 2011). Again, this field is supplemented with a related "RelIchgld" field in which reliability of the ichnoguild determination is recorded.

The final three fields within the Subsite table include "Comments", "UserID", and "ValidStatusID". The "Comments" field has been included to record information unique to the

subsite being entered, as an additional method of identifying the provenance of the data source. The "UserID" field serves the same purpose in the Subsite table as it does in the previously discussed section, which is to record the user entering the data. The final lookup field in this table is "ValidStatusID". The values entered within this field are sourced from the Valid_Status table, which provide several scenarios regarding the status of a record. This first scenario is the indication that the record has been entered, but not verified by a qualified database user. The second scenario is the indication that the record has been verified by a qualified database user. The final scenario is if there has been an objection raised regarding the record of interest. If an objection has been recorded, the record is submitted to additional review by a qualified database user and a judgement will be passed regarding the specific objection recorded.

As previously mentioned, the final table in Figure 2.2 is Subsite_Ichnotax, which is related to the Subsite table through the "IchSiteID" and "SubsiteID" fields. The details of this table are outlined in the following section.

2.3.3 Subsite Ichnology Section

A database relationship diagram for the Subsite Ichnology section is shown in Figure 2.3, following the same format described in the previous two sections. This section completes the data entry process, associating trace fossils observed with the relevant subsite. Among the details captured are the ichnotaxonomic determination, ichnotaxonomic certainty, reference, and possible ichnotaxonomic revisions following the original publication.

The primary table of focus in this section, Subsite_Ichnotax, is related to the rest of the database through the Subsite and Site tables. The Subsite_Ichnotax table has a composite primary key, composed of the "IchSiteID", "SubsiteID", "OrgIchspID", and "CurIchspID" foreign key fields. This specific combination of foreign keys was selected to avoid duplicate entries of an ichnotaxa at any given locality.

The "IchSiteID" lookup field functions in exactly the same fashion as described in the Subsite table of the previous section, with values extracted from the Site table. The "SubsiteID" field records the stratigraphic unit of interest at a particular locality. Combining these two ID's relates all captured data in the previous two sections of the database to the trace fossils entered in this section.

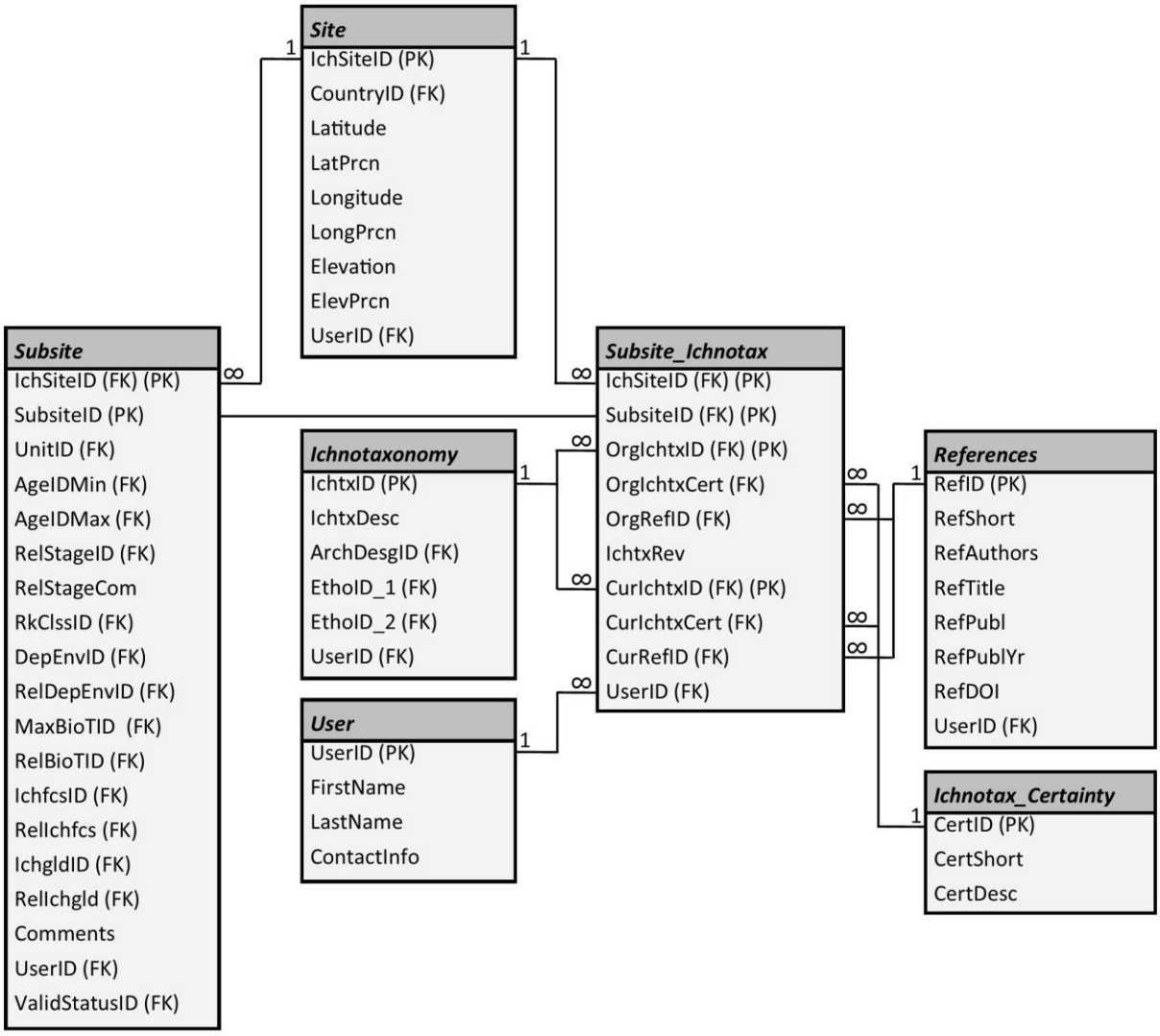


Figure 2.3 Relationship diagram of the Subsite Ichnology section, which captures the trace fossils recorded at the previously described locality. PK = Primary Keys; FK = Foreign Keys. All relationships except for Subsite "SubsiteID" to Subsite_Ichnotax "SubsiteID" are 1 to many (∞). The previously mentioned relationship is many (∞) to many (∞).

With a clear reference to the site and formation of interest, the Subsite_Ichnotax table then focuses on the trace fossils reported at the aforementioned site with the "OrgIchtxID" lookup field. The values for the "OrgIchtxID" field are extracted from the Ichnotaxonomy table, where all ichnotaxonomy is recorded and given a unique numeric ID. This table includes several other categorizations associated with the specific ichnotaxonomy, including the architectural design of each ichnotaxon entered, as defined by Buatois et al. (2017), and the dominant

ethological determinations for each entry (for review see, Buatois & Mángano, 2011 and Vallon et al., 2016).

The next lookup field in the Subsite_Ichnotax table is “OrgIchtxCert”, which records any taxonomic qualifiers listed in the publication. These values are extracted from the Ichnotax_Certainty table, where qualifiers such as aff. and cf. can be recorded. This table also contains an ID for ichnotaxa that have not been included in the figures of the publication, which adds a level of uncertainty as the database user cannot attempt to confirm proper identification by the authors of the publication.

The final lookup field dealing with original ichnotaxonomic classification in the Subsite_Ichnotax table is the “OrgRefID”. This field cites the author of the publication from which the data is being extracted. The values for the “OrgRefID” are extracted from the tbl_References table.

The “IchtxRev” field is a Yes/No field used to indicate if the original classification of the trace fossil has been changed in subsequent publications or ichnotaxonomic reviews. There are many examples of such reviews for Ediacaran and Cambrian trace fossils, such as Seilacher et al. (2005) and Jensen et al. (2006). These publications, among many others, have resulted in frequent alteration of original ichnotaxonomic assignments. Qualified database users can also reclassify original observations if they deem necessary, in which case a reference citing unpublished observations by said user would be entered.

All fields with a “Cur” prefix (i.e. “CurIchtxID”, “CurIchtxCert”, and “CurRefID”) act in the exact same fashion as their “Org” counterparts. This duplication of the aforementioned fields is to capture any subsequent ichnotaxonomic revisions following the original publication. If no revisions have been applied, these fields will hold the exact same ID values as the “Org” fields. The inclusion of ichnotaxonomic revisions is again following the previously described logic of Markwick & Lupia (2002). These details allow users to observe trends in ichnotaxonomic revisions through time, or search for junior synonyms should that be required.

The final lookup field in the Subsite_Ichnotax table is “UserID”. As described in previous sections, the values for this field are extracted from the User table. This field captures the user that entered the data, providing an audit trail for administration or querying purposes.

2.4 Discussion

IchnoDB was created using Microsoft Access, which was the selected database management system (DBMS) due to availability as part of the Microsoft Office suite on Windows platforms. MS Access does present serious limitations, such as the lack of concurrent users and a lack of online availability. To properly address these problems, the database design and content would have to be implemented into a DBMS that supports concurrent, online use (e.g., Firebird, MySQL). The purpose of this publication is not to promote the design and use of IchnoDB on a specific DBMS. Rather, the aim is to promote ichnological data digitization and subsequent use through the description of a database structure. It is observed that paleontology databases share many similar design aspects (Markwick & Lupia, 2002), with PBDB (Peters & McClennen, 2016) and GBDB (Fan et al., 2013) being no exception. Current paleontology databases need only include a field indicating the branch of paleontology to which the record pertains to. In doing this, the ontological differences between body fossils and trace fossils are recognized, and equivalent concepts can be captured within the database structure. The lone exceptions are ethological and architectural design classifications, to which body fossil equivalents do not exist. In these instances, two new tables would have to be created, which would enable proper categorization associated with each ichnotaxa.

The inclusion of detailed ichnology data into established paleontology databases would provide many benefits, open access to data being primary among them. Through the use of exploratory data analysis, there are many possible applications for this information, with the only limitations of its use being the creativity of the analyst (Morgenthaler, 2009). In addition, a database would address data preservation issues (e.g., Uhen et al., 2013) where unpublished supplementary material is at risk of being lost. Further, the time required to compile ichnology datasets would be significantly reduced. Although the ultimate goal of a database is to contain a comprehensive dataset, it is likely that researchers will still be required to compile individual publications to supplement datasets while less time consuming methods are being developed (e.g., PaleoDeepDive (Peters et al., 2014)). Finally, the inclusion of detailed ichnology data in an established paleontology database would solve previous funding issues cited for the lack of an ichnology database.

To illustrate the utility of the database structure presented, the design has been tested by entering ichnology data from published peer-reviewed articles reporting trace fossils observed in formations of an Ediacaran to early Cambrian age, as initially summarized elsewhere (e.g., Mángano & Buatois, 2014; Buatois et al., 2016). Although these tests focused primarily on ichnotaxa and related paleoecological data, the inclusion of additional ichnology specific details at individual sites (e.g., ichnofacies, ichnoguilds, bioturbation intensity) provide a platform for meaningful future analysis. Presently, IchnoDB contains 712 distinct localities recorded globally (Figure 2.4), with 3758 trace fossil records distributed between the localities. Obvious applications of this data include visual and statistical analysis of ichnology specific details such as the evolution of ichnofaunas, paleoecological trends, and paleogeographic distributions, or the supplementary use of the data in other geology or biology related studies.

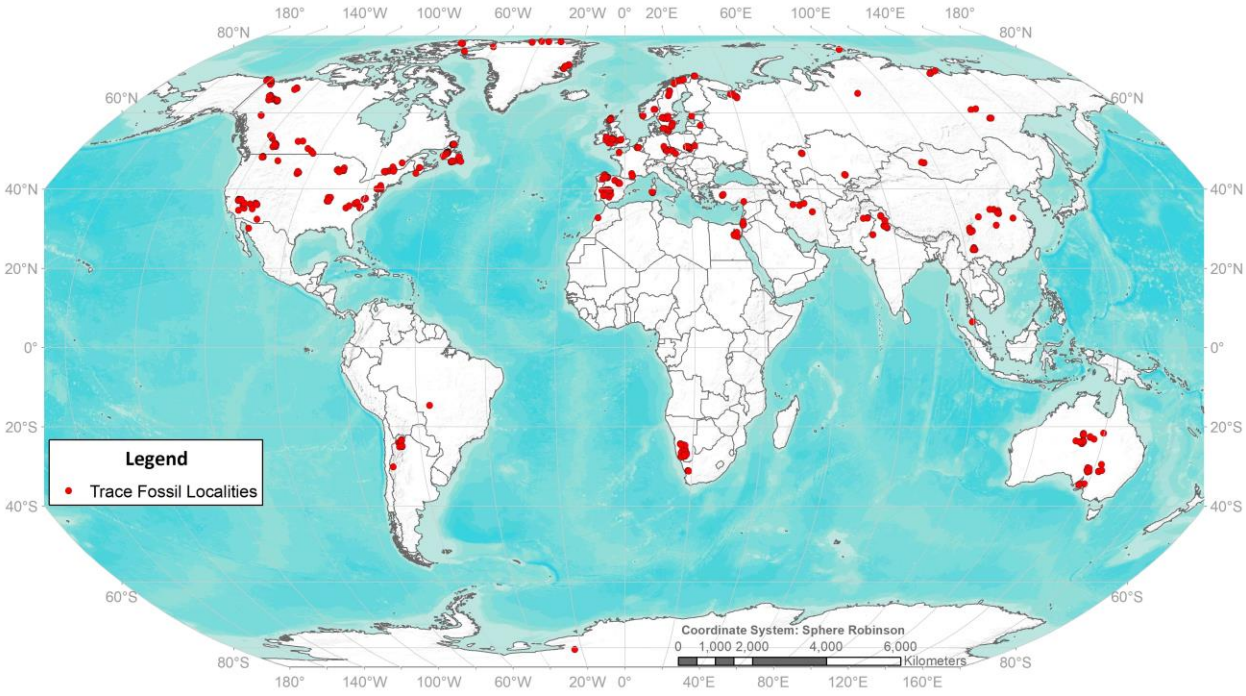


Figure 2.4 Global distribution of all trace fossil localities contained in Ediacaran to early Cambrian strata, stored within IchnoDB.

2.5 Conclusions

Although it is possible to capture ichnology data in current online paleontology databases, this practice is flawed and does not account for the fundamental differences between trace fossils and body fossils. These differences are based on the nature of trace fossils which record the organism behavior in its interaction with the substrate, while body fossils are fundamentally remains that record the phylogenetic affinity of past organisms. Without a clear distinction between these two very different paleontologic objects, the potential of trace fossils may be underexploited. To address this issue, current paleontology databases must include a field within the existing database structure that indicates the branch of paleontology to which the record is referring to. This field then permits different, but equivalent concepts to be recorded for trace fossils (e.g., ichnofacies, ichnoguilds, bioturbation intensity). The detailed ichnology records captured in this way allow users to take full advantage of the distinct characteristics associated with the separate fields of paleontology, and their particular objects of study. In addition to this, simply adding several new fields to a database maintains its original efficient structure.

Further, the database structure presented in this paper is relatively simple. The presentation of this database design is meant to serve as a starting point to which improvements can be made, and more importantly, it is intended to serve as a potential catalyst in the amendment of current paleontology databases. As previous stand-alone ichnology databases have not received overwhelming support, inclusion of an ichnology specific section within existing databases may provide a solution to this issue. The benefits of a possible open access ichnology section within a paleontology database would not only apply to the ichnology community, but to all academics and industry professionals interested in the use of paleontology data. Hypothetically the resulting ichnology data could be used to study ichnology specific questions, or it could be used as a supplement to issues involving other branches of geology or paleobiology such as paleontology, sedimentology, sequence stratigraphy, and paleogeography. Finally, open access to a large dataset would reduce the time prohibitive issue with current ichnology data use, and it supports exploratory data analysis, both important factors in future research.

3. Addition of Supplementary Geological Data to Support Visual Analysis

The purpose of this thesis is to investigate the association between depositional environments and trace fossil occurrence, which is now possible with access to the digitized ichnology data contained in IchnoDB. The intention is to use both visual and statistical data mining techniques to look for patterns or trends within the data. The statistical data mining techniques need no supplemental data apart from what can be extracted through IchnoDB. Visual analysis is inherently less quantitative, therefore the addition of supplementary datasets may provide assistance in evaluation and interpretation. Bonham-Carter (1994) encouraged the addition of supplementary geological datasets for visual analysis, having observed that additional datasets may support pattern recognition that would not have been apparent otherwise.

An important observation to consider in the visual analysis of paleontological data is the non-uniformity of fossil preservation both spatially and temporally (Holland, 2016). With a more thorough understanding of what creates this preservation bias, appropriate supplementary datasets can be selected to supplement analysis. Holland (2016) indicated that the primary control of fossil preservation non-uniformity is the stratigraphic record, as fossils are contained within sedimentary rocks. Tectonic subsidence and eustasy are described as primary factors in the accumulation and destruction of strata, which ultimately control preservation.

Tectonic subsidence is an important factor in the creation of sedimentary basins, imparting a strong influence on the extent and type of the resulting basin (Holland, 2016). Subsidence through Earth history is observed to be a result of two tectonic mechanisms. The first is stretching and cooling of the continental lithosphere, leading to the creation of rifts, cratonic basins, and passive margins. The basins resulting from this type of tectonic activity are observed to be the longest lived and the most spatially extensive through Earth history. The second tectonic process in which basins are formed is continental lithospheric flexing and bending (Holland, 2016). Foreland, forearc, and backarc basins are all produced from this type of lithospheric deformation. As these basins occur in more active tectonic regimes, they are shorter lived and less likely to survive into deep Earth history because of the high probability that they will be destroyed through metamorphism or subsequent uplift and erosion (Holland, 2016).

Finally, this author observed that all basins experience changes in subsidence and eustatic sea level change over time, the principles of which are the basis of sequence stratigraphy.

The type of basin produced by tectonic subsidence and the subsequent extent and longevity are the primary control of fossil preservation (Holland, 2016). This author observed that basin types show a wide range of sizes, which in turn effects the sampling of fossils, particularly those restricted to few biogeographic localities. The longevity of basins also controls preservation, with basins created by lithospheric cooling persisting for an order of magnitude longer than basins created by lithospheric stretching or flexing (Holland, 2016). Clearly the fossil sampling opportunities within the later type of basin will be much greater than in basins that existed for much shorter periods of time. The final observation made by Holland (2016) was the preference in preservation of different depositional environments between basins. Marine environments dominate basins controlled by cooling, whereas basins formed by lithospheric flexing and stretching commonly preserve terrestrial depositional environments.

Beyond the primary control of fossil preservation, Holland (2016) listed multiple secondary factors that control the non-uniformity of preservation. The influence of sequence stratigraphy, in both subsidence and eustatic sea level changes are observed to influence fossil preservation. The cyclic nature of sequence stratigraphy, observed in the supercontinent cycle, influences both sea level changes and basin formation (Meyers & Peters, 2011; Nance & Murphy, 2013; Holland, 2016). The supercontinent cycle, often referred to as the Wilson cycle, results in non-uniform basin production (Holland, 2016). It is observed that during periods of supercontinent break-up, continental lithosphere stretching and cooling produces rifts, cratonic basins, and passive margins. Whereas during supercontinent formation, collisions between continental plates results in lithospheric flexing and production of foreland basins. Further, it is observed that long-term sea level curves are strongly correlated to the Wilson cycle (Meyers & Peters, 2011). Another factor influencing fossil preservation is the influence of eustasy on a global scale (Holland, 2016). This author noted that basins of a similar age and type will have similar subsidence rates, and they will therefore be influenced by eustatic sea level changes in the same way. This results in correlated stratigraphic stacking patterns and depositional environment preservation, controlling the availability of sampling from particular environments. The final stratigraphic factor controlling the temporal effects of fossil preservation is the gradual

loss of “doomed” sediments through time (Holland, 2016). These sediments are described as those deposited in regions where subsidence is not taking place, resulting in erosion through time. Doomed sediments can also include basins that form within tectonically active regions, such as convergent margins. Erosion of these sediments result in an irreversible decline in the probability of fossil preservation through time. Finally, Holland (2016) suggested that biological factors also impose a bias in the preservation of fossils through time. This is described as a function of species evolution and extinction. Recently evolved species are geographically limited, and through time geographic range and abundance increase, followed by gradual geographic reduction until extinction occurs.

With a better understanding of the factors controlling the non-uniformity of fossil preservation both spatially and temporally, including additional datasets to display these influences will support increasingly robust visual analysis. The inclusion of detrital zircon data is particularly suitable for the analysis of paleontological datasets. Examples of applicable detrital zircon data use include its application to evaluate paleogeographic models, infer tectonic settings in which deposition occurred, and in the analysis of orogenic evolution. Ensuring the accuracy of paleogeographic reconstructions through detrital zircon provenance analysis benefits any data visualized using paleogeographic models. Additionally, detrital zircon data can be used to infer the tectonic setting in which a sediment was deposited (Cawood et al., 2012). Including this indicator with a paleontology dataset would provide a visualization of the preservation bias between tectonic settings. The last clear benefit of using detrital zircon samples to supplement the visualization of paleontology data is through the investigation of geodynamic evolution. Understanding the geodynamic regime adds additional context to the area of investigation.

In addition to detrital zircon data, paleogeographic reconstructions add another level of context for visual analysis. Visualizing data points in a paleogeographic context provides a relevance not possible in the observation of current geographic localities. Further, including relatively simple paleogeographic settings such as orogenic belts, terrestrial, shallow and deep marine regions, provides additional context to paleogeographic localities.

4. Use of Detrital Zircon Data to Test Plate Tectonic Models and Infer Tectonic Settings

4.1 Introduction

Over the past 30 years there has been a dramatic increase in the number of geochronology studies conducted, with zircon as the mineral of interest (Spencer et al., 2016). This increase has been facilitated by the development of laser ablation techniques along with cost reduction to conduct analysis (Košler et al., 2002; Košler & Sylvester, 2003; Spencer et al., 2016). Zircon is often the mineral of choice for geochronology analysis due to its inclusion in a range of lithologies, its resistance to weathering (Fedo et al., 2003), and a high closure temperature to resist metamorphic resetting (Cherniak & Watson, 2000). With the increasing volume of geochronology studies conducted, many have been focused on detrital zircon. As a result, there has been an increase in the amount of raw data produced, and it is this data that will be used to supplement visual analysis of ichnology data.

Large compilations of previously published detrital zircon geochronology datasets have been made publicly available for research purposes (e.g., Belousova et al., 2010; Voice et al., 2011; Dhuime et al., 2012; Roberts & Spencer, 2015; Puetz et al., 2018). These datasets are not without issues, most notably the lack of spatial coordinates for individual samples and the lack of lithostratigraphic detail. The largest and by far most detailed of the aforementioned datasets is that compiled by Puetz et al. (2018), in which the authors capture isotope data, U-Pb ages, limited lithostratigraphic information, and GPS coordinates. Despite the detail captured, the lack of an approximate deposition age poses challenges when using this data in conjunction with paleogeographic models and in the inference of a tectonic setting. Although it is possible to provide a maximum deposition age through the identification of the youngest detrital zircon grain age, this is often a poor estimation (Fedo et al., 2003). Considering this issue, previously published detrital zircon geochronology data will be compiled manually as part of this thesis to conduct visual provenance analysis and for the inference of tectonic settings.

In compiling supplementary data from previously published data, several challenges arise. Principal among them is the prohibitive nature of the time required to extract and compile the data in a standard format. This issue stems from the inconsistent nature in which data is

reported, and the file types in which the information is shared. Considering these obstacles, it was decided that data compilation would focus on publications from North America. In addition to the detrital zircon geochronology compilation, a compilation of igneous and metamorphic ages was extracted from the DateView database (http://sil.usask.ca/Databases.htm; Eglington & Armstrong, 2004; Eglington, 2004) to facilitate provenance analysis. The igneous and metamorphic geochronology data will be used to indicate possible source localities for the compiled detrital zircon dataset.

With a compiled detrital zircon dataset, there are two ways in which this will supplement visual analysis of a paleontology dataset. As it is the intention of this study to conduct visual analysis in GPlates, detrital zircon provenance analysis is used to assess the quality of the plate model being used. This is accomplished through the visualization of accepted provenance theories to ensure the plate model in use does not present any conflicts with the theories being observed. Further, by using the model presented by Cawood et al. (2012), tectonic settings are inferred for the samples that contributed to the detrital zircon analysis. Plotting these results in a paleogeographic reconstruction supports the capability of analyzing the geodynamic regime during a defined period of time, whether that is narrowly defined or across the broad history of a region.

4.2 Methods

4.2.1 Data Gathering

The data required for this study were extracted from supplementary files supplied with previously published articles. The supplementary files often excluded sample information that was not strictly related to the U-Pb ratios extracted from zircon analysis. Therefore, additional details such as the approximate deposition age of the unit from which the sample was taken, lithology, and spatial coordinates were extracted from the publication and included. These datasets were then compiled into a consistent file type and format for data preprocessing. The final compiled dataset contains sample ID, analytical technique, lithology, approximate lithostratigraphic deposition age, spatial coordinates, $^{206}Pb/^{238}U$, $^{207}Pb/^{235}U$, $^{207}Pb/^{206}Pb$ isotope ratios and ages, associated one sigma uncertainties for ratios and ages, calculated discordance, and when available $^{176}Lu/^{177}Hf$, $^{176}Hf/^{177}Hf$ isotope ratios, calculated $\varepsilon_{Hf(t)}$ values, and calculated

Hf two-stage depleted mantle (T_{2DM}) model ages. Although not all aspects of the data have been utilized in this study, the decision to capture as much detail as possible was influenced by database guidelines suggesting that detailed data is easy to reduce, whereas sparse data is impossible to refine (e.g., Markwick & Lupia, 2002).

After the data was compiled, the individual analyses were subjected to a set of criteria to ensure that only high-quality data was being used for further analysis. A best age is determined from $^{206}Pb/^{238}U$ and $^{207}Pb/^{206}Pb$ measured ratios, where the $^{206}Pb/^{238}U$ calculated age and uncertainty is used for results less than 1.0 Ga, and the $^{207}Pb/^{206}Pb$ calculated age and uncertainty is used for results over 1.0 Ga. Spencer et al. (2016) observed that the precise cutoff used between the two isotopic ages varies in the literature, leading to the selection of 1.0 Ga, as it appears to be the most common of the values selected. Furthermore, detrital zircon analyses were excluded if the uncertainty of the selected best age was greater than or equal to 30 Ma. Techniques for selecting an appropriate cut-off value regarding uncertainties are discussed by Puetz et al. (2018); however, a value of 30 Ma was selected due to the observed decrease in detail beyond this point and simplicity. The amount of discordance calculated for each detrital zircon grain was the final constraint regarding subsequent use. Analyses were excluded if discordance values exceeded 10%, which is described by Spencer et al. (2016) as a standard for this value.

4.2.2 Visualizing Probability Density Plots

Detrital zircon studies often visualize the resulting probability density plots (PDPs) or kernel density estimates (KDEs) of their samples by stacking them on top of each other (e.g., Gehrels & Pecha, 2014; Hadlari et al., 2014; Rainbird et al., 2017). This method is effective with a small number of samples (e.g., twelve), beyond this point such plots are cumbersome and visually overwhelming (Vermeesch, 2013). To address this issue, samples are plotted in FitPDF (Eglington, 2018b), which uses a different technique for viewing multiple PDPs at the same time.

FitPDF allows you to plot hundreds of PDPs simultaneously by altering the viewing axis. This software plots each PDP as a line along the appropriate deposition age of the sample. As a result, the PDP is viewed from above, with the grain age plotted on the x axis and the deposition

age plotted on the y axis (Figure 4.1). The software then uses colour to represent the height of each probability curve, which have all been normalized to 100 for this type of plot. As is demonstrated in Figure 4.1, where a PDP is above 80%, that segment of the line is coloured red, where the PDP is above 60% that segment of the line is coloured orange, etcetera. This viewing technique promotes pattern recognition on a much larger scale as compared to the traditional method of stacking PDPs.

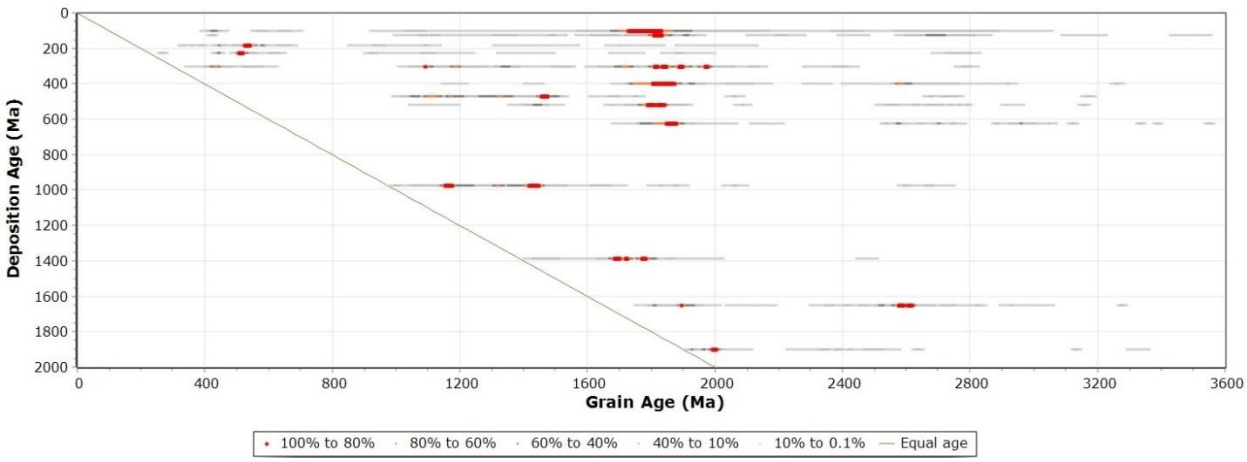


Figure 4.1 Visualization of a multiple plot created in FitPDF. In this example there are 13 samples with deposition ages ranging from the Paleoproterozoic to the Mesozoic, with a total of 890 individual detrital zircon grain analyses. Deposition ages (Ma) are plotted on the y axis, while grain ages (Ma) are plotted on the x axis. In this view from "above" the exact shape of the probability density plot for each sample cannot be observed; however, colour symbology displays where each PDP crosses user specified thresholds (e.g., red line segments represent where each PDP is above 80%, orange line segments are where the PDP is between 60% and 80%, etc.). Additionally, an equal age line is plotted where grain age and deposition age are equal. This assists in identifying samples that may have been affected by metamorphism or where the deposition age assigned to a sample is likely incorrect.

4.2.3 Supplementary Datasets

Datasets extracted from the DateView database (http://sil.usask.ca/Databases.htm) are an essential source of supplementary data for visual provenance analysis. DateView contains over 110,000 geochronology records available to the public, of which 39,000 metamorphic and igneous crystallization ages with associated spatial coordinates were extracted from across the globe. These data points will highlight potential sources for the detrital zircon grains when plotted simultaneously in GIS software. In addition to geochronology information,

thermochronological data was also extracted from DateView. This data was fundamental in the identification of areas likely undergoing denudation as a result of tectonic uplift, further highlighting potential source regions of the detrital zircon grains.

An additional supplementary dataset included for visual provenance analysis was extracted from a database of paleocurrent information. This database compiled by Brand et al. (2015) contains paleocurrent data from around the world, spanning the Proterozoic to Phanerozoic eons. The extracted data was plotted using GIS software and transformed into line segments to represent the paleocurrent direction at the time of deposition using the bearing distance to line tool in ArcGIS. Despite a lack of temporal resolution within this dataset during the Proterozoic, the data provides a general idea of sediment transportation direction.

4.2.4 Associating Possible Provenance Sources

With the large number of samples compiled in the detrital zircon dataset, associating possible provenance sources for every grain age within a PDP would be visually overwhelming. Instead, only significant grain age peaks of each sample were selected for visual provenance analysis. To define these significant grain age peaks, FitPDF was used. This software allows the user to define a cutoff value for each PDP that has been normalized to 100%, above which the range of grain ages will be defined and exported in the form of an excel spreadsheet. For this study, a value of 80% was defined as the baseline for a significant grain age peak. The resulting grain age ranges for each sample where this criterion was met were then exported. The exported grain age ranges of interest can then be used in conjunction with any dataset of potential source localities, typically using SQL queries.

The igneous and metamorphic geochronology dataset extracted from DateView was used to identify possible sources for the compiled detrital zircon grain analyses. SQL queries in Microsoft Access were used to identify all matches between the DateView data and the grain age ranges extracted from the detrital zircon data. Considering the temporal aspect of visual analysis, the deposition age of the sample to which provenance sources have been matched must be included for proper use in GPlates. This newly compiled file was then exported to GIS software and spatial join tools were used to identify the tectonic plates on which the sources were located to be used with the PalaeoPlates Model for visualization.

4.2.5 Using Plate Tectonic Models

With the data gathered to this point, our final step is to utilize the spatial and temporal details of the detrital zircon samples for visual provenance analysis. Due to the ever-changing location of Earth's tectonic plates, paleogeographic reconstructions are used to add spatial and temporal context to this type of analysis. To achieve this goal the paleogeographic reconstruction software, GPlates, along with a plate tectonic model, PalaeoPlates (Eglington, 2018c) is used to view all datasets simultaneously.

Several steps are required to prepare the data for use in GPlates. The first step is to address the temporal aspect of the data. For crystallization or metamorphism ages, minimum and maximum ages are added to the datasets by subtracting and adding the uncertainty from the recorded age. The detrital zircon samples are viewed in context of their deposition age, to which an uncertainty is not typically recorded. It was deemed reasonable to infer an uncertainty of 10 Ma for deposition ages of the detrital zircon samples. After this has been completed, the spatial aspect of the data must be addressed. GPlates requires that each data point specify the tectonic plate on which it is located within the model used. GIS overlay tools are used at this point to conduct a spatial join between the point data and the polygon shapefile of the plates used within the model of choice. After the spatial join has been completed, the data is ready to be exported as a shapefile to plot in GPlates for visual analysis.

4.3 Results and Discussion

The dataset compiled for this study contains 36,418 individual U-Pb analyses conducted on detrital zircon grains, extracted from 403 samples across North America (Figure 4.2). After applying the criteria described in the previous section to the detrital zircon U-Pb dataset, a total of 22,239 analysis are carried forward to be used in pattern recognition and visual provenance analysis. Using the FitPDF software to plot all of the PDPs simultaneously (Figure 4.3), clear patterns emerge which have been described previously for North America (e.g., Voice et al. (2011), Figure 3). Grain ages cover a wide spectrum, but the most common are those from 200 Ma to 600 Ma representing the Cordilleran & Appalachian Orogenies, from 1000 Ma to 1500 Ma

representing the Grenville Orogeny, from 1600 Ma to 1900 Ma representing the Trans-Hudson Orogeny, and from 2600 Ma to 2750 Ma representing the Superior Orogeny.

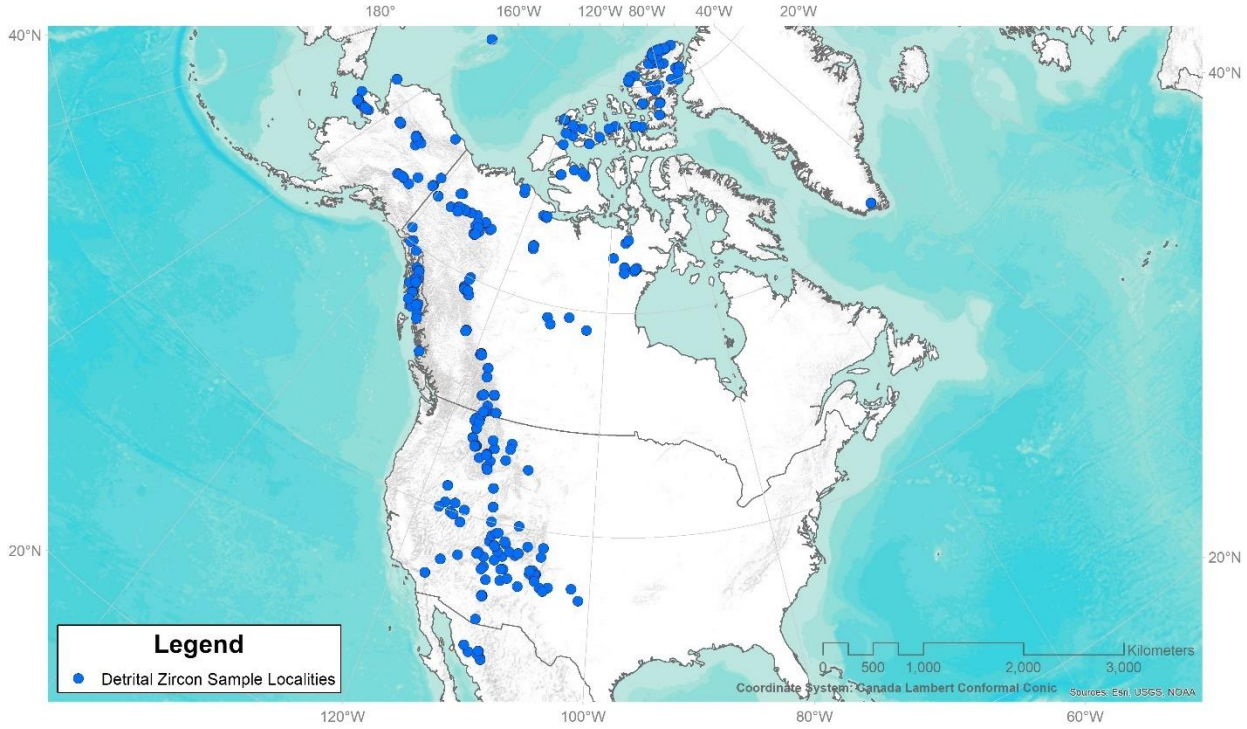


Figure 4.2 Map of North America displaying the location of all 403 samples from which detrital zircon data has been compiled.

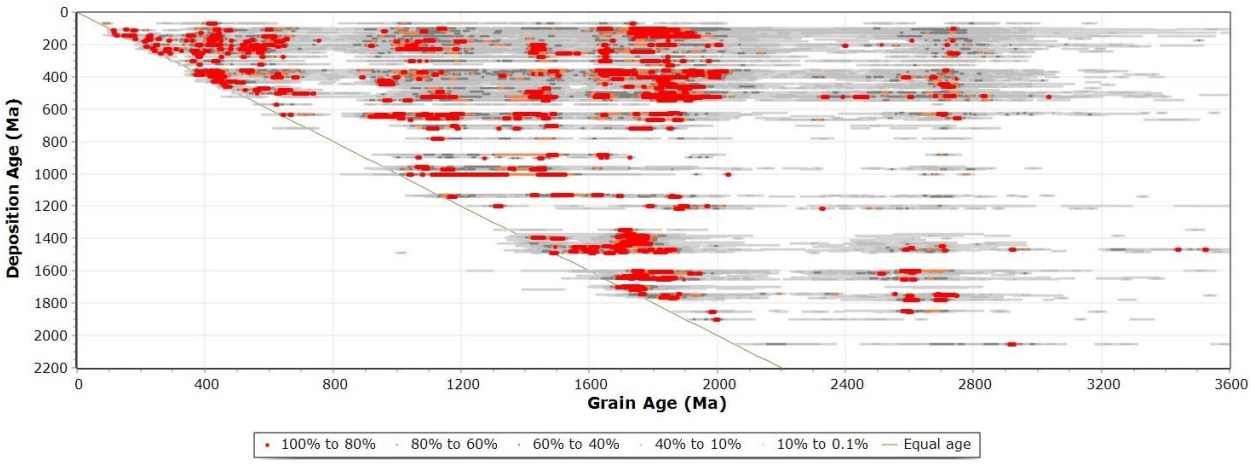


Figure 4.3 Multiple probability density plot from the compilation of previously published data across North America. A total of 22,239 individual detrital zircon grain ages are retained to construct 403 unique PDPs.

With this compiled detrital zircon data and the use of supplementary datasets (e.g., crystallization, metamorphic, and cooling ages, and paleocurrent data) paleogeographic reconstruction software can be used for visual provenance analysis. This type of analysis can identify flaws in the model if provenance cannot be explained, or it can provide support should no discrepancies exist. Further, tectonic settings inferred from a sample that has been used to extract and analyze detrital zircon grains can provide indicators for comparison against plate tectonic models.

4.3.1 Visualizing Established Detrital Zircon Provenance Theories to Support the Validity of Plate Tectonic Models

The Amundsen Basin and the Mackenzie Mountains contain abundant Neoproterozoic siliciclastic sequences that have been thoroughly studied in an attempt to determine their provenance (Young, 1979; Rainbird et al., 1992, 1997, 2017). Early studies of Neoproterozoic sediments in Northern Canada suggested a correlation between the Shaler Group on Victoria Island, and the Mackenzie Mountains Supergroup in Northern Canada (Young, 1979; Young et al., 1979). Young (1979) also observed that paleocurrent directions within these Neoproterozoic sediments were predominantly west to northwest, ultimately leading to the suggestion that the Grenville Orogeny may have been a provenance source. This theory was tested further by Rainbird et al. (1992) where they used U-Pb data from detrital zircon grains within the Shaler Group to support the suggestion that much of the sediment was supplied by the Grenville Orogeny over 3,000 kilometers away. Later studies used more U-Pb detrital zircon data and Sm-Nd isotope data (Rainbird et al., 1997), along with improved geochronology techniques (Rainbird et al., 2017) to further support the theory of a vast Neoproterozoic drainage system across Laurentia.

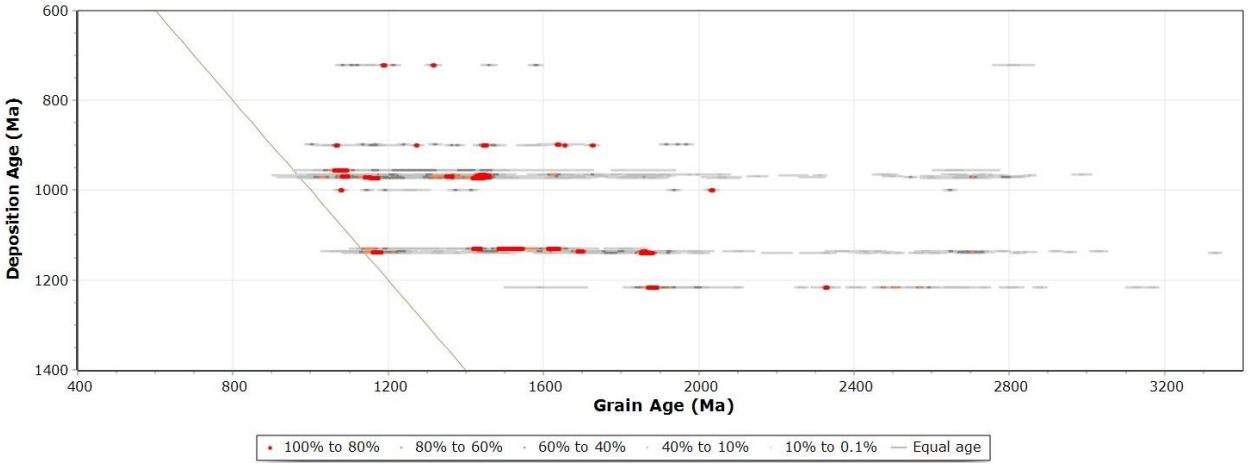


Figure 4.4 Multiple probability density plot of the 11 samples extracted from Rainbird et al. (1997) and Rainbird et al. (2017).

Using the data of interest, together with paleogeographic reconstruction software, this theory can be visualized. The raw geochronology data collected from Rainbird et al. (1997) and Rainbird et al. (2017) results in 11 samples within the Amundsen Basin and the Mackenzie Mountains. Using the FitPDF software to display these 11 probability density curves simultaneously, recurring high probability peaks between 1.1 to 1.2 Ga, 1.4 to 1.5 Ga, 1.6 to 1.7 Ga, and 1.8 to 1.9 Ga are observed (Figure 4.4). Extracting the high probability (<80%) peaks from FitPDF and using this data in conjunction with the crystallization and metamorphic data extracted from DateView, files were created for use in GPlates.

The result of this process (Figure 4.5) suggests similar provenance sources of the detrital zircon grains across Laurentia throughout the Neoproterozoic. The first reconstruction at 1130 Ma shows that the Grenville orogeny is the probable source of the two younger zircon grain ages found within samples of this age based on the colour ramp applied to source localities. The older grain age peak for this sample does not appear to be supplied from the Grenville. Rather it is possible that the darker brown source localities (~1800 Ma to 1900 Ma) were exposed and eroded through fluvial processes, or sediment recycling is responsible for the inclusion of these grains. This reconstruction, and supporting paleocurrent data, suggest sediment transport was directed across Laurentia in an east to west / northwest direction relative to current North American geography. The remaining two reconstructions at 980 Ma and 720 Ma both show very similar results, with the Grenville orogeny being the location of possible source rocks. It is

worth noting that these successively younger sedimentary rocks do not contain large proportions of older Trans-Hudson detrital zircon grains. Possible explanations for this distinction range from the exhaustion of a source from which to erode or recycle the zircon grains, to fluvial channel re-routing. Visual provenance analysis for these samples provides support for the configuration of the plate tectonic model in use.

Figure 4.5 Paleogeographic reconstructions created in GPlates using the PalaeoPlates Model, plotting possible provenance sources and paleocurrents throughout the Neoproterozoic. The Grenville orogeny appears to serve as the primary provenance source, with the paleocurrent data supporting this theory. All images are centered around Laurentia, with the remaining continents and the poles places accordingly in successive reconstructions. A global free air gravity map (Bonvalot et al., 2012) is faintly visible behind the continental plates of these models. **(A)** Paleogeographic reconstruction at 1130 Ma, displaying Grenville and Trans-Hudson provenance sources. **(B)** Paleogeographic reconstruction at 980 Ma, displaying dominantly Grenville provenance sources. **(C)** Paleogeographic reconstruction at 720 Ma, displaying dominantly Grenville provenance sources.

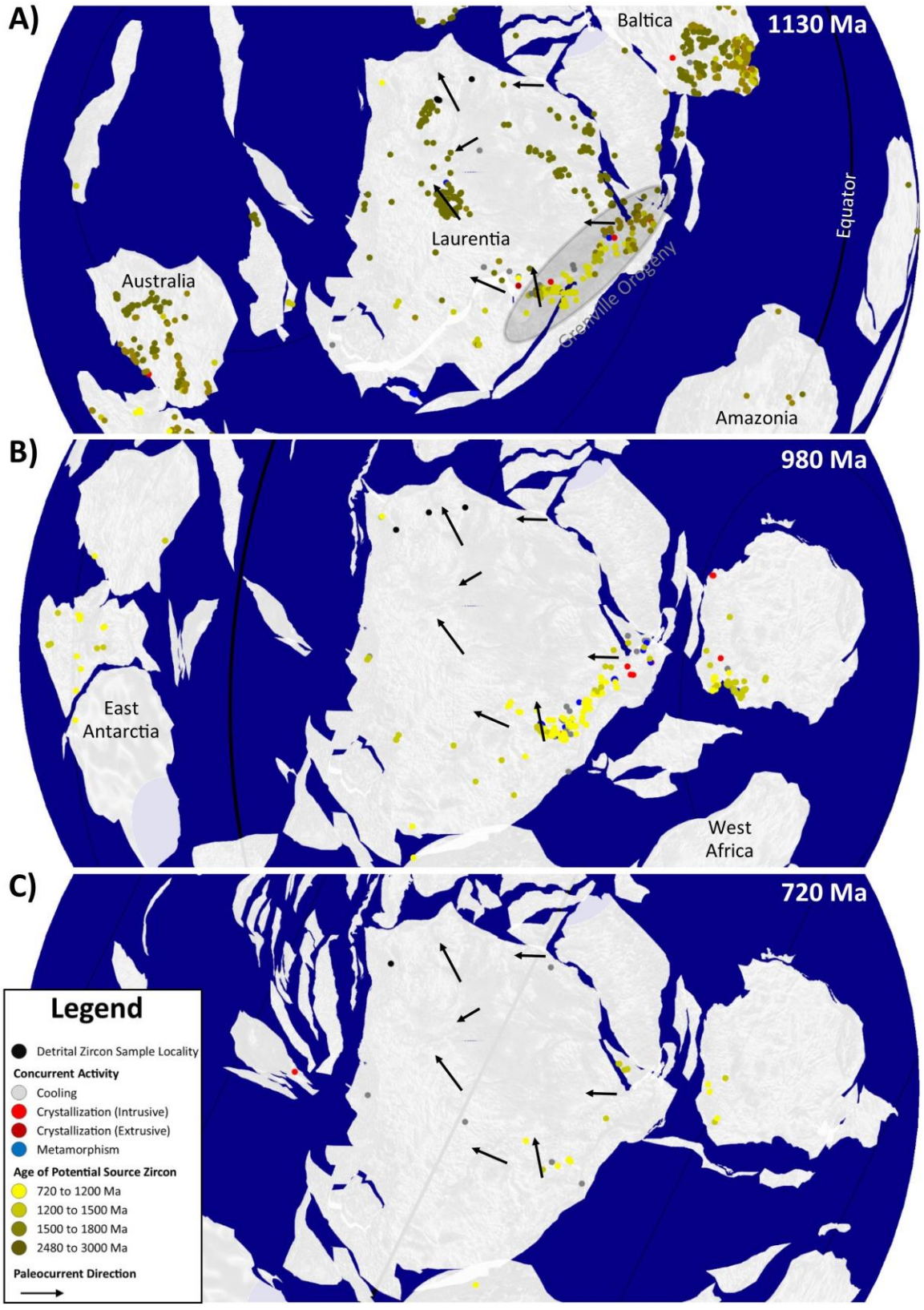


Figure 4.5 Caption on previous page.

An additional example of visual provenance analysis used to validate the plate tectonic models can be observed in the examination of Belt-Purcell Supergroup sediments. The Belt-Purcell Supergroup is a large, thick Mesoproterozoic sedimentary basin that is exposed in southeast British Columbia and parts of the northwestern United States (Harrison et al., 1974; Frost & Winston, 1987; Ross et al., 1992; Lewis et al., 2010). Early attempts to define the source of sediments within this basin by both Harrison et al. (1974) and Frost & Winston (1987) led to the conclusion that parts of the Belt-Purcell Basin were supplied by an unknown source along the western margin of Laurentia. Harrison et al. (1974) suggested a western sediment provenance, originally referred to as "Belt Island," for the siliciclastic material of the Ravalli Group based on thickness and grain size patterns. Frost & Winston (1987) used Sm-Nd isotopic results from samples in the Belt-Purcell basin to suggest possible provenance from a source, possibly the Siberian platform, along the western margin of Laurentia that subsequently rifted away in the late Precambrian. U-Pb and Sm-Nd isotopic analysis of detrital zircon was later conducted by Ross et al. (1992) to further refine possible sources of this western land mass. These authors concluded that a distinct grain age between 1590 Ma and 1600 Ma within the samples could not be explained by a North American source, as there was very little magmatic activity during this time. The distinct lack of magmatic activity on Laurentia is observed to be significantly larger than the 10 Ma gap described by Ross et al. (1992), with the period between 1490 Ma and 1610 Ma referred to as the "North American Magmatic Gap" (NAMG) by Lewis et al. (2010) (Figure 4.6). Ross et al. (1992) addressed this problem by suggesting that Australia is the likely provenance source, as the ages and Sm-Nd isotopic signatures of several Australian cratons match the ages and isotopic signatures of the detrital material analyzed.

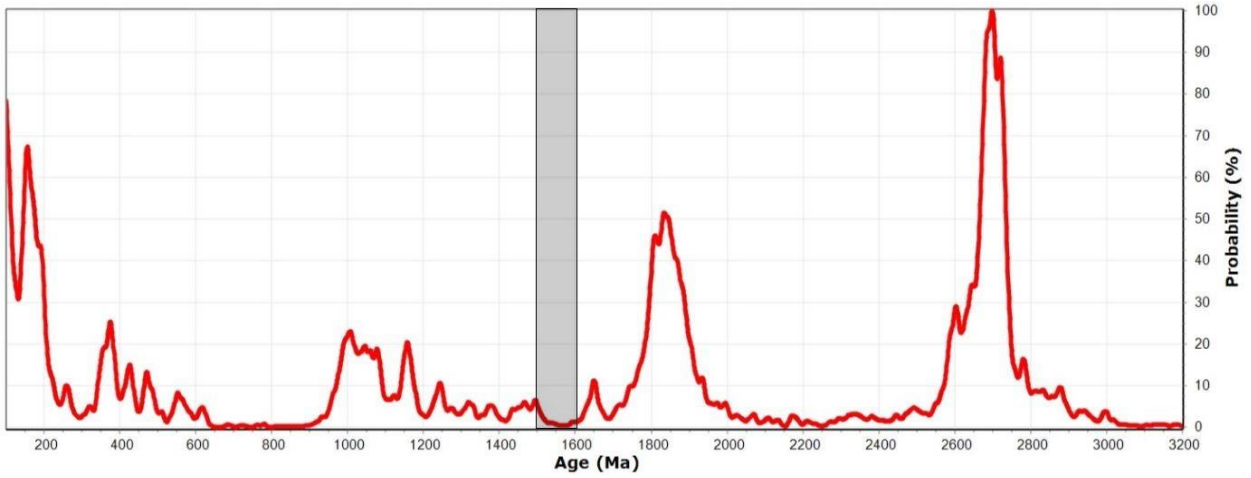


Figure 4.6 Single probability density plot created in FitPDF using the crystallization and metamorphic ages recorded across North America, extracted from DateView. The North American Magmatic Gap (NAMG) occurring from 1490 Ma to 1610 Ma is highlighted by a dark grey box to emphasize the lack of zircon sources.

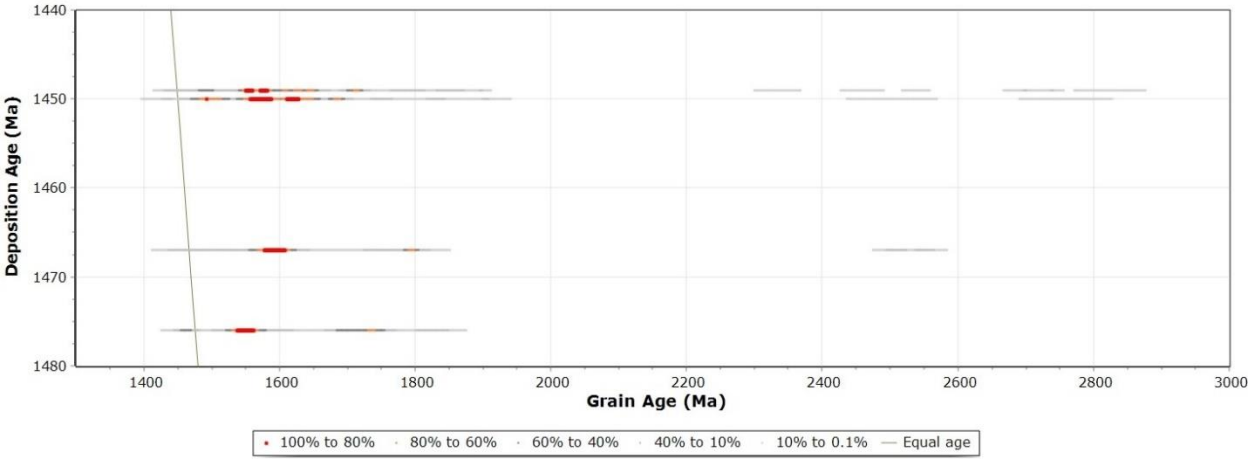


Figure 4.7 Multiple probability density plot of samples in the Belt-Purcell Basin and the Yankee Joe Basin that show high probability peaks during the North American Magmatic Gap (NAMG) between 1490 Ma and 1610 Ma.

Utilizing FitPDF to visualize the PDPs from previously published analyses, high probability grain age peaks are visible during the NAMG (Figure 4.7). This includes samples from both the Belt-Purcell Supergroup in the northwest United States (Ross & Villeneuve, 2003; Lewis et al., 2010) and the Hess Canyon Group in the southwestern United States (Doe et al., 2012). Plotting this data in GPlates using the PalaeoPlates Model, along with possible sources and other supplementary data previously described, a visual model for Australian provenance is provided (Figure 4.8). A paleogeographic reconstruction at 1590 Ma emphasizes the lack of activity on Laurentia, while showing a distinctly different scenario occurring on the Australian cratons during that time. Although the model does not show direct contact between Laurentia and Australia, it implies that they are adjacent to one another. The second paleogeographic reconstruction at 1520 Ma is included to indicate that the PalaeoPlates Model suggests Australia and Laurentia began to rift apart around 1540 Ma. Finally, the paleogeographic model at 1460 Ma shows deposition of the samples in the Belt-Purcell and Hess Canyon groups, along with potential sources as moderate to light yellow points. As displayed in the paleogeographic reconstructions, few sources exist on Laurentia, while there is an abundance of potential source localities on Australia. The series of paleogeographic snapshots in Figure 4.8 illustrates the validity of the PalaeoPlates Model with our current knowledge, supporting the use of this plate tectonic reconstruction for visual analysis of ichnology data.

Figure 4.8 Paleogeographic reconstructions created in GPlates using the PalaeoPlates Model during the Mesoproterozoic. This series of reconstructions illustrates a foreign provenance source for certain sediments in the Belt-Purcell and Yankee Joe basins. All images are centered around Laurentia, with the continents and poles placed accordingly in successive reconstructions. A global free air gravity map (Bonvalot et al., 2012) is faintly visible behind the continental plates of these models. **(A)** Paleogeographic reconstruction at 1590 Ma emphasizing magmatic activity on Australian cratons during this time. **(B)** Paleogeographic reconstruction at 1520 Ma showing the rift between the Central Australian, South Australian, and East Antarctica cratons and Laurentia. **(C)** Paleogeographic reconstruction at 1460 Ma showing deposition of sediment into the Belt-Purcell and Yankee Joe basins, with possible source regions.

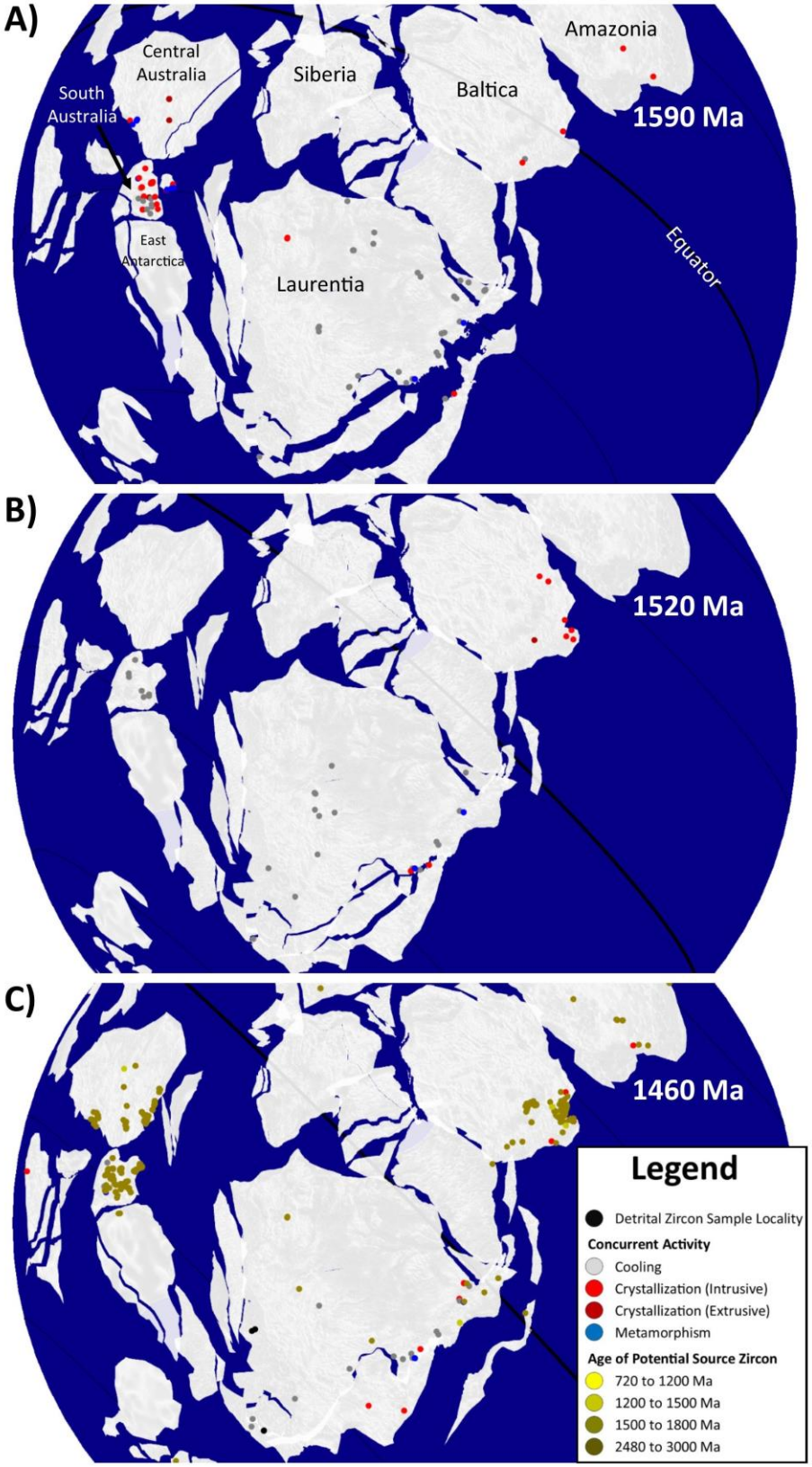


Figure 4.8 Caption on previous page.

4.3.2 Inferring Tectonic Settings from Detrital Zircon

With the reduced cost of analysis and refinement of laser ablation U-Pb measurement techniques, detrital zircon studies produce hundreds of individual detrital zircon grain ages. The many grains analyzed from an individual sample produce unique grain age distributions that can be used to infer the tectonic setting in which a sediment has been deposited (Cawood et al., 2012). These authors described a model in which deposition age of the sample is subtracted from the grain ages, followed by the use of cumulative proportion diagrams to determine the tectonic setting. It is important to note that the use of cumulative proportion curves in this model result in a sensitivity to both sample size, and the accuracy of the deposition age.

The model presented by Cawood et al. (2012) differentiates between three tectonic settings, convergent, collisional, and extensional. Extensional tectonic settings produce lithospheric stretching and cooling, which results in the formation of rifts, passive margins, and cratonic basins (Holland, 2016). Due to the longevity of sediment accumulation and the size of the hinterland from which these basins draw sediment, grain age distributions in these settings are dominated by detrital zircon ages that are significantly older than the sample deposition age (Cawood et al., 2012; Holland, 2016). Considering these observations, the first step of the tectonic setting identification model presented by Cawood et al. (2012) is the identification of extensional settings. Samples that display less than 5% of zircon grains formed within 150 Ma of sample deposition are inferred to represent this type of tectonic setting.

Convergent tectonic settings are associated with supra-subduction zones, where forearc and backarc basins are formed (Cawood et al., 2012). As basins within this type of setting are the result of lithospheric flexing and bending, their longevity and hinterland size are both shorter and smaller in comparison to extensional settings (Holland, 2016). In addition to this, syndepositional igneous activity is common in these settings, resulting in a grain age spectrum that contains a high proportion of detrital zircons with ages close to the time of sample deposition (Cawood et al., 2012). Therefore, the second step of the tectonic setting identification model is the identification of convergent settings. The criteria determined by these authors for step two is the identification of samples that contain over 30% of zircon grain ages that are within 100 Ma of the sample deposition age.

The last tectonic setting identified by this model are collisional settings, where foreland basins are formed (Cawood et al., 2012). These basins, like those of convergent settings, are produced through lithospheric flexing and bending (Holland, 2016). Cawood et al. (2012) observed that foreland basins contain a variable amount of detritus from syn-collisional magmatism, magmatic arc material related to ocean closure, and detritus eroded from the orogenic welt and cratonic foreland. Clearly this variability makes the identification of collisional settings less certain than the previous two, and as a result, all samples not classified by the first two criteria of the model are considered to be collisional.

Applying this model produced by Cawood et al. (2012) to the samples compiled in this study, tectonic settings can be plotted in conjunction with additional datasets. Taking advantage of GPlates, the geodynamic regime can be visualized on Laurentia before, during, and after the Precambrian to Cambrian time period of interest (Figure 4.9). Although the temporal aspect of the data was dramatically altered to display all samples deposited during the vast time periods defined in Figure 4.9, these paleogeographic reconstructions illustrate important trends. The paleogeographic reconstruction summarizing the Paleo- and Mesoproterozoic displays all three tectonic settings across Laurentia. This reconstruction depicts the complicated nature of the geodynamic regime on Laurentia at this time, representative of accretion and orogenic events. The paleogeographic reconstruction during the Neoproterozoic and Cambrian illustrates a more stable geodynamic regime across Laurentia. During this time period, the western and northern margins of Laurentia relative to present-day geography are dominated by passive margins. The last paleogeographic reconstruction represents the remaining Paleozoic era, which displays a similar geodynamic environment of passive margins along the western and northern margins of Laurentia, with an increasing amount of detrital zircon samples inferring collisional environments. In addition to this, the production of arcs can be seen, which eventually collide with the western margin of Laurentia during the Mesozoic and Cenozoic.

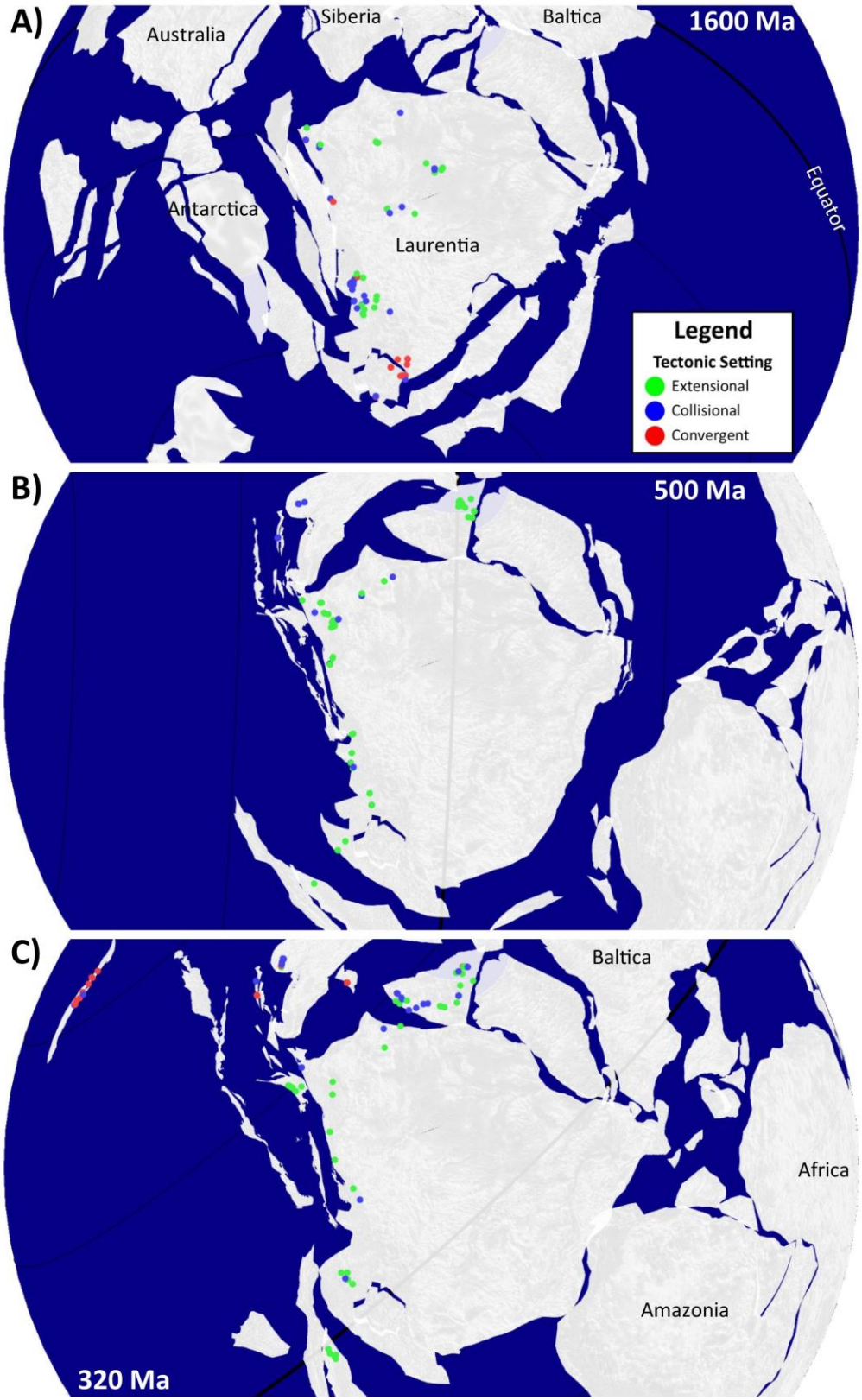


Figure 4.9 Caption on following page.

Figure 4.9 Paleogeographic reconstructions created in GPlates using the PalaeoPlates Model, plotting the inferred tectonic setting for each sample based on the method described by Cawood et al. (2012). All images are centered around Laurentia, with the continents and poles placed accordingly in successive reconstructions. A global free air gravity map (Bonvalot et al., 2012) is faintly visible behind the continental plates of these models. **(A)** Paleogeographic reconstruction at 1600 Ma, displaying samples deposited throughout the Paleo- and Mesoproterozoic. **(B)** Paleogeographic reconstruction at 500 Ma, displaying samples deposited throughout the Neoproterozoic and Cambrian. **(C)** Paleogeographic reconstruction at 320 Ma, displaying all samples deposited throughout the Ordovician to Permian.

4.4 Discussion

An effective method for visualizing many detrital zircon PDPs simultaneously has been demonstrated with the use of FitPDF. This software allows users to overcome the limitations encountered using the traditional method of stacking PDPs. The ability to view hundreds of PDPs simultaneously is a feature that will be required with access to the ever-increasing size of detrital zircon datasets. Pattern recognition from the plots produced with FitPDF may also prove to be useful when studying new datasets.

It has been demonstrated that visual provenance analysis can be used to support the validity of plate tectonic models. This type of analysis does not provide indisputable proof that the plate model is robust, rather it lends support for its use. The visualization of provenance analysis during the Neoproterozoic across Laurentian (e.g., Rainbird et al., 1992, 1997) and the provenance analysis of zircons with a distinct NAMG signature (e.g., Ross et al., 1992) illustrate the plate model supporting these theories. This use of visual provenance analysis is not meant to displace traditional methods of detailed provenance analysis. The visual analysis described in this section is only concerned with the provenance of zircon grains present in samples with a probability greater than 80% as a starting point for understanding overall provenance trends. Whereas traditional methods describe sources for all grains present in a sample and consider additional issues such as sediment recycling and biases present in the detrital zircon record.

The second function that detrital zircon data provides to the visual analysis of paleontology data is the inference of tectonic settings. Using the model created by Cawood et al. (2012), the dataset compiled provides detail regarding the geodynamic environment on Laurentia during the time period of interest. The results of this analysis suggest stability across the western

and northern margins of Laurentia relative to the present North American geography during the Ediacaran to Cambrian. Considering this observation, fossil preservation across northern and western Laurentia during this time is favorable and clustering of data across these regions is expected. Further, observing the geodynamic settings across Laurentia both before the Precambrian to Cambrian transition, and after, provide an indication of the geodynamic evolution.

To this point the detrital zircon dataset has only been used for visual analysis. Methods to produce a more objective, quantified production of patterns within detrital zircon datasets has been proposed by Vermeesch (2013), where Multidimensional Scaling (MDS) has been utilized. This work has been supplemented with an analysis of dissimilarity methods compatible with MDS (Vermeesch, 2018), and the publication of a template for conducting this technique in the statistical programming environment `R` (Vermeesch et al., 2016). Supplementary data (e.g., mineralogical, geochemical) can be included with the detrital zircon grain age PDPs through the use of Generalised Procrustes Analysis (GPA) and Individual Differences Scaling (INDSCAL), providing a more robust outcome than MDS on its own (Vermeesch & Garzanti, 2015). In addition to quantifying and visualizing the differences between samples, MDS can be used to investigate the evolution of geodynamic environments in which samples have been deposited (Spencer & Kirkland, 2016). Although this work is encouraging, the examples used with the described methods have been limited to relatively small sample sizes. Further work investigating the reliability of these techniques when used with larger datasets, covering larger geographic regions would provide useful details regarding possible limitations in their application.

5 Statistical Analysis of Detrital Zircon Datasets

The compiled detrital zircon dataset has been shown to be useful in testing plate tectonic models by way of visual provenance analysis, and the visualization of tectonic evolution. Visualization is undoubtedly useful; however, a more quantifiable technique may be desirable. One such technique, Multidimensional Scaling (MDS), is a statistical method that can be used to conduct provenance analysis (e.g., Vermeesch, 2013) and the investigation of orogenic evolution (e.g., Spencer & Kirkland, 2016). Examples where MDS has been applied to detrital zircon datasets have been encouraging, although the size and geographic extent of examples has been relatively small and restricted.

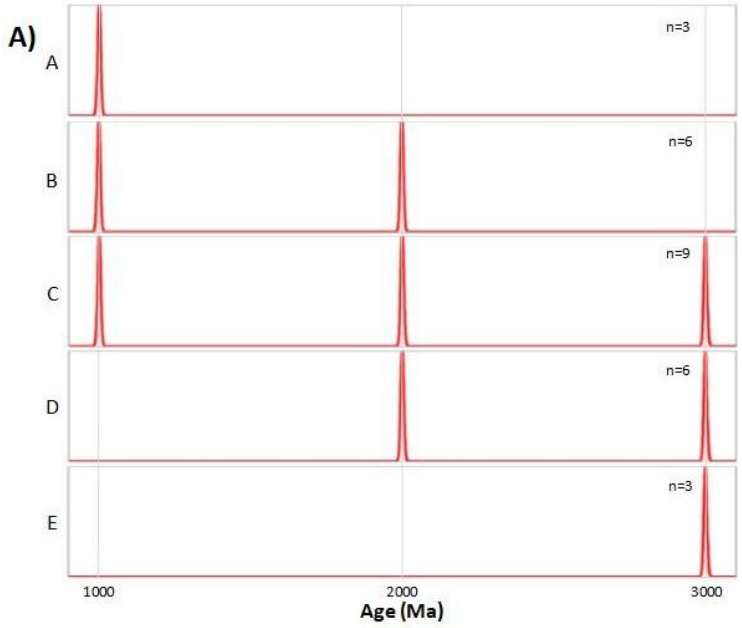
Previous chapters in this thesis have emphasized the utility of data mining. As statistical techniques are fundamental to data mining (Han et al., 2012), the compiled detrital zircon dataset from Laurentia can be used to investigate how MDS works in association with detrital zircon analysis and therefore describe the limitations of this technique. To illustrate possible issues with this technique several factors are analyzed, including the application of dissimilarity measures previously disregarded (e.g., Vermeesch, 2018), the algorithms used to conduct MDS, the number of dimensions in which ordination is conducted, and the effect of sample size on the goodness of fit (Kruskal, 1964). Defining these limitations will provide guidance regarding how this technique can be used to supplement the visual analysis techniques for testing plate tectonic models and investigating geodynamic evolution.

5.1 Limitations of using MDS in Association with Detrital Zircon

5.1.1 Dissimilarity Measures

A thorough review of dissimilarity measures commonly used for MDS with detrital zircon analysis is provided by Vermeesch (2018). This author described several dissimilarity measures including: Chi-squared statistic, Kolmogorov-Smirnov statistic, Kuiper statistic, Cramér-von-Mises statistic, Likeness parameter, Cross-correlation coefficient, and the Sircombe-Hazelton L2-norm. Observing published examples of MDS use with detrital zircon data (e.g., Vermeesch, 2013; Spencer & Kirkland, 2016), it is clear that there is a bias towards the use of the Kolmogorov-Smirnov (K-S) statistic. Vermeesch (2018) explained that the preference for

this dissimilarity measure stems largely from the fact that data does not need to be pre-treated, and subtle differences between age distributions can be identified.



B)

	A	B	C	D	E
A	0	0.50	0.66	1.0	1.0
B	0.50	0	0.33	0.50	1.0
C	0.66	0.33	0	0.33	0.66
D	1.0	0.50	0.33	0	0.50
E	1.0	1.0	0.66	0.50	0

Figure 5.1 (A) Probability density plots (PDPs) created from the synthetic detrital zircon grain ages using FitPDF. All plots are normalized to 100%. The sample names are represented by letters on the left, and the number of samples (n) is listed on each PDP. **(B)** Kolmogorov-Smirnov dissimilarity values calculated between all five samples in R using the workflow described by Vermeesch (2013), where values range from 0 (exactly the same) to 1 (nothing in common).

To illustrate the effects of the Kolmogorov-Smirnov (K-S) statistic, a simple synthetic detrital zircon dataset has been created. The synthetic dataset contains varying amounts of hypothetical zircon grains of approximately 1000 Ma, 2000 Ma, and 3000 Ma. Utilizing code provided by Vermeesch (2013) to calculate K-S dissimilarities in R, the comparison of each synthetic grain age distribution is observed in Figure 5.1. A close inspection of this figure clarifies how dissimilarity values summarize pair-wise PDPs. Take for example the resulting K-S dissimilarity between samples B & C, and C & D. Both comparisons result in a K-S dissimilarity value of 0.33. Unfamiliarity with dissimilarity values may lead to the inference that samples B and D are the same, with approximately one third of their hypothetical zircon grains being dissimilar from those in sample C. Observing a dissimilarity value of 0.5 between samples

B & D in Figure 5.1 B, the initial inference is disproved. Visualization of these detrital zircon grain distributions in Figure 5.1 A provides assistance when trying to understand the results of this dissimilarity measure.

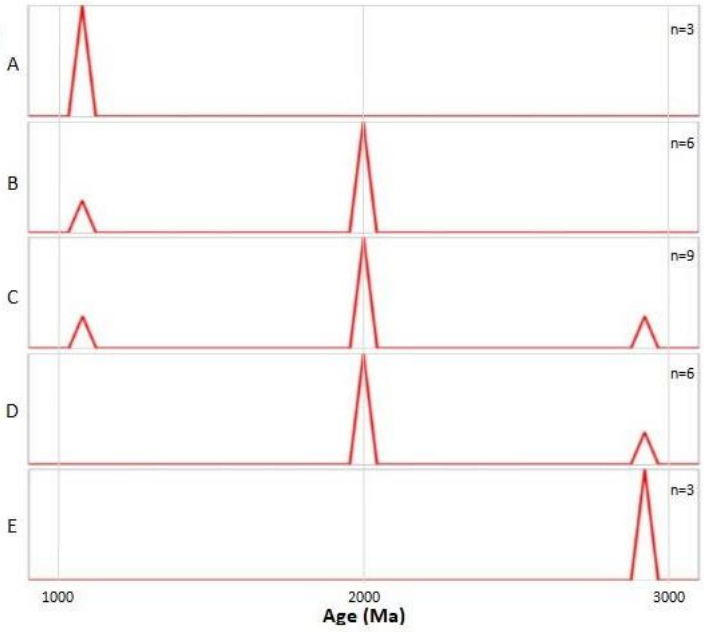
To provide clarity and confidence in the methods selected for MDS, the differences and limitations between the non-parametric and parametric dissimilarity measures must be understood. The use of parametric dissimilarity measures is discussed by Vermeesch (2018), with the observation of two important limitations regarding their use with detrital zircon provenance studies. The first of which is the influence of sample size, an issue that is recognized here but not studied further. The second issue is that parametric techniques require binning which often results in the loss of detail. With the Laurentian dataset compiled as part of this thesis, the many samples with complex PDPs may produce MDS ordinations with unintelligible patterns. As an ordination is simply a technique in which data is reduced to a simplified form, perhaps further simplification through the use of parametric dissimilarity values will produce results that are easier to interpret.

To investigate this question, the Bray-Curtis dissimilarity method has been selected as the parametric dissimilarity measure. This dissimilarity measure is commonly applied in ecology, as it is well suited for abundance data, reaching a maximum value of 1.0 when objects have nothing in common and ignoring dual absences (Quinn & Keough, 2002). Detrital zircon data is analogous to ecological abundance data, through the creation of grain age bins to categorize the zircon grains. These bins define a pseudo-species, thereby counting any detrital zircon grain falling within the bin range as a pseudo-species occurrence. Further, this dissimilarity method allows for the inclusion of associated sigma values with each reported U-Pb age obtained for the detrital zircon grains, a factor not accounted for in the K-S dissimilarity measure.

Unlike non-parametric dissimilarity tests, parametric tests require data pre-treatment in the form of defining both the placement and size of age bins (Vermeesch, 2018). This is a notable disadvantage, as this process is not a simple task, requiring the use of multiple computer programs (e.g., FitPDF, Excel, Access) and knowledge of basic SQL queries. Age bins are defined via FitPDF (Eglington, 2018b), followed by exporting the data and reformatting it for use within R. The result of binning with the synthetic detrital zircon dataset are observed in both the PDPs and the Bray-Curtis dissimilarity values (Figure 5.2). The placement and size of bins

will have a large influence on the amount of detail lost through the use of parametric tests (Sircombe, 2004; Vermeesch, 2018), which is clearly observed in the PDPs of Figure 5.2 A. Samples B, C, and D all show a reduction in peak size for the grains occurring at 1000 Ma and 3000 Ma. This example used 50 age bins, with a range beginning at 800 Ma and ending at 3200 Ma. As a result, each bin represented 48 Ma and the range for these bins in turn determined their placement. These criteria established a division in bins at exactly 2000 Ma, leading to the emphasis of this grain age peak over the other two. The results of the Bray-Curtis dissimilarity illustrate the loss of detail as a product of bin size and placement when compared to the results of the K-S dissimilarity in Figure 5.1.

A)



B)

	A	B	C	D	E
A	0	0.748	0.776	1.0	1.0
B	0.748	0	0.100	0.223	1.0
C	0.776	0.100	0	0.100	0.776
D	1.0	0.223	0.100	0	0.748
E	1.0	1.0	0.776	0.748	0

Figure 5.2 (A) Probability density plots (PDPs) created from the synthetic detrital zircon grain ages using FitPDF, with the bin size altered to 50 Ma. All plots are normalized to 100%. The sample names are represented by letters on the left, and the number of samples (n) is listed on each PDP. **(B)** Bray-Curtis dissimilarity values calculated between all five samples in R using the vegdist function within the vegan package.

5.1.2 MDS Algorithms

To conduct MDS, the statistical programming software R (R Core Team, 2018) is used. As this is open source software, individuals are free to create packages which contain tools for conducting statistical analysis. To produce MDS ordinations, two such packages are commonly

utilized. The first of these is the MASS package (Venables & Ripley, 2002), which allows users to conduct MDS through the use of their isoMDS function. This has been the algorithm of choice in previously published examples where MDS is used with detrital zircon data (e.g., Vermeesch, 2013; Spencer & Kirkland, 2016). The second package used for MDS ordination is vegan (Oksanen et al., 2018), with their algorithm referred to as the monoMDS function.

To properly evaluate MDS algorithms, the quality of ordination must be quantified. This quantification is described as the stress value, from which the analyst can determine a corresponding "goodness of fit" (g.o.f) resulting from the ordination, as observed in Figure 5.3 (Kruskal, 1964). Stress values are reported with both MDS functions in `R`, therefore the ordination of identical datasets between the two algorithms can be compared based on these values.

g.o.f	Poor	Fair	Good	Excellent	Perfect
S	0.2	0.1	0.05	0.025	0

Figure 5.3 A guide to interpreting stress values (S) from Kruskal (1964) in determining the goodness of fit (g.o.f) between dissimilarities and distances of MDS ordinations (modified from Vermeesch, 2013). Stress values may also be presented in percentages.

Two datasets have been selected to analyze the results obtained from the MDS algorithms being investigated. Each function within `R` is applied to the individual datasets multiple times to emphasize the characteristics unique to isoMDS and monoMDS. The resulting stress values are recorded in Table 5.1, where it can be observed that the monoMDS function provides variable results, while the isoMDS function does not. This important difference is explained by the initial configuration used within each MDS algorithm. The isoMDS function defines the same initial starting configuration every time it is used, therefore providing the same stress values after every iteration. Contrarily, the monoMDS function creates a random starting configuration for every iteration, producing a unique stress value each time. It is observed that MDS is capable of producing an ordination that appears to be the optimal solution, termed a local minima, when in fact it is not truly the best solution (Holland, 2008). Therefore, running multiple trials with the

monoMDS function decreases the likelihood of obtaining a local minima, as recommended by Oksanen et al. (2018). Considering these results, the monoMDS function provides an algorithm with a preferable starting configuration and therefore producing a more robust ordination.

Table 5.1 Stress values reported after successive iterations of running a dataset through the isoMDS and monoMDS functions in R. The quality of ordination is determined via the stress value reported after running the MDS algorithm. Cells highlighted in grey represent the lowest stress value and therefore the best result.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
isoMDS								
Verm	2.6178	2.6178	2.6178	2.6178	2.6178	2.6178	2.6178	2.6178
Laurentian	18.2381	18.2381	18.2381	18.2381	18.2381	18.2381	18.2381	18.2381
monoMDS								
Verm	4.0281	2.5570	2.6482	2.5520	4.0278	2.8701	2.6483	3.4957
Laurentian	12.9177	12.9210	12.2981	12.6200	13.2324	12.3508	12.6072	12.2325

5.1.3 K Value Selection

Multidimensional scaling is unique from other ordination methods in that the number of dimensions (K) resulting from the ordination is determined before analysis (Holland, 2008). As a result, there are no hidden axes of variation, as is the case with correspondence analysis and principal components analysis. Holland (2008) noted that the selection of too few dimensions will result in the ordination displaying multiple axes of variation on a single plotted dimension. Regardless of this observation, most MDS analyses select two dimensions as the number of axes due to the ease of plotting and interpretation (Vermeesch & Garzanti, 2015).

Although the selection of two dimensions for MDS ordination of detrital zircon analysis is the standard for published examples, it is possible to visually determine the appropriate number of dimensions in which to conduct MDS. This can be accomplished by conducting MDS with a progressively higher number of dimensions and plotting the resulting stress values in a scree plot, as observed in Figure 5.4 (Holland, 2008). This author suggested that the dimension after which stress does not significantly drop will be the appropriate number of dimensions with which to conduct MDS. The scree plot in Figure 5.4 would suggest that two dimensions is appropriate, despite the anomalous peak observed with four dimensions.

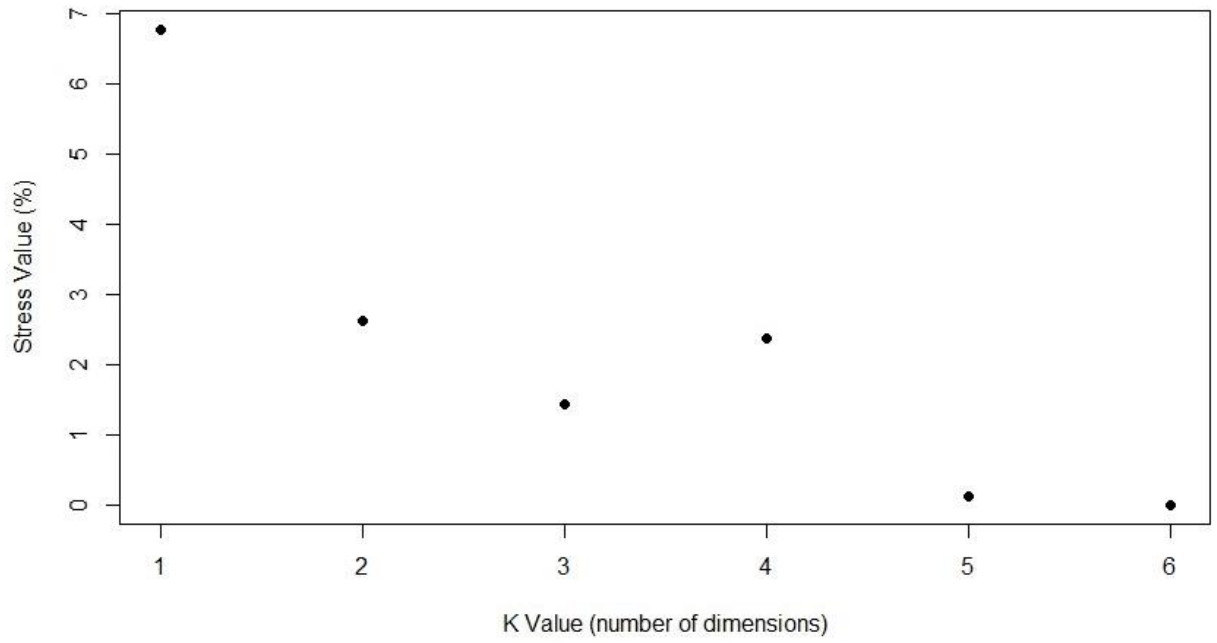


Figure 5.4 Scree plot displaying the reduction of stress as a result of MDS ordination with K dimensions. Conducted using the Vermeesch (2013) Chinese dataset with the isoMDS function in R.

5.1.4 Number of Samples Analyzed

The results presented in Table 5.1 display a large disparity in the stress values between the Laurentian and Verm datasets. The obvious distinction between these two datasets is the sample size, with the Verm dataset consisting of 15 samples, whereas the Laurentian dataset consists of 140 samples. It would seem logical to suggest that this difference is controlling the vastly different stress values produced between datasets. To further investigate the influence of sample size the compiled detrital zircon dataset has been split into various sample sizes ranging from 10 to 400. The isoMDS function with two dimensions is then used to produce MDS ordinations from each dataset. The resulting stress values are then recorded and plotted to visualize the effect of sample size on the goodness of fit (Figure 5.5). The stress values in Figure 5.5 show an increase with sample size up to 50, at which point the stress values begin to plateau.

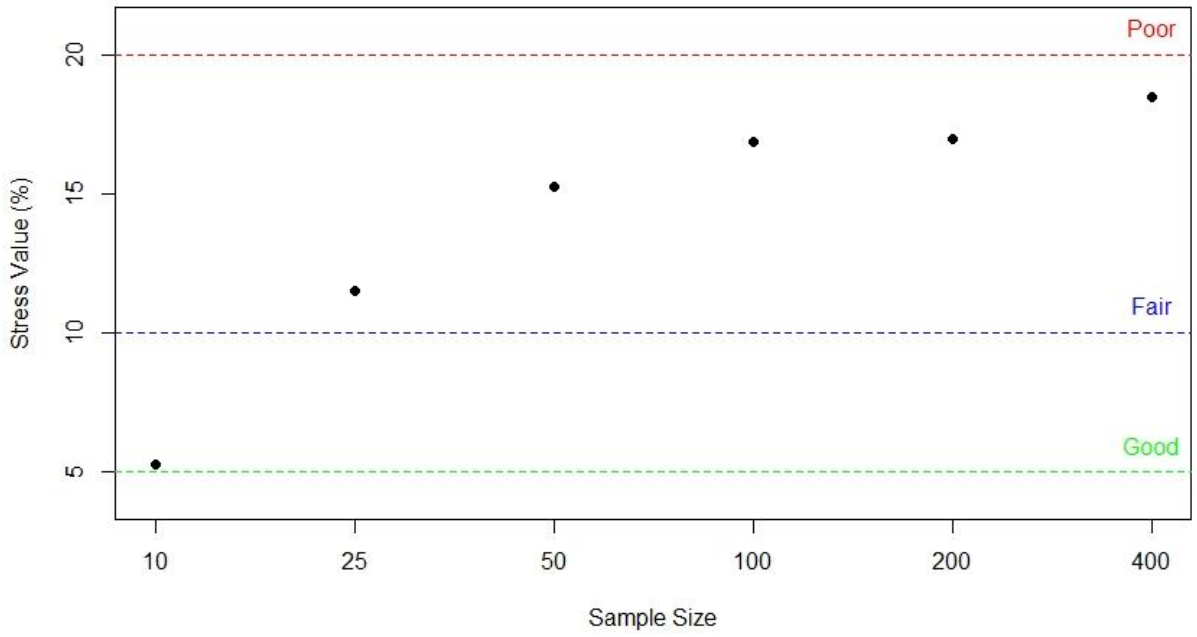


Figure 5.5 Scree plot displaying the effect of increasing sample size on the resulting stress value reported with the MDS ordination, where the number of dimensions selected is two. Coloured dashed lines indicating the divisions between goodness of fit as defined by Kruskal (1964) are included.

The positive correlation between sample size and stress values in Figure 5.5 provides the false impression that these values are related. Rather, this correlation is produced by conducting MDS on an increasingly complex dataset from across North America without changing the number of dimensions in which to produce the ordination. Holland (2008) observed that it is possible to over simplify an MDS ordination with the selection of too few axes, whereby additional axes of variation are forced onto one axis. With an increasing number of samples, this dataset is including samples from disparate geographic regions across North America which in turn increases the complexity as these represent different geological processes. Therefore, the sample size and stress value are independent of each other, and entirely dependent on the nature of the data itself. A dataset may be extremely large, but it may only represent a very simple relationship that is sufficiently displayed on two axes.

5.2 Comparing MDS Techniques on a Large Scale

Taking into consideration the strengths and weaknesses of dissimilarity measures used, the differences in MDS algorithms, and the number of dimensions selected for use, this study will compare different techniques to analyze the results of MDS ordination on larger datasets than described in the literature. Detrital zircon data subjected to MDS as a tool for provenance analysis aims to identify clusters of samples, from which it is assumed that these clusters share a provenance source (e.g., Vermeesch, 2013). MDS can also be used to display the transition from different types of geodynamic settings (e.g., Spencer & Kirkland, 2016). This technique is not guaranteed to display the evolution of geodynamic settings, as interbasin communication (e.g., sediment recycling) can produce indecipherable signatures in an MDS ordination. Considering these possibilities with the ordination of detrital zircon data, the effectiveness of previously published methods (i.e. using the K-S statistic with two dimensions) in the production of clusters or trends with a larger dataset are analyzed. In addition to this, a second method using parametric dissimilarities is used to investigate possible advantages of non-traditional techniques.

5.2.1 Methods

To conduct the MDS ordination trials, a portion of the compiled detrital zircon data from across the northern and western margins of North America were selected (Figure 5.6). Due to the size of the dataset and the unfamiliar process involving data pre-treatment and analysis, only one third of the compiled samples were used in these tests. With this dataset, two variations of MDS ordination are conducted to an effort to gain a better understanding of this statistical technique.

The first variation of MDS uses methods described in previously published detrital zircon studies, where the isoMDS function in `R` is used to produce an ordination from a matrix containing Kolmogorov-Smirnov dissimilarities. In contrast, the second variation will use the monoMDS function to produce an ordination from a matrix containing Bray-Curtis dissimilarities. As previously observed, this technique requires data pre-treatment in the form of creating bins into which the age of every detrital zircon grain analysis will be sorted. The selection of bin size was based on results plotted within FitPDF (Figure 5.7). Visual analysis of

the bin sizes tested led to the inference that the selection of 50 age bins simplifies the data, while still retaining the overall pattern observed without binning. Finally, both variations will define a K-value of two for ordination, as this is the observed standard and easily interpreted.

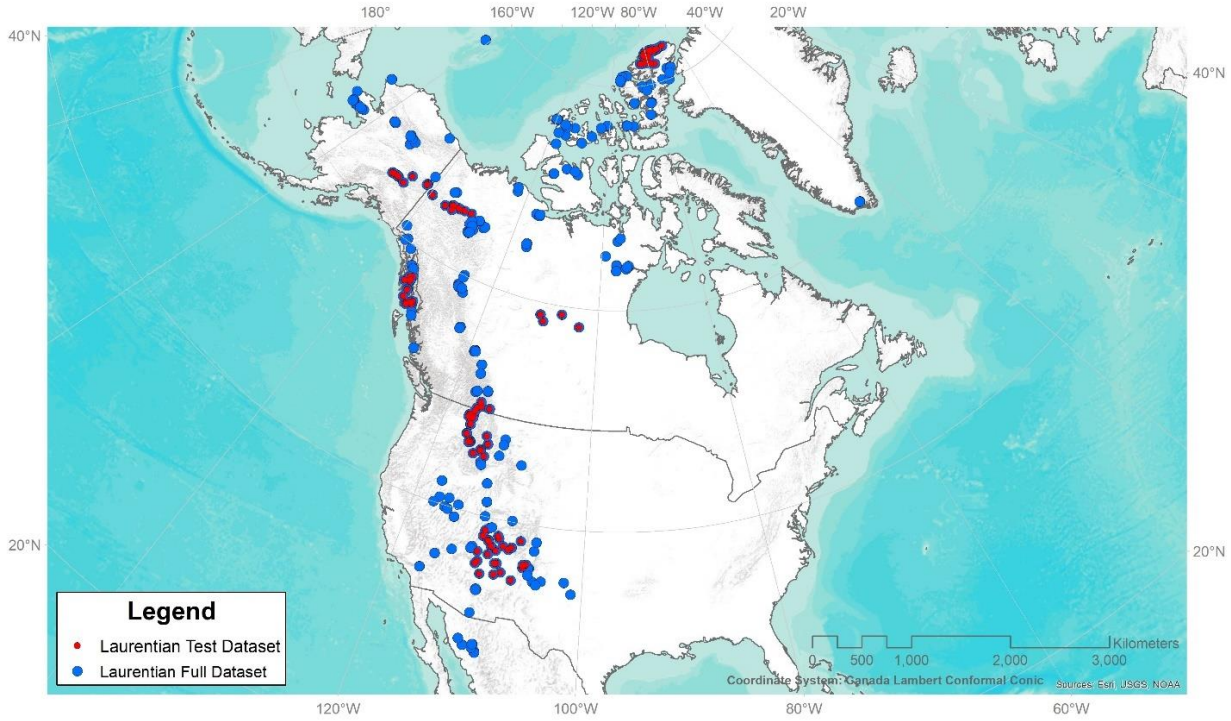


Figure 5.6 Map of North America displaying the location of all samples compiled in the Laurentian dataset. The data points represented by red points are those that have been extracted for use with MDS, broadly representing all geographic regions of the dataset.

Following MDS ordination, cluster analysis will be conducted within `R` to supplement the identification of potential patterns. K-means clustering, a non-hierarchical clustering method (Legendre & Legendre, 2012) has been selected for further analysis. This is a divisive method of clustering, in which the data begins as a single group and is split into subgroups until the user defined number of clusters has been reached (Borcard et al., 2011). These authors explained that K-means clustering runs iteratively with the objective of producing clusters that contain the smallest within-groups sums-of-squares. Further, it is observed that K-means clustering identifies high-density regions within the data.

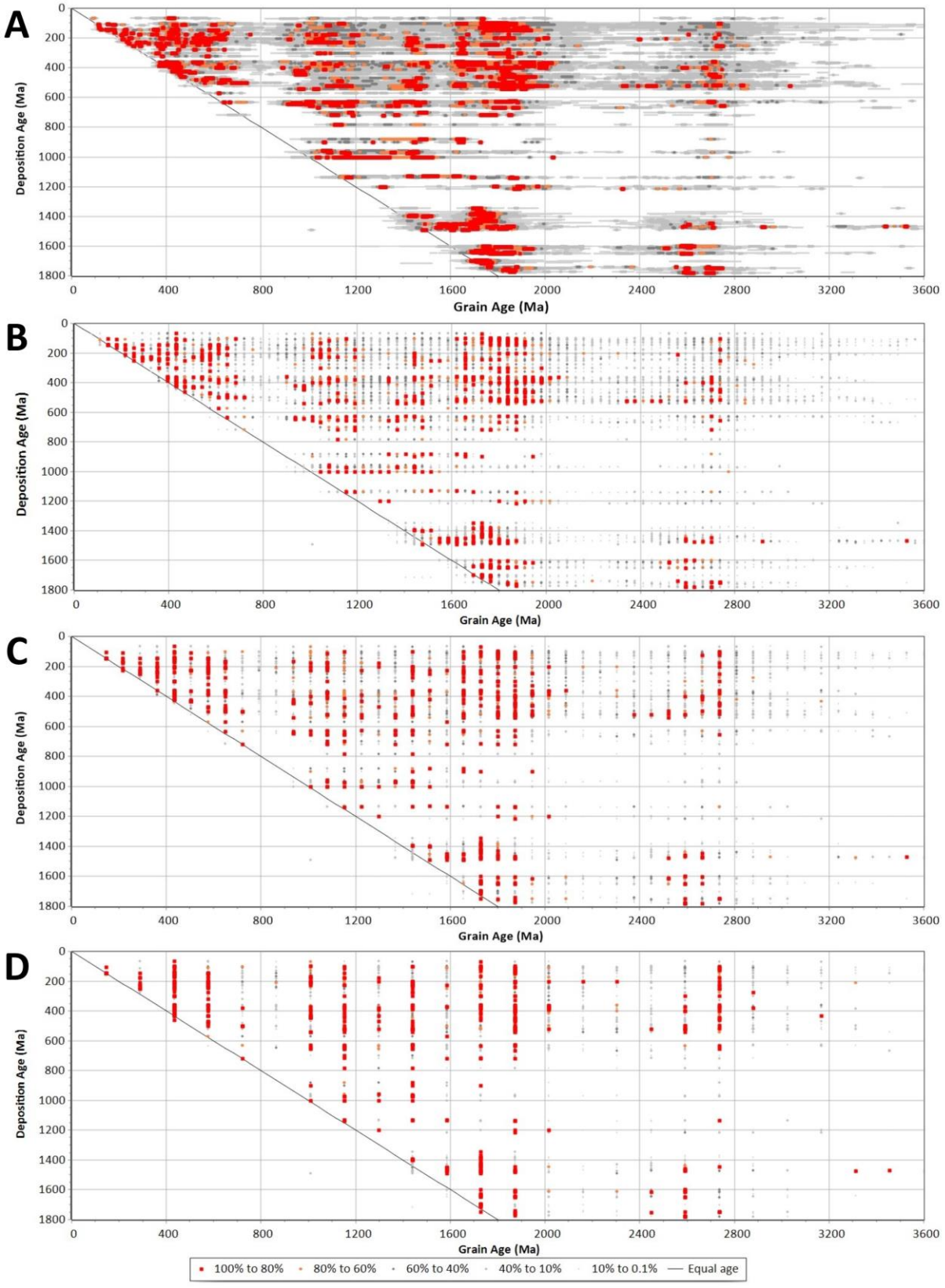


Figure 5.7 Caption on following page.

Figure 5.7 Multiple probability density plots (PDPs) created in FitPDF from the Laurentian dataset used. **(A)** Multiple PDPs with no bin size restriction. **(B)** Multiple PDPs with bin size set to 100. **(C)** Multiple PDPs with bin size set to 50. **(D)** Multiple PDPs with bin size set to 25.

Following the production of ordination and clusters, supplementary data will be required to assist in pattern recognition. Sample metadata to be appended through the use of database queries includes all grain ages contained within a sample, zones of high probability determined in FitPDF, inferred tectonic setting, and sample deposition age. Although R does provide excellent graphics packages, these can become unwieldy with the desired level of variables displayed in the symbology. Considering this, analysis was conducted in ioGAS, as this software supports complex symbology required for ease of interpretation. A detailed outline regarding the data preprocessing, all datasets used in this section, and the code used in R studio are available as supplementary material.

5.2.2 Results

5.2.2.1 MDS Ordination Using the Kolmogorov-Smirnov Statistic

The Kolmogorov-Smirnov statistic as a dissimilarity measure used with the isoMDS algorithm provides baseline results using methods which have been considered acceptable in previous publications. Applying these parameters to the Laurentian dataset, the resulting stress of this ordination is 18.24%, resulting in a fit that is described as fair, approaching poor. Further, this ordination produces a trend, rather than distinct clusters of samples as observed in the work conducted by Vermeesch (2013). To aid interpretation and provide an explanation for these results, several different sets of symbology will be applied to the MDS plot.

The first MDS plot created (Figure 5.8 A) emphasizes the dominant grain ages within individual samples. The symbology applied to this plot displays what is interpreted to be the primary influence on the trend produced, high probability grain age peaks. Initially, only a colour ramp was applied to the high probability (i.e. >80%) grain age peaks within each sample which displayed strictly unimodal grain age distributions. To account for this issue, four size divisions were imposed on the points within the plot to further visualize grain ages characteristic of the major orogenic events on Laurentia (i.e. the Cordilleran, Appalachian, Grenville, Trans-Hudson, and Superior orogenic events). Admittedly, should a sample display a bimodal

distribution within a single defined age spectrum (e.g., two peaks, one at 200 Ma, the other at 400 Ma) this would still appear as a unimodal distribution with one small blue point on the MDS plot based on the defined symbology criteria. Despite the reduction in details observed, a high probability grain age trend is apparent, with samples containing the oldest high probability peaks plotted in the bottom left quadrant. This is followed by samples with progressively younger grain ages plotting towards positive dimension 1 values, and then a final shift towards positive dimension 2 values. The inclusion of a multiple probability density plot displaying only high probability grain ages within the samples (Figure 5.8 B) provides further detail as to how the defined criteria for symbology are displayed within the samples.

Figure 5.8 (A) MDS ordination of the test dataset from Laurentia with Kolmogorov-Smirnov dissimilarities. Results are produced from the isoMDS function within the MASS package in R. These results are plotted in ioGAS, with criteria for symbology as follows: Only grain age peaks above 80% probability are displayed, a blue (young) to red (old) colour ramp is applied to the peaks, and a symbol size division is applied, promoting the visualization of multiple high probability grain age peaks within a single sample. **(B)** A multiple probability density plot created in ioGAS from the samples used to produce the MDS ordination, with only high probability (>80%) grain age peaks displayed.

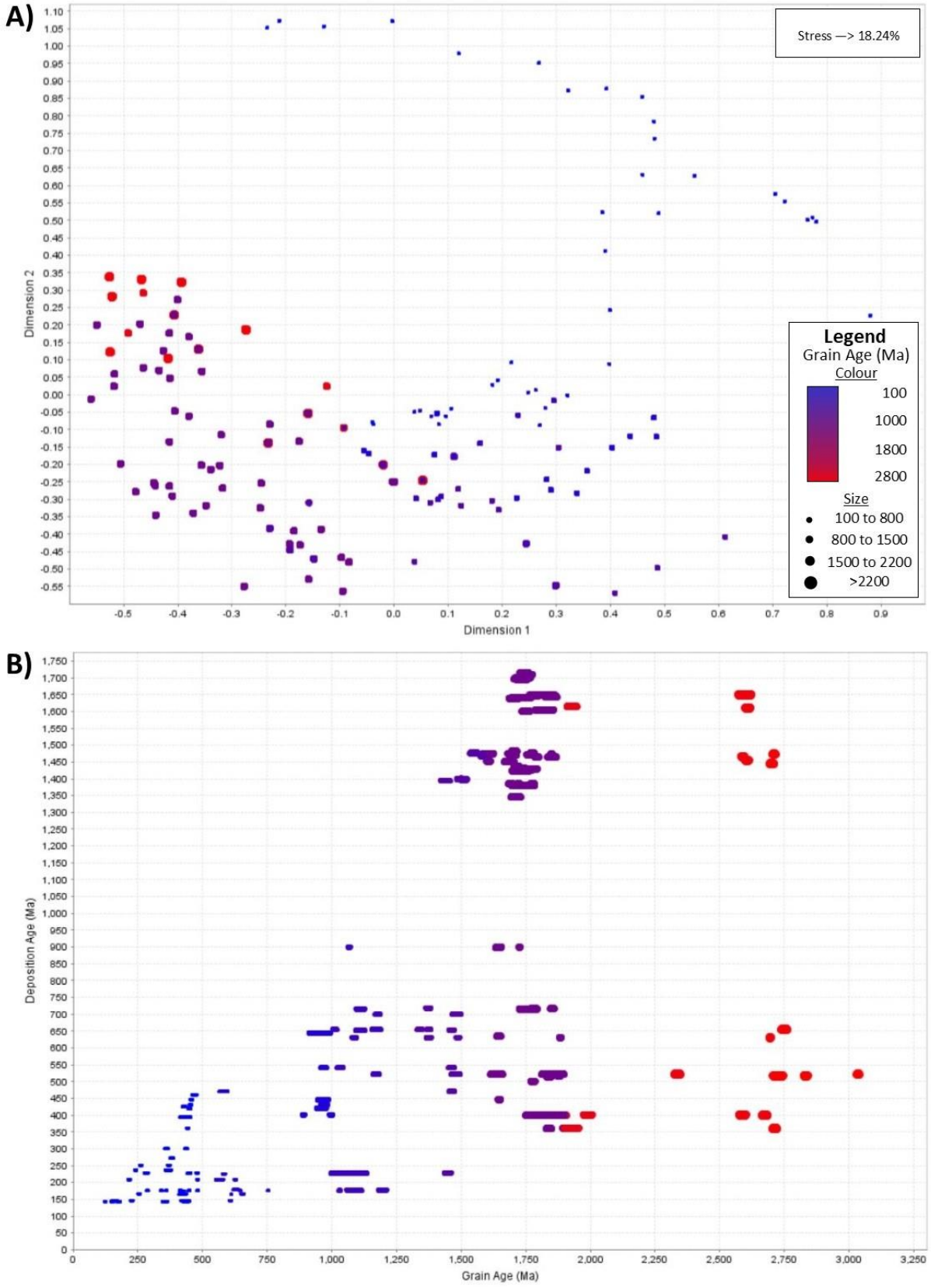


Figure 5.8 Caption on previous page.

An additional MDS plot created (Figure 5.9 A) displays K-means clusters produced from the kmeans function in the stats package (R core team, 2018) in R. An arbitrary number of eight clusters was created simply to divide the samples into a manageable number for further analysis. A multiple probability density plot was included (Figure 5.9 B) with this MDS plot to supplement the analysis of resulting clusters. The K-means clusters defined from this ordination provide additional support to the high probability grain age trend observed in Figure 5.8. The MDS plot of Figure 5.9 displays K-means clusters in the bottom left quadrant that are defined by samples with the oldest high probability grain age peaks. The previously described trend continues with clusters along increasingly positive dimension 1 values containing samples with progressively younger high probability peaks. Finally, the trend is completed by moving towards increasingly positive dimension 2 values on the MDS plot, again displaying samples with progressively younger high probability grain age peaks.

To further investigate the characteristics of the K-means clusters, all samples from each cluster were imported into FitPDF and single probability density plots were produced (Figure 5.10). With a summary of each cluster, these stacked single PDPs assist in describing the ordination. As it was previously interpreted, the dominant grain age distributions within each sample determine the overall trend of the ordination. Clusters 7 and 8 represent the samples dominated by the oldest high probability grain age peaks ranging from 1600 Ma to 2800 Ma. The next three K-means clusters of the grain age trend are clusters 1, 3, and 6. These clusters show distinct and dissimilar PDPs. Most notably, of these three clusters, number 6 appears to be an amalgamation of samples with grain ages across the entire grain age spectrum. This cluster is located around the origin of the graph which is the result of these samples being similar to the other clusters, but they are dissimilar enough to plot in a different cluster area. Cluster 6 may be representative of sediment recycling, with the origin on an MDS plot being an area to look for samples of this type. The other two clusters in this part of the MDS ordination both represent samples with grain ages ranging from 900 Ma to 1600 Ma. The final three clusters 2, 4, and 5 all display dominantly unimodal distributions with young grain ages decreasing from 600 Ma to 100 Ma.

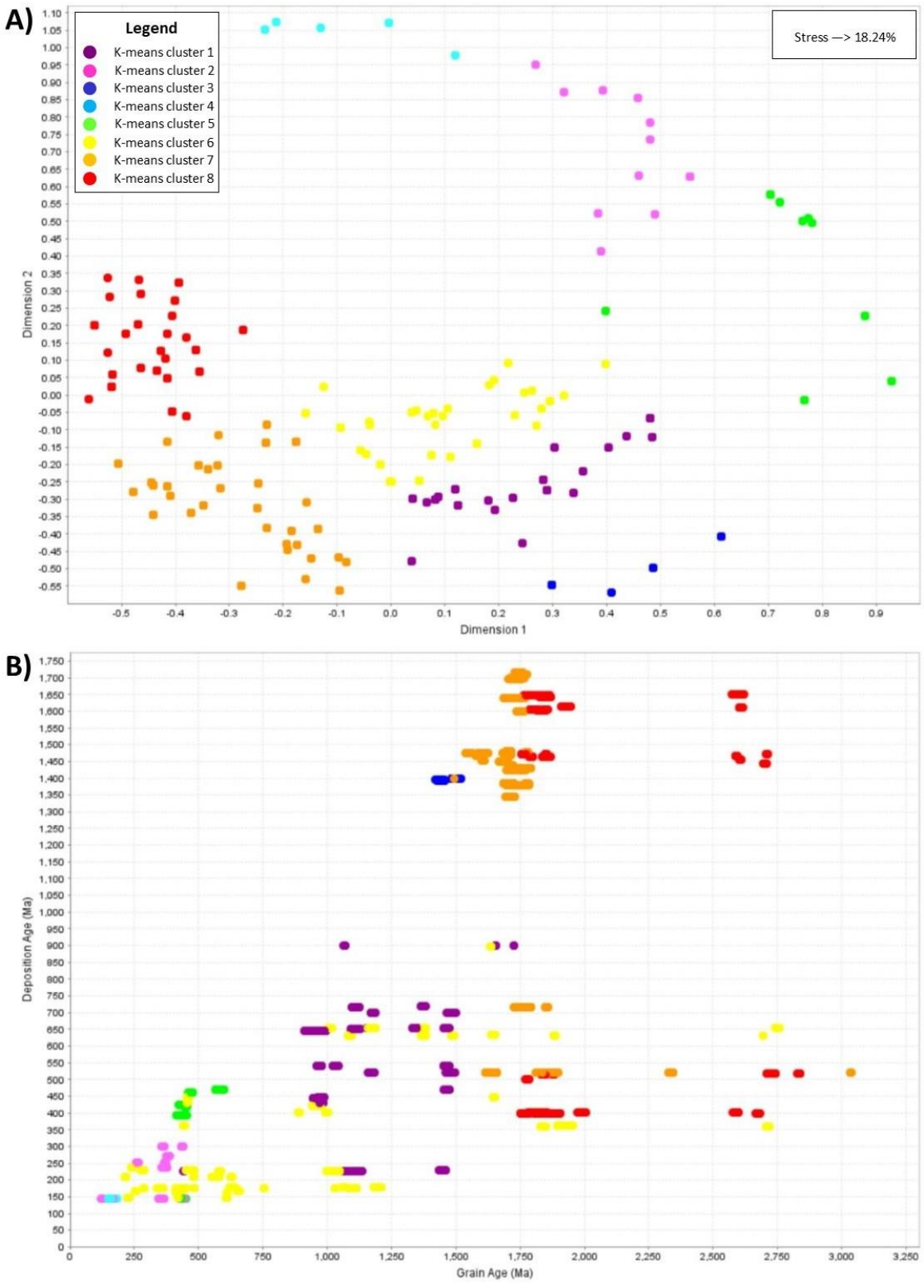


Figure 5.9 Caption on following page.

Figure 5.9 (A) MDS ordination of the test dataset from Laurentia with Kolmogorov-Smirnov dissimilarities. Results are produced from the isoMDS function within the MASS package in R. These results are plotted in ioGAS, displaying K-means clusters, produced from the stats package in R. **(B)** A multiple probability density plot created in ioGAS from the samples used to produce the MDS ordination, providing context to dominant grain ages within the samples of each k-means cluster.

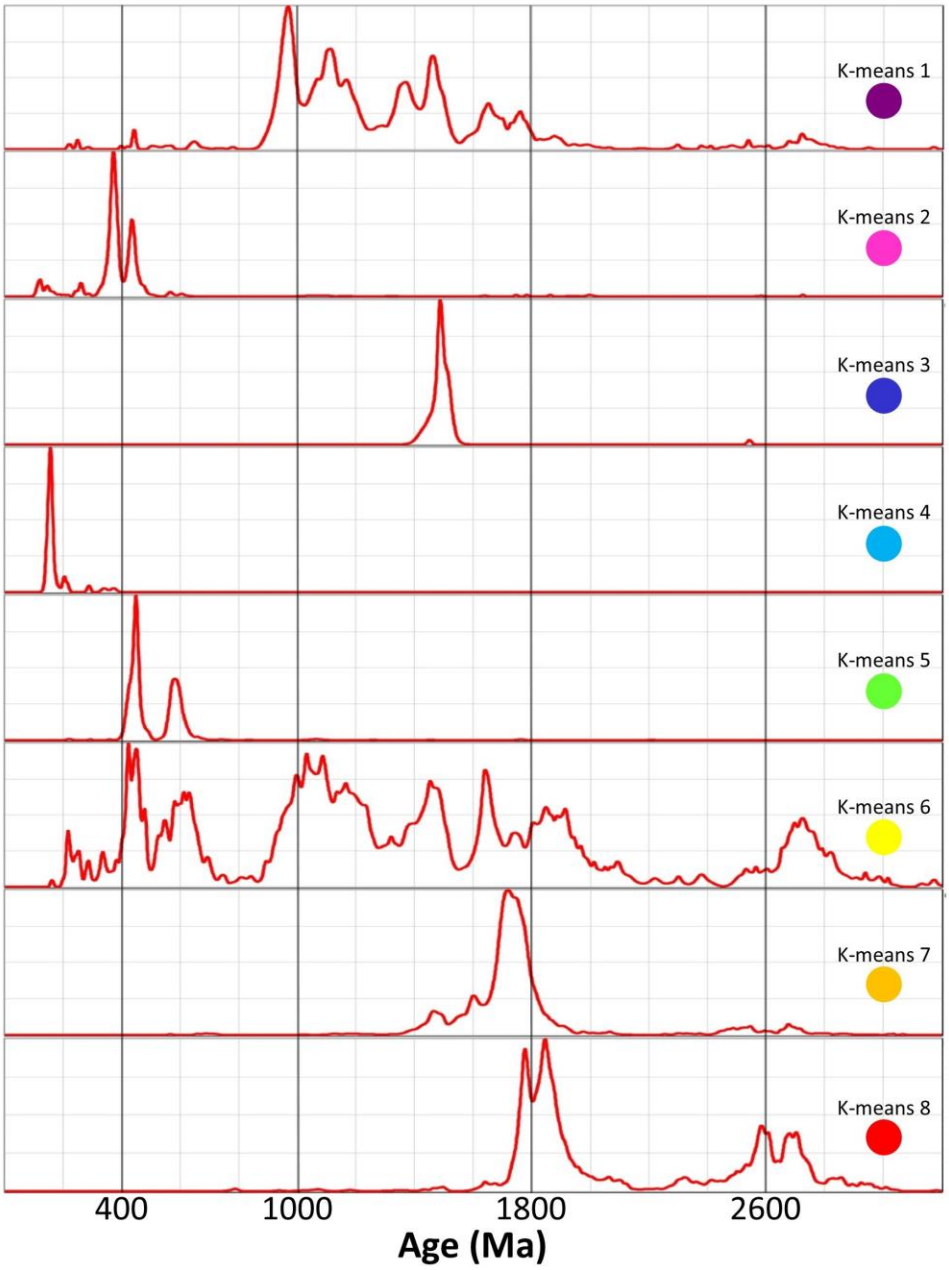


Figure 5.10 Stacked PDPs normalized to 100%, created in FitPDF from the samples of each K-means cluster of the Kolmogorov-Smirnov MDS ordination.

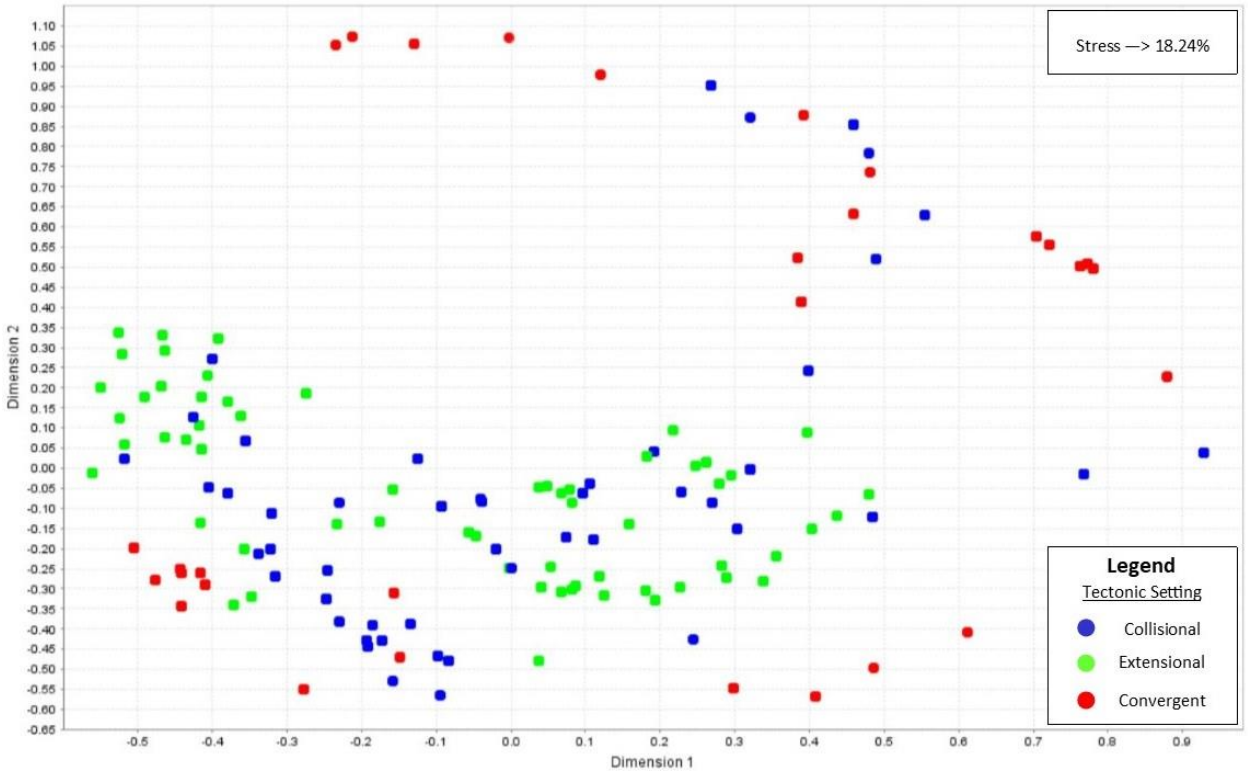


Figure 5.11 Plot of the MDS ordination of K-S dissimilarities using isoMDS in R. Each point is coloured to represent the geodynamic setting inferred from the grain age spectra, as per the method described by Cawood et al. (2012).

The final MDS plot was produced (Figure 5.11) to investigate possible patterns in geodynamic evolution from the ordination of samples. As observed by Spencer & Kirkland (2016), MDS ordinations can produce results that display the evolution of geodynamic settings in sedimentary successions. This MDS ordination produces a distinct cluster of extensional tectonic settings along the bottom half of dimension 2 within the plot, while collisional and convergent settings display no obvious clustering. Although this pattern produced is by no means conclusive when compared to those displayed by Spencer & Kirkland (2016), the grain age trend described can offer some assistance in providing an interpretation. This pattern can be explained by the northern and western margins of Laurentia being dominated by extensional tectonic settings from 1700 Ma to 350 Ma. Regional collisional and convergent tectonic settings are also observed across the aforementioned margins much less frequently. At approximately 350 Ma the tectonic regime across the northern and western margins of Laurentia changes to primarily convergent and collisional tectonic settings.

5.2.2.2 MDS Ordination Using the Bray-Curtis Statistic

The second variation of an MDS ordination was conducted using the Bray-Curtis statistic as a dissimilarity measure in conjunction with the monoMDS function in R. This ordination was produced to compare results with the baseline methods described in previous detrital zircon publications. Using this method with the Laurentian dataset produced a stress value of 15.50%, resulting in a goodness of fit described as fair. As with the first MDS variation, this method does not produce distinct clusters of samples, instead there appears to be a trend in the ordination that is produced from the dominant grain age peaks within samples.

The production of an MDS plot through the application of a colour ramp and symbol size variations, as explained in the previous section, displays a clear trend controlled by the high probability grain age peaks within each sample (Figure 5.12 A). Included with the MDS plot is a multiple probability density plot (Figure 5.12 B), where high probability (i.e. >80%) grain age peaks are subjected to the symbology used with the MDS plot to both visualize the effects of symbology and understand the dispersal of grain ages within the samples used. The high probability grain age trend displays samples containing the oldest grain age peaks plotted in the bottom right quadrant of the MDS plot, followed by samples with progressively younger high probability grain age peaks towards positive dimension 2 values. This trend then progresses to negative dimension 1 values where the bulk of samples dominated by young high probability grain ages are plotted. The trend observed with this variation of MDS ordination does not appear to be an improvement when compared to the trend observed in Figure 5.8 A. Further, comparing the multiple PDP from this method to that of the baseline multiple PDP (Figure 5.8 B), there is an obvious loss of detail in the high probability grain ages within each sample.

Figure 5.12 (A) MDS ordination of the test dataset from Laurentia with Bray-Curtis dissimilarities calculated with 50 bins. Results are produced from the monoMDS function within the vegan package in R. These results are plotted in ioGAS, with criteria for symbology as follows: Only grain age peaks above 80% probability are displayed, a blue (young) to red (old) colour ramp is applied to the peaks, and a symbol size division is applied, promoting the visualization of multiple high probability grain age peaks within a single sample. **(B)** A multiple probability density plot created in ioGAS from the samples used to produce the MDS ordination, only displaying high probability (>80%) grain age peaks.

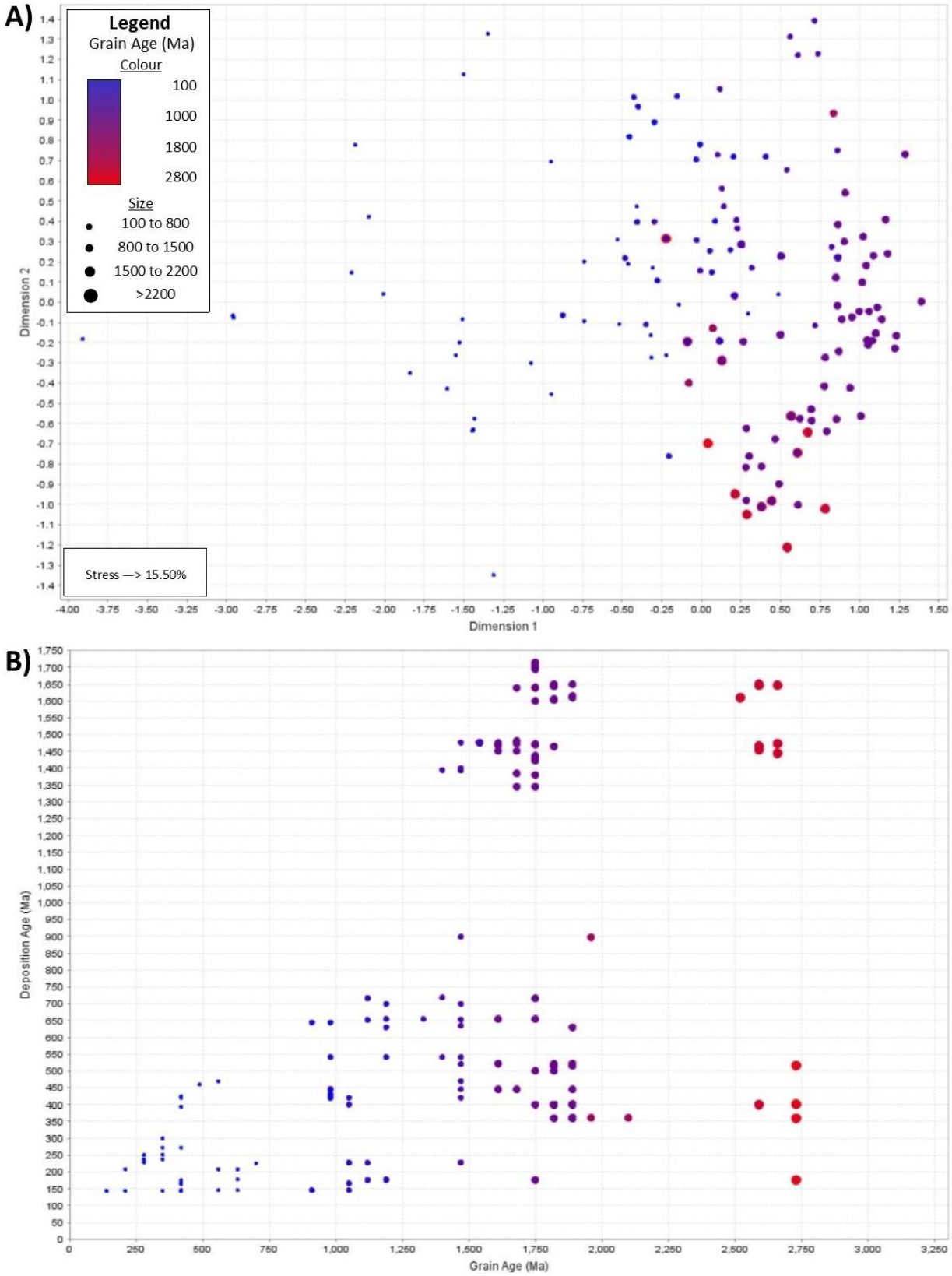


Figure 5.12 Caption on previous page.

An additional MDS plot (Figure 5.13 A) displays the supplementary K-means clusters produced in R. For the sake of consistency, a total of eight clusters are produced for further analysis of the ordination. In addition to the ordination plot, a multiple PDP (Figure 5.13 B) was included to supplement the analysis of clusters produced. The plots in Figure 5.13 support the observation of a high probability grain age trend, although to a lesser extent than that observed in the first variation of MDS ordination. K-means cluster 8 provides an example of decreased clarity, observed in the bottom right quadrant of the MDS plot where the oldest high probability grain ages are located. The inclusion of a sample with a distinctly younger high probability grain age of approximately 1100 Ma within this cluster is odd. As previously emphasized, clusters are selected to produce the lowest within-group sum-of-squares from the two dimensional ordination in which the samples are plotted (Borcard et al., 2011). Therefore, this is not a flaw of the K-means clustering, rather the methods selected for the second variation of MDS ordination appear to produce less distinct trends when compared to the first method.

To supplement the analysis of K-means clusters, stacked single PDPs were produced to summarize the samples within each cluster (Figure 5.14). The summarization of each cluster provides additional support for the high probability grain age trend observed in Figure 5.12 A. The first two K-means clusters, 3 and 8, display nearly identical PDPs and display the initiation of the grain age trend. Samples within these two clusters contain the largest quantity of detrital zircon grains ranging from 2600 Ma to 2800 Ma. The description of the trend continues with K-means clusters 4 and 5, which also share a similar PDP. These clusters display a roughly unimodal distribution of grain ages, ranging from 1600 Ma to 1800 Ma. The decreasing dominant grain age trend is completed with K-means clusters 1, 7, and 6 progressively. K-means cluster 1 displays a bimodal distribution between 900 Ma to 1500 Ma. The remaining two clusters, 7 and 6, both show roughly unimodal grain age distributions of detrital zircon grains from 400 Ma to 600 Ma. K-means cluster 2 was not mentioned in this trend as it lies at the origin of the plot (0,0) and displays a PDP that contains grain ages across the entire spectrum. A similar cluster was observed in the first variation of MDS ordination, and it is thought that this may be representative of sediment recycling to some degree. The placement of these samples around the origin is logical, as they contain grain ages throughout the trend described.

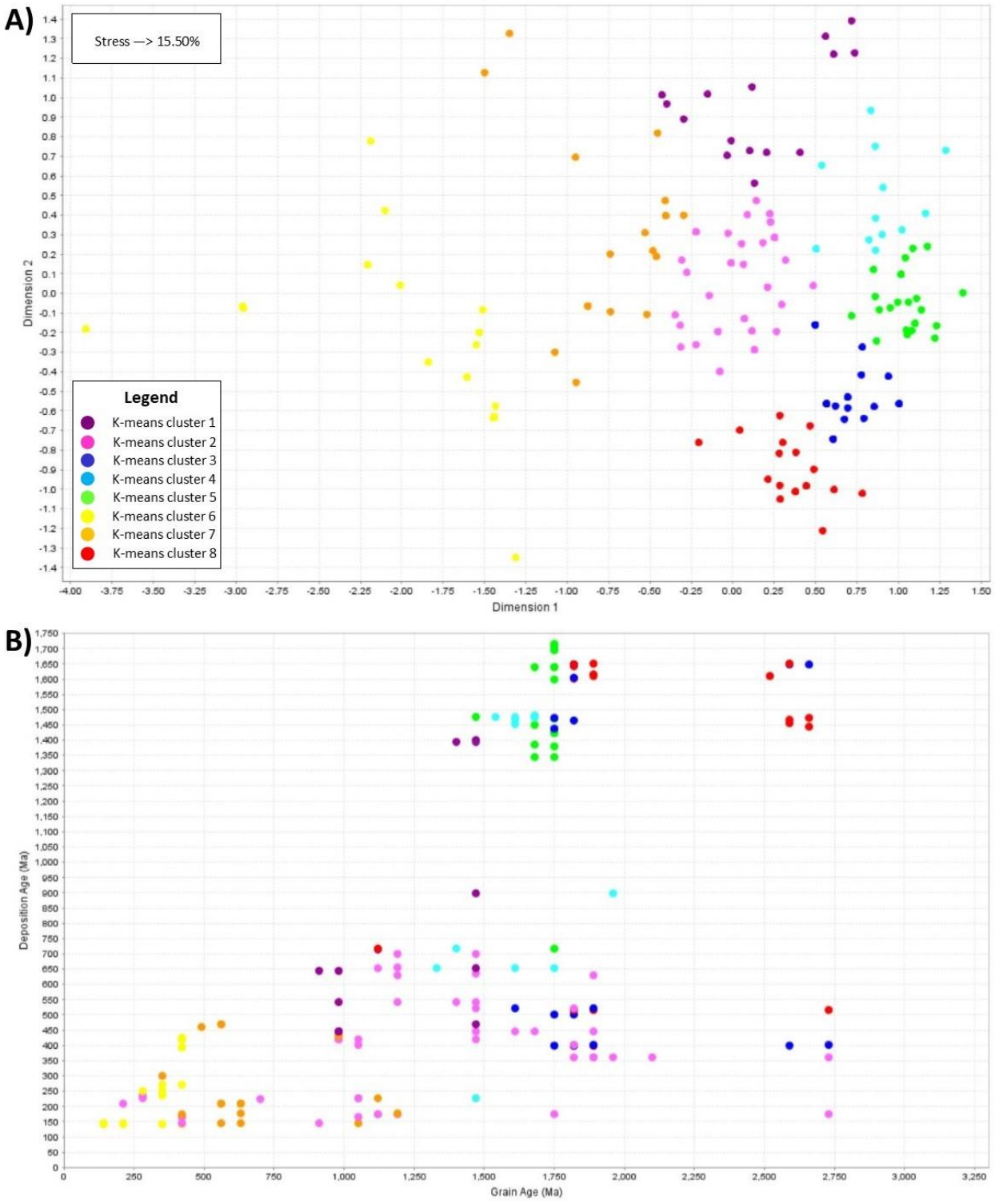


Figure 5.13 Caption on following page.

Figure 5.13 (A) MDS ordination of the test dataset from Laurentia with Bray-Curtis dissimilarities. Results are produced from the monoMDS function within the vegan package in R. These results are plotted in ioGAS, displaying K-means clusters, produced from the stats package in R. **(B)** A multiple probability density plot created in ioGAS from the samples used to produce the MDS ordination, providing context to dominant grain ages within the samples of each k-means cluster. The multiple PDP only displays high probability (>80%) grain age peaks.

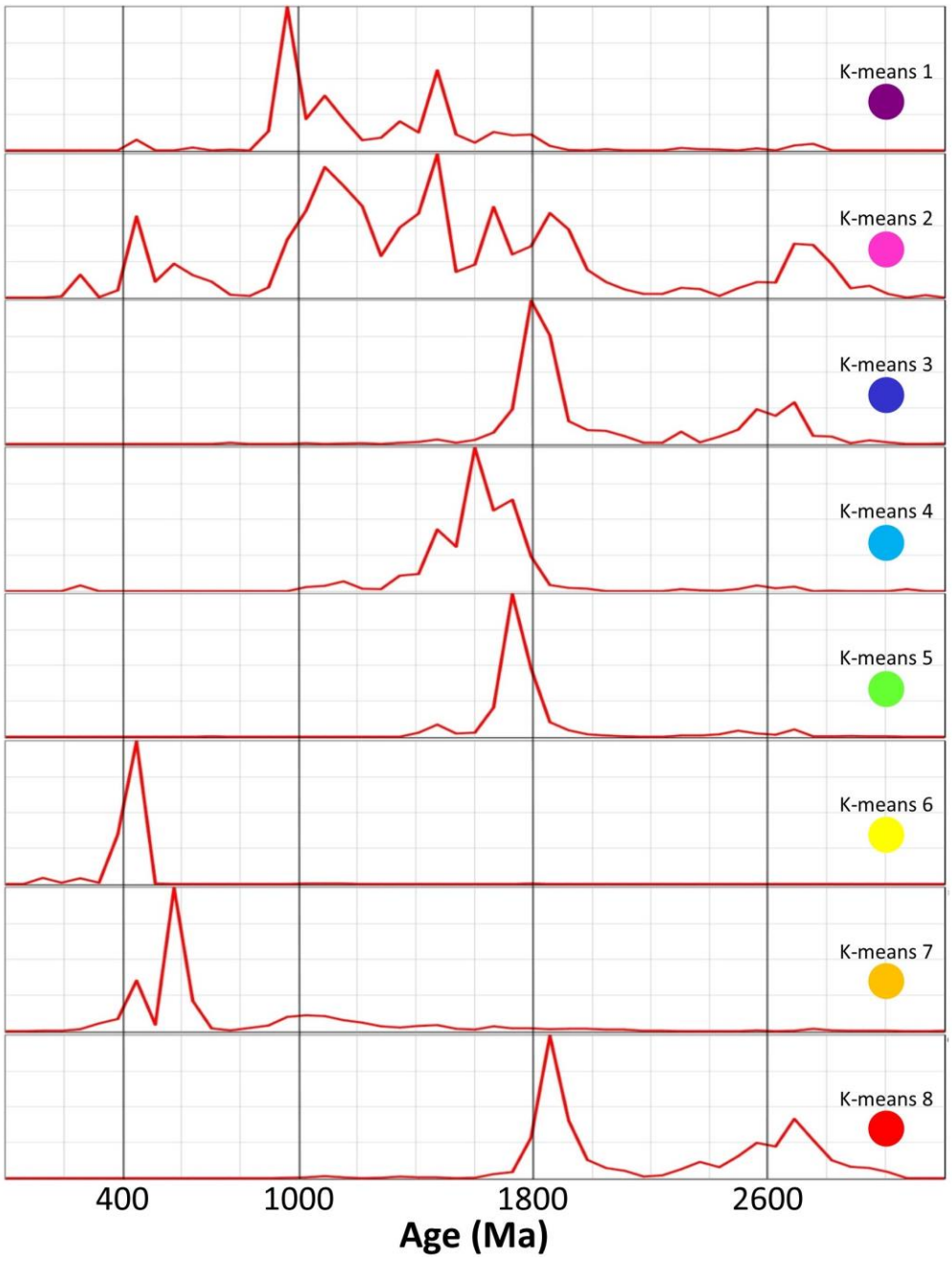


Figure 5.14 Stacked PDPs normalized to 100%, created in FitPDF from the samples of each K-means cluster of the Bray-Curtis MDS ordination

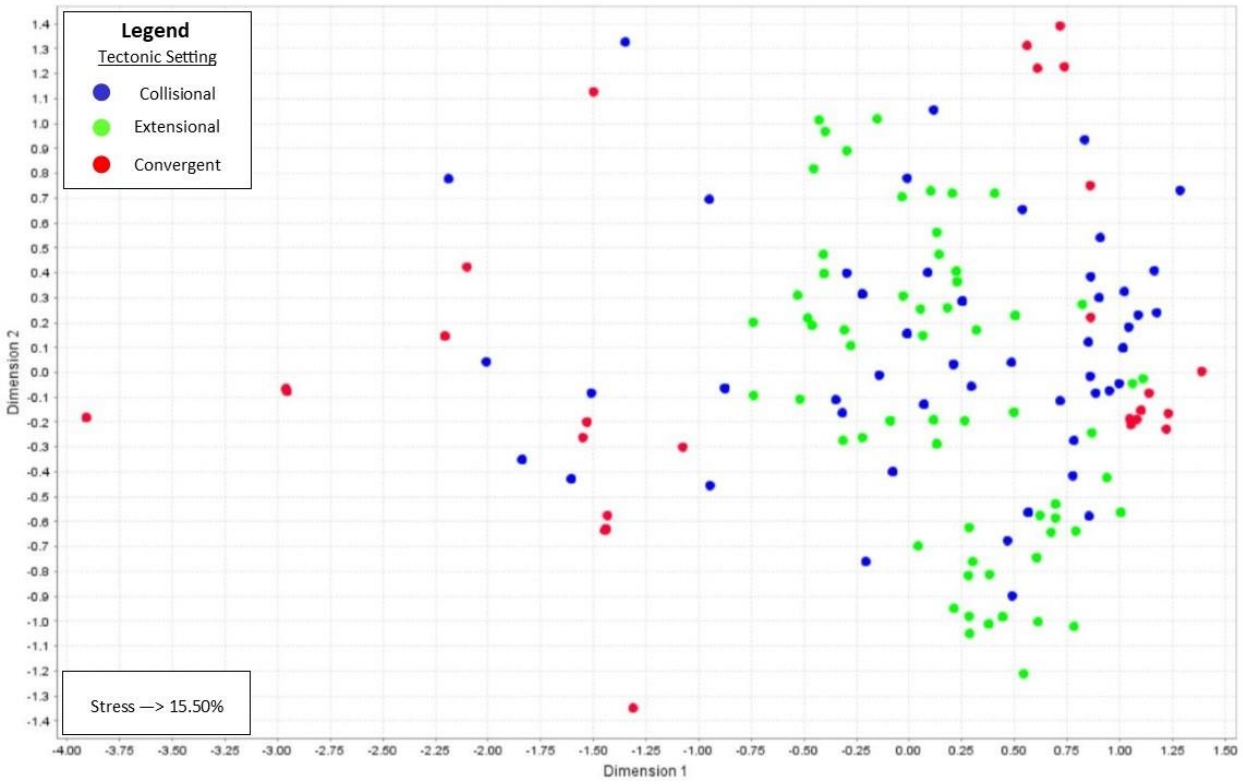


Figure 5.15 Plot of the MDS ordination of Bray-Curtis dissimilarities using monoMDS in R. Each point is coloured to represent the geodynamic setting inferred from the grain age spectra, as per Cawood et al. (2012).

The final MDS plot (Figure 5.15) was created to look for obvious trends in geodynamic evolution from the ordination produced. Using the method described by Cawood et al. (2012) to observe general geodynamic settings produces a similar pattern to that observed in the first MDS variation. Figure 5.15 displays a clear restriction of extensional tectonic settings to the right half of dimension 1. As described in the equivalent plot with the first variation of MDS ordination (Figure 5.11), this is interpreted as the northern and western margins of Laurentia being dominated by extensional tectonic settings from approximately 1700 Ma to 350 Ma, with regional collisional and convergent settings throughout that time. Beginning at 350 Ma the tectonic regime across these margins on Laurentia no longer display extensional settings, instead they are dominated by convergent and collisional settings.

5.2.3 Discussion

Having analyzed both variations of MDS ordination with the Laurentian dataset, this statistical technique is ill-suited for use with geographically extensive detrital zircon datasets. In terms of identifying sample clusters as demonstrated by Vermeesch (2013), the resulting ordinations did not clearly identify specific sample groupings. This lack of clear sample clustering was supplemented by K-means clustering to identify any sort of pattern within either ordination. The resulting K-means clusters did not provide meaningful aid, rather this technique emphasized the previously identified trend of dominant high probability grain ages within samples. Therefore, MDS is not effective for provenance analysis when datasets containing samples from vast geographic regions and multiple basins are analyzed in conjunction with each other. This statistical method is most effective when used for the analysis of smaller geographic regions, restricted to samples from a single sedimentary basin.

MDS has also been used to investigate geodynamic evolution (e.g., Spencer & Kirkland, 2016), where trends resulting from the ordination display a transition of tectonic environments through time. Although both MDS variations produced trends from the samples analyzed, the interpreted geodynamic evolution was much more of a generalization, observing the dominance of extensional settings across the northern and western margins of Laurentia from the Mesoproterozoic until the Carboniferous. These results are not entirely unexpected, with the amalgamation of samples from across the western and northern margins of North America being influenced by differing levels of sediment recycling and many different provenance sources. Further, the examples discussed by Spencer & Kirkland (2016) are from single orogenic belts, a stark contrast to the samples representative of multiple orogenic events in the Laurentian dataset. Again, the use of MDS for the investigation of geodynamic evolution is most effective when used with datasets restricted to a smaller geographic region or single orogenic events.

In determining which MDS method is preferred, the data preparation and results are both considered. Regarding data preparation, the first variation with the Kolmogorov-Smirnov dissimilarity measure does not require data pre-treatment, whereas the Bray-Curtis dissimilarity measure does. Data pre-treatment has proven to be time consuming and is therefore considered a large downside to the second variation of MDS ordination. The results of both MDS variations produce similar results, displaying dominant grain age trends and vague geodynamic evolution patterns. Although the high probability grain age trend and geodynamic evolution pattern

produced from the K-S dissimilarity measure were easier to distinguish. As a result, there are no clear benefits to using the Bray-Curtis dissimilarity measure when compared to the K-S dissimilarity measure. The only limiting factor of the first MDS variation is the use of the isoMDS function. In terms of being able to reproduce results this algorithm is beneficial; however, the risk of producing an ordination representative of a local minima is high. Consequently, the use of the K-S dissimilarity measure with the monoMDS function is recommended for MDS ordination of detrital zircon data.

An additional issue to consider with the two variations of MDS conducted on the Laurentian dataset is the number of dimensions selected. The ease in which conducting analysis and deriving interpretations with a two dimensional ordination is appealing, and as such it is the standard for conducting MDS on detrital zircon data (e.g., Vermeesch & Garzanti, 2015). Despite the convenience of selecting two dimensions, it is probable that this selection is over simplifying the Laurentian dataset, resulting in multiple axes of variation forced on to a single plotted dimension (Holland, 2008). Creating a scree plot of the resulting stress values from ordinations with increasing dimensionality (Figure 5.16) appears to confirm the over simplification of the Laurentian data. With the scree plots of Figure 5.16 as a guide, the selection of either three or four dimensions would be a better representation of the data. Moving forward, either the number of dimensions should be increased for analysis, or a more geographically restricted analysis should be conducted in which two dimensions is appropriate for the dataset to be analyzed.

This statistical technique was not used in conjunction with the visual analysis of detrital zircon to supplement paleontology datasets. Although, MDS would likely prove to be beneficial on a geographically smaller scale than that of the Laurentian dataset. To take advantage of this technique, multiple individual basin analyses would have to be conducted, which was determined to be too large of a burden considering the time frame in which this project was to take place.

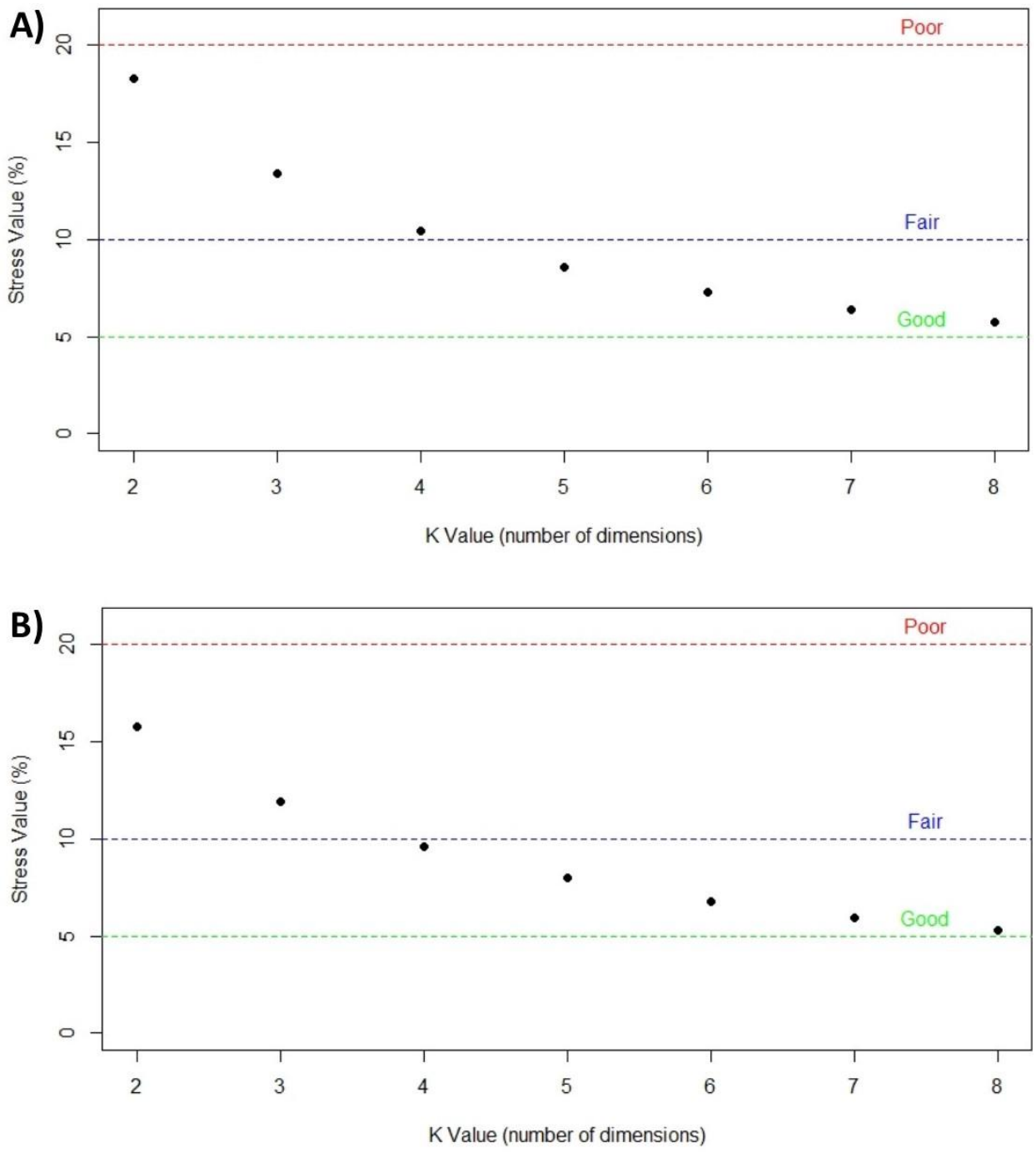


Figure 5.16 Plots created in R displaying the reduction in stress with an increasing number of dimensions selected to produce MDS ordinations from the Laurentian dataset. Divisions between the goodness of fit definitions described by Kruskal (1964) are included. **(A)** Reduction of stress using the Kolmogorov-Smirnov statistic with the isoMDS algorithm. **(B)** Reduction of stress using the Bray-Curtis statistic with the monoMDS algorithm.

6. Association between Depositional Environments and Trace Fossil Occurrence during the Ediacaran to Cambrian Transition

The purpose of this study is to investigate methods in which possible associations between depositional environments and trace fossils can be observed. To facilitate this objective, an ichnology database has been designed and populated for use with data mining techniques. Further, a detrital zircon dataset has been compiled as supplementary material for visual analysis as it can be used to validate plate tectonic models and indicate tectonic settings (Cawood et al., 2012). With access to these large datasets, the initial step in the knowledge discovery from data (KDD) process, selection of data is addressed (Han et al., 2012). With this data, the remaining steps in the knowledge discovery process can be conducted. This chapter will focus on both visual and statistical pattern recognition techniques in detail.

Following data selection, the knowledge discovery process proceeds to data transformation, data mining, and finally interpretation of results (Han et al., 2012). Although some aspects of data mining have been explored in previous chapters (e.g., visual analysis of detrital zircon, and the application of MDS), this chapter will focus on visual and statistical analysis of ichnology data and their depositional environments. Visual analysis will consist of spatial and temporal analysis with paleogeographic reconstructions, while statistical analysis is conducted using correspondence analysis.

6.1 Visual Analysis

As part of the KDD process, visualization of data using GIS software is fundamental when investigating global phenomena. With the aid of GIS software and file or personal geodatabases in Microsoft Access, large volumes of data can be plotted and manipulated. Although this software addresses our need for spatial and temporal analysis, it does not provide a paleogeographic context. Published paleogeographic reconstructions from the Neoproterozoic to Cambrian suggest that the configuration of the continents was dramatically different compared to their current positions. To address this issue, plate tectonic reconstruction software is used for the best possible approximation of spatial, temporal, and paleogeographic relationships.

A large dataset has been compiled, with more than 2.1 million geological map polygons, and over 88,000 geologically significant data points. This data is used to supplement a trace fossil dataset extracted from IchnoDB to look for patterns in the data both spatially and temporally. To achieve this objective, the supplementary data is used to create paleogeographic reconstructions at predetermined time intervals. Due to our specific and narrow time frame of interest, the Ediacaran to Cambrian, many of the geological map polygons and geologically significant data points are not relevant to this study. Using queries in both ArcGIS and Microsoft Access, it is possible to dramatically reduce the size of the supplementary datasets by filtering out any data not directly relevant to the time period defined. Through the application of these age criteria, the reduced datasets consist of roughly 190,000 geological map polygons, and 44,000 geologically significant data points. Finally, geoprocessing tools within ArcGIS (e.g., spatial join, union, intersect) enable the alteration of shapefiles for use in plate tectonic software, GPlates (Boyden et al., 2011; Qin et al., 2012).

The paleogeographic reconstructions created in GPlates broadly define environments (e.g., deep marine, shallow marine, terrestrial, orogenic belts) that may influence the spatial distribution of trace fossils. Additional information in the supplementary dataset provides approximate paleoclimatic zones, general paleocurrent directions, and the tectonic setting of global sedimentary basins through the Ediacaran to early Cambrian. In the integration of the multiple datasets into a single model, visual analysis of spatial and temporal patterns within the trace fossil data is more robust.

6.1.1 Data Collection and Use for Visual Analysis

Although paleogeographic reconstructions for the Ediacaran and Cambrian exist (e.g., Golonka et al., 1994; Golonka, 2012; Li et al., 2013), these maps are not easily converted in to a digital format for visualization or manipulation in GIS software. The process of digitization has been described by Cao et al. (2017), where these authors fit multiple digitized paleogeographic reconstructions onto a plate tectonic model for manipulation. This method of transposing paleogeographic reconstructions on to alternative plate tectonic models is not without issues, as the dissimilarities between two models can create confusing results (Figure 6.1). To avoid visually discontinuous paleogeographic reconstructions resulting from alternative plate tectonic

models, the compilation of geological maps and geologically significant points were used to define broad environmental regions. The distribution of deep marine, shallow marine, terrestrial, and orogenic belts were determined with the assistance of the compiled datasets.

As these paleogeographic reconstructions will supplement identification of possible spatial patterns within the trace fossil data compiled in IchnoDB, time periods applicable to this study had to be defined. Three divisions of time will be used, based on observable increases in the ichnodiversity spanning the Ediacaran to Cambrian periods. This phenomenon was described by Mángano & Buatois (2014), with a visual summarization of these events in Figure 6.2. The first clear increase in both ichnodiversity and ichnodisparity is observed between the Ediacaran and Terreneuvian. Figure 6.2 also shows a less dramatic increase in ichnodiversity and ichnodisparity between the Terreneuvian and Series 2 within the Cambrian. Based on this summarization of the Ediacaran to Cambrian transition in terms of trace fossils, the Ediacaran, Terreneuvian, and Epoch 2 were the epochs selected for analysis. For effective visualization of paleogeographic reconstructions in GPlates, a single date had to be assigned for each epoch. The visualization of the Terreneuvian reconstruction at 530 Ma was selected based on the approximate midpoint of the epoch. A date of 510 Ma was selected for Epoch 2, as this assists in producing a paleogeographic reconstruction that is distinct from that of the Terreneuvian. Determining an appropriate date for the Ediacaran was less straightforward due the complexity in defining possible divisions within the temporally largest defined geological period (Xiao et al., 2016). A date of 550 Ma was selected, as *Cloudina* are frequently observed both below Ediacaran trace fossils (e.g., Jensen et al., 2007) and co-occurring with Ediacaran trace fossil (e.g., Cai et al., 2014). *Cloudina* are often described as indicative of the terminal Ediacaran (e.g., Jensen et al., 2007; Xiao et al., 2016) which begins at approximately 550 Ma. Therefore, the selection of 550 Ma as the date to summarize Ediacaran trace fossils likely provides an appropriate time frame for a significant portion of trace fossils extracted from IchnoDB, and it is also ideal in that it keeps a consistent spacing between the latter two dates.

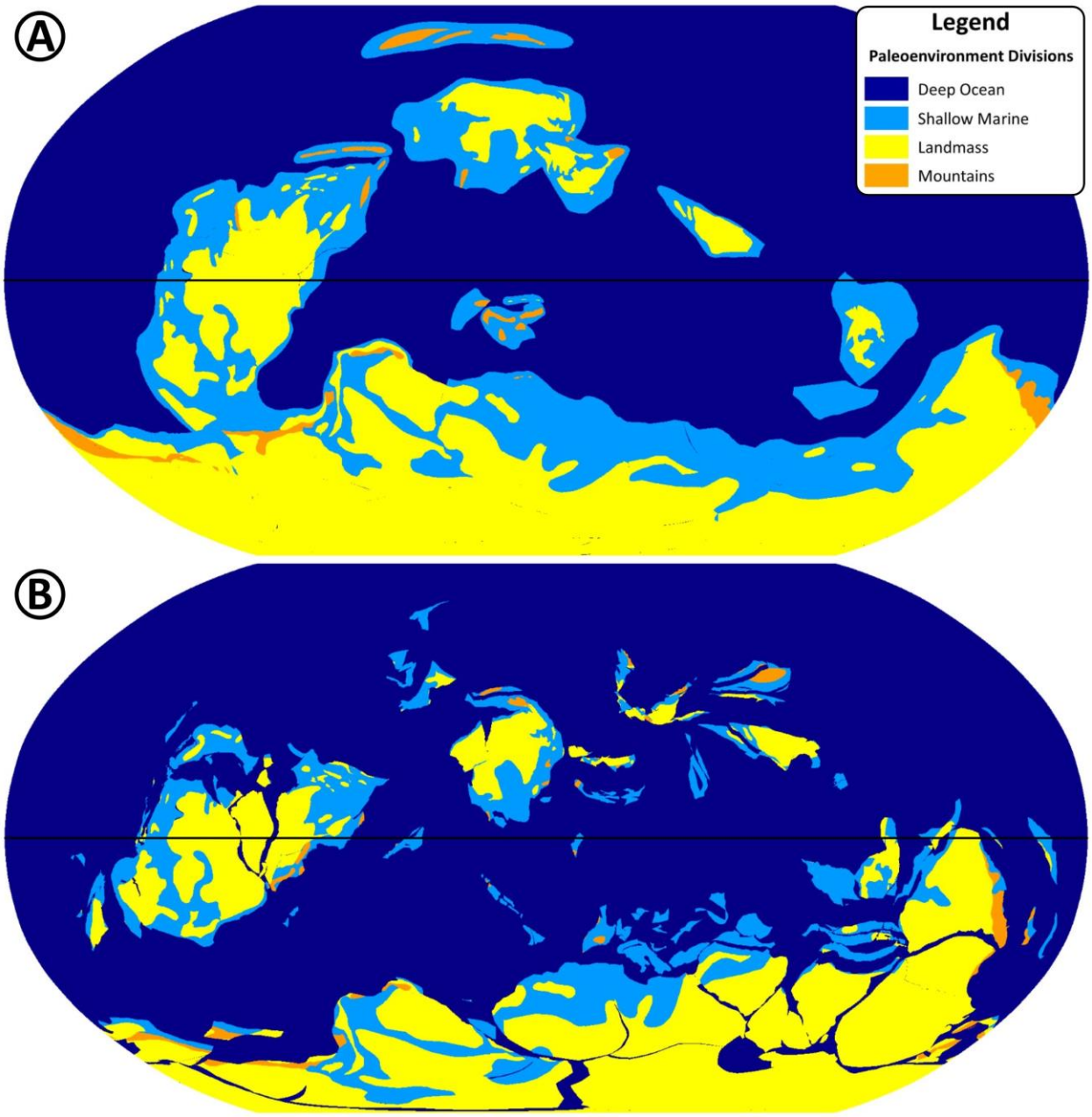


Figure 6.1 (A) Model provided by Cao et al. (2017), using a plate tectonic model from Matthews et al. (2016) in GPlates during the Early Devonian. The paleogeographic reconstructions discussed by these authors recognize five paleoenvironments: deep ocean, shallow marine, landmass, mountains, and ice sheets (not applicable to the Early Devonian). **(B)** Transposing the paleogeographic reconstruction from Cao et al. (2017) onto the PalaeoPlates Model provided by Eglington (2018c). The same four paleoenvironments are displayed in this model as above, but differences between the plate tectonic models produces a visually discontinuous reconstruction.

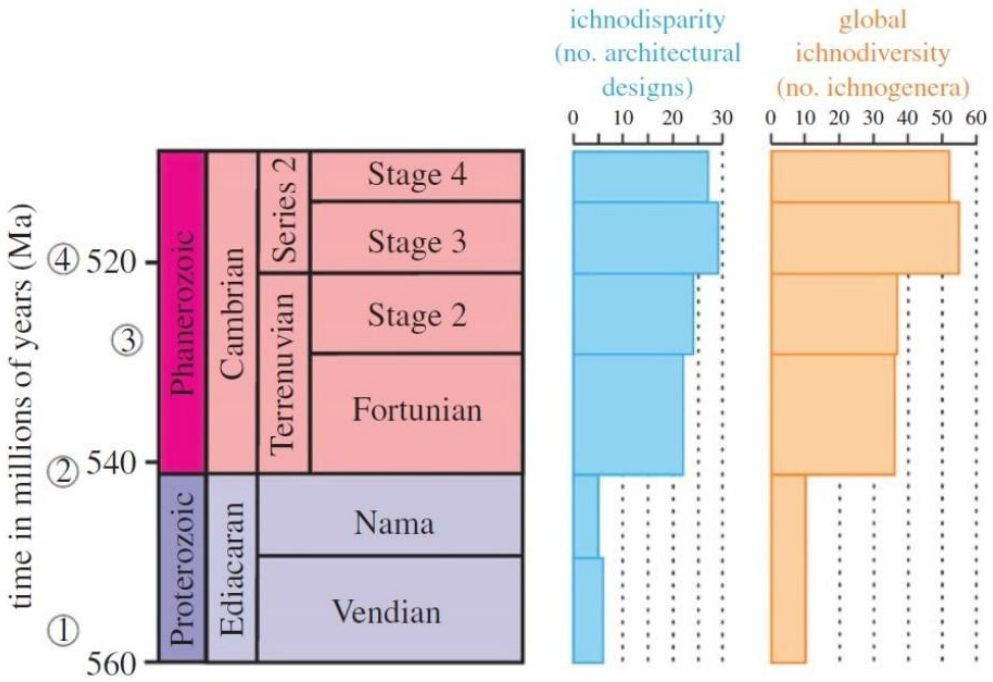


Figure 6.2 Summary diagram of changes in global ichnodisparity and ichnodiversity during the Ediacaran to Cambrian transition (modified from Mángano & Buatois, 2014).

With clearly defined intervals in which to create paleogeographic reconstructions, it is necessary to describe how the data collected is used to define each of the four paleoenvironments. A large portion of data compiled was used to define shallow marine settings. The geological maps collected were no exception, providing almost exclusively shallow marine indicators. There were two types of geological maps obtained, those with inferred depositional environments within the metadata, and those without. The geological maps of the Arctic (Harrison et al., 2011), Australia (GA, 2012), Namibia (CGMW, 2000a), South Africa (CGMW, 2000b), and China (Zheng & Hu, 2010) all included depositional environment information. With the metadata from these maps it was possible to define terrestrial, shallow marine, and deep marine environments with simple queries in Microsoft Access. The remaining geological maps for North America (GSA, 2005), South America (CGMW, 2001), and Asia (CGMW, 2017) were supplemented with lithology data provided by the global lithological map, GLiM (Hartmann & Moosdorf, 2012) through the use of geoprocessing tools in ArcGIS.

Where the compiled map polygons lack depositional metadata, the GLiM supplement is used to infer shallow marine settings through the indication of carbonates. It is observed that

carbonates throughout the geological record typically represent shallow marine, warm water conditions (Bosence & Wilson, 2003). Both thermodynamic and biogenic factors suggest carbonate formation within warm water conditions, with the presence of photosynthetic organisms defining the bathymetric conditions as they are restricted to the photic zone (Bosence & Wilson, 2003; Li et al., 2013). Cool and cold water carbonates also exist; however, it is observed that their production rates are roughly one order of magnitude less than those of warm water carbonates (James & Jones, 2016). Therefore, without additional data to distinguish between the paleoclimatic conditions, all carbonates are assumed to represent shallow marine warm water conditions.

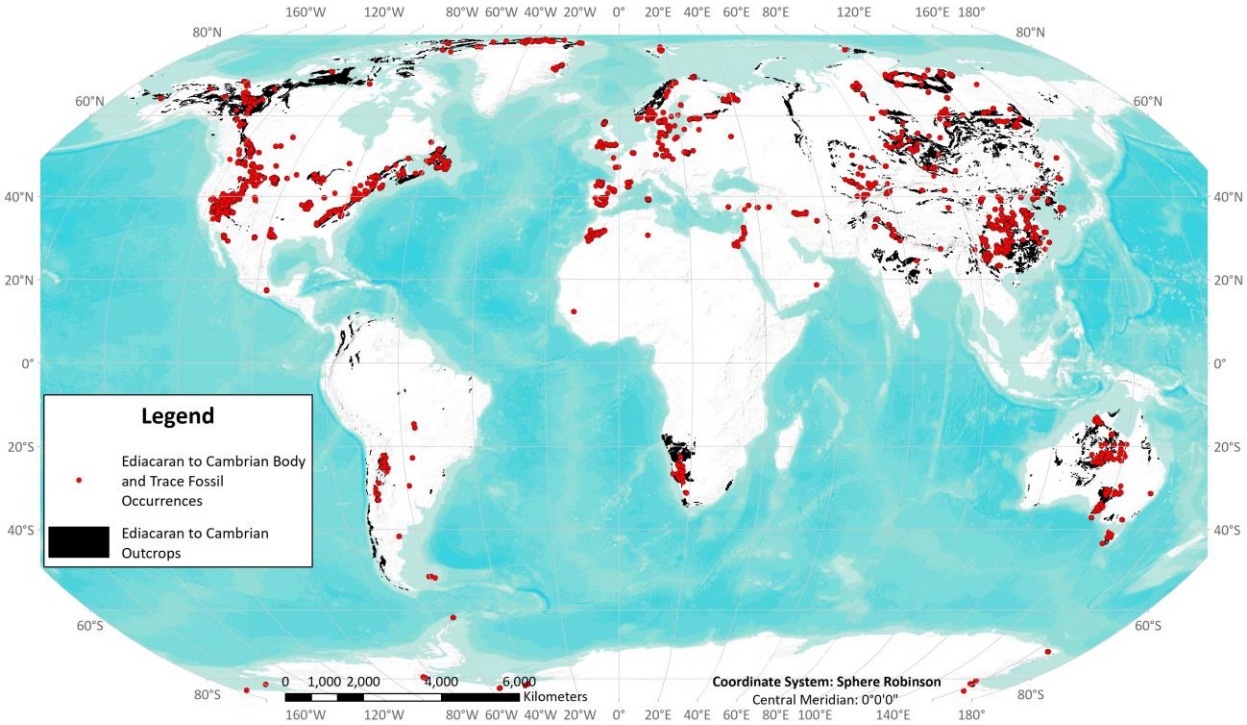


Figure 6.3 Global map displaying Ediacaran to Cambrian sedimentary geological units. Including Ediacaran to Cambrian body and trace fossil data extracted from the Paleobiology Database and IchnoDB emphasizes regions where relevant geological map data is not present, most notably Europe, northwestern Africa, the Middle East, and Antarctica.

The global coverage of the compiled geological maps is significant, although including the paleontology data highlights a lack of relevant Ediacaran to Cambrian map data in northwestern Africa, Europe, Antarctica, and the Middle East (Figure 6.3). It should be noted that the scales for the different geological maps ranged from 1:1,000,000 to 1:5,000,000. As a result of the small scales, broad geological age ranges are reported in some of the map polygons. Where there was enough data, polygons were restricted to more precisely defined age ranges to limit instances where data might be conflicting (e.g., units described as strictly "Cambrian" were preferred over those that may have been described as "Cambrian to Devonian").

The Paleobiology Database (PBDB) was used to create a dataset for the identification of paleoenvironments during the Ediacaran to Cambrian time period of interest. The use of PBDB data to refine paleogeographic reconstructions was first detailed by Cao et al. (2017). Although these authors focused on earlier periods in the Phanerozoic, the method is equally applicable during the Ediacaran to Cambrian. Cao et al. (2017) focused primarily on the definition and alteration of shallow marine paleoenvironments. Through this process these authors emphasized that living habits of the fossils must be considered to ensure the selection of strictly marine paleoenvironment data points are selected (i.e. no fluvial or terrestrial paleoenvironments should be included). During the Ediacaran to early Cambrian this is not an issue, as this point in geological history macrofauna were restricted to marine or marginal marine environments (Buatois & Mángano, 1993; Minter et al., 2017). Terrestrial colonization was initiated by prokaryotes and eukaryotes during the Precambrian (Minter et al., 2017); however, macrofauna did not begin colonization until the late Cambrian. This was initiated through temporary expansion into nonmarine environments during the late Cambrian to early Ordovician, evidenced by arthropod trackways on eolian sand dunes and burrow systems associated with ancient soils (Buatois & Mángano, 1993; MacNaughton et al., 2002). These trace fossil occurrences record colonization much earlier than body fossils, with the earliest body fossils described in rocks of a late Silurian age (Jeram et al., 1990). Therefore, the metadata provided with the PBDB dataset was taken at face value, providing both shallow and deep marine indicators, with missing metadata treated as undefined shallow marine depositional environments.

IchnoDB was used in a similar fashion to the body fossil data for the process of defining paleoenvironmental regions. Datasets were extracted for each time period of interest, excluding

trace fossils found in lithostratigraphic units where geological age ranges resulted in an occurrence plotting in more than one paleogeographic reconstruction. As with body fossils, trace fossils can provide information regarding the depositional environment in which they were formed (Frey & Pemberton, 1984; Pemberton et al., 1992a). Further, the autochthonous nature of trace fossils provides a more robust indicator of contemporary ecological conditions than that of body fossils, supporting the refinement of paleogeographic reconstructions using the methods employed by Cao et al. (2017). As observed with the body fossils, the trace fossil record does not record terrestrial colonization during the Ediacaran to Epoch 2 (Buatois & Mángano, 1993; MacNaughton et al., 2002; Minter et al., 2016b, 2017). Rather, the trace fossil record exhibits the initial colonization of marginal marine environments (i.e. tidal flats) by macrofauna during the Fortunian and continuing throughout the Cambrian (Mángano & Buatois, 2004; Collette et al., 2010; Mángano & Buatois, 2015). The first record of nonmarine trace fossils occurs in rocks of late Cambrian to early Ordovician age (MacNaughton et al., 2002). As a result, the depositional environment metadata from IchnoDB does not require any alterations or restrictions, and it will be used to provide indicators of marginal, shallow, and deep marine environments.

A dataset of paleoclimatic indicators was compiled from multiple sources (Boucot et al., 2013; Li et al., 2013), to further define deep marine, shallow marine, and terrestrial paleoenvironments. Shallow marine environments were defined based on previous interpretations and climatically sensitive lithologies. Of the shallow marine indicators, there were many instances of evaporites, in which two caveats should be included for a better understanding of these chemical precipitates. First, it is observed that evaporites typically form in shallow water (<10 m) settings within restricted basins (Schreiber & El Tabakh, 2000; James et al., 2010). Second, Boucot et al. (2013) noted that the term “marine evaporite” is misleading, and rather these deposits should be referred to as “shoreline region evaporites.” Deep marine paleoenvironment indicators within this dataset were uncommon, but where present they were defined by basinal shales and diamictites interpreted as unconfined debris flows. The remaining paleoenvironment indicators identified terrestrial settings through the presence of fluvial deposits, observed unconformities, and lithologies related to soil forming processes (e.g., kaolinite, bauxite, calcrete).

The final dataset used for creating paleogeographic reconstructions was extracted from the DateView geochronology database (Eglington & Armstrong, 2004; Eglington, 2004). From this dataset four geologically significant radiometric age types are defined: regional metamorphism, extrusive crystallization, intrusive crystallization, and cooling. The cooling ages recorded in this dataset provide details regarding the thermochronological history of the samples, which are used to infer the timing of exhumation (Reiners et al., 2005). These points are used to suggest subaerial exposure, typically representing orogenic events. The regional metamorphic radiogenic ages within DateView provide information similar to the cooling points, except that metamorphism is the result of continental thickening, typically as a result of orogenic activity (England & Thompson, 1984). Finally, the intrusive and extrusive crystallization radiogenic ages signify the production of new continental crust. Extrusive igneous geochronology records are used to imply subaerial exposure and the presence of terrestrial paleoenvironments. It is recognized that the presence of extrusive igneous records does not exclude the possibility of submarine eruptions, although as a generalization this is not accounted for. Intrusive igneous geochronology records are used primarily to imply orogenic paleoenvironments and the presence of subduction zones (Frost & Frost, 2014). As igneous plutons form at depth, they offer no indication of subaerial exposure. Rather, intrusive igneous rocks are associated with convergent margins (Frost & Frost, 2014), and where continental collisions result in magma generation (England & Thompson, 1984). Frost & Frost (2014) observed that igneous intrusions may form in the absence of the two previously mentioned criteria, as layered mafic intrusions or major anorthosite complexes, although notable occurrences of such plutons are not recorded during the Ediacaran to Cambrian.

Two datasets were collected as supplementary material to the paleogeographic reconstructions created. The first additional dataset was a compilation of published detrital zircon analyses. The detrital zircon dataset was composed from two sources, the first was a compilation of detrital zircon analyses from the north and west coast of North America, with details outlined in Chapter 4 of this thesis. The second, and much larger component of the detrital zircon dataset was extracted from a database published by Puetz et al. (2018). The detrital zircon data is used to infer the tectonic setting of the sedimentary basins in which the samples are deposited, using a process defined by Cawood et al. (2012). These authors proposed a model to define three generalized tectonic settings in which the sedimentary basin formed,

being convergent, collisional, or extensional settings. It is observed that this calculation requires an accurate deposition age of the sediment from which the zircon grains have been extracted for reliable results. The detrital zircon dataset compiled for Laurentia contains detailed deposition ages for all samples compiled, while the data compiled by Puetz et al. (2018) is missing this particular information. To account for this flaw, the maximum deposition age determined by the youngest recorded detrital zircon grain age within a sample was used as the deposition age (Fedo et al., 2003). This "fix" will result in a dataset that is biased towards convergent and collisional tectonic settings, as the maximum depositional age will in most cases provide an age that is older than the true deposition age. As both trace and body fossils are associated with sedimentary rock, the occurrence of these fossils is controlled by the accumulation of sediments, which is observed to be non-uniform throughout geological history (Holland, 2016). This author underscored the fact that tectonic settings have a significant influence on the preservation potential of sedimentary basins, with a strong bias towards extensional tectonics settings compared to convergent tectonic settings. The inclusion of this dataset will help visualize this bias in the ichnology data.

The second dataset added to the paleogeographic reconstructions contains paleocurrent information. This data was extracted from a database published by Brand et al. (2015), in which paleocurrent information from previously published papers and unpublished theses from around the globe is compiled. With the addition of paleocurrents to the paleogeographic reconstructions it may be possible to identify large drainage systems and suggest areas where marginal marine (e.g., deltas, estuaries) conditions were present. The inclusion of this dataset was plagued with issues, most importantly is a poorly defined time scale. Specifically, the entire Proterozoic eon is grouped together, precluding the extraction of meaningful data for the Ediacaran. In addition, the paleocurrents are referenced to present-day north, requiring manipulation for use in a paleogeographic model where cardinal directions have shifted.

6.1.2 Paleogeographic Reconstruction Methods

Defining paleogeographic environments was conducted in a systematic process. The first step of this process was distinguishing deep marine paleoenvironments from shallow marine paleoenvironments. Broad distinctions were made by using the PalaeoPlate model (Eglington,

2018c) in GPlates. Where this model displayed the absence of continental crust, deep marine settings were defined. This method is based on the theory of plate tectonics (Condie, 1997; Rowley, 2002), whereby seafloor spreading is described as a conveyor belt. It is explained that new oceanic lithosphere is created at mid-ocean ridges and old oceanic lithosphere returns to the mantle at subduction zones to accommodate the newly created lithosphere. Multiple ages are reported for the oldest in-situ oceanic crust on Earth, with Müller et al. (2008) reporting an age of approximately 270 Ma, while Rowley (2002) reported an age of approximately 180 Ma. Regardless of this inconsistency, both reported ages are significantly younger than the time period of interest. Through inductive reasoning, it is implied that the absence of continental crust in a plate tectonic model during the Ediacaran to Cambrian transition suggests the presence of oceanic crust. This inductive argument cannot be proven; however, the purpose of this argument is not to demonstrate the truth of our conclusion, but rather to suggest this conclusion to be probable (Copi, 1978). Inductive logical arguments are used throughout the definition of the remaining paleoenvironments.

The assumption that the presence of continental crust eliminates the possibility of deep marine paleoenvironments from existing within these plate boundaries is an over simplification. Where supplementary data is available additional criteria are imposed to further refine the distinction between these paleoenvironments. Criteria were created for both siliciclastic and carbonate systems. For the purpose of consistency and simplicity, this study defines the division of shallow and deep marine in siliciclastic systems at the shelf break (Arnott, 2010).

Carbonate sedimentary systems provide a challenge for defining shallow marine environments during the Ediacaran to Cambrian transition. Modern carbonates are typically split based on the depth, with the recognition of shallow marine and deep marine carbonates (Bosence & Wilson, 2003). These authors observed that shallow marine carbonates are split further to account for climatic conditions, with the recognition of warm and cool water conditions. Warm water carbonates provide the most useful definition of shallow marine environments as they are produced by organisms that rely on photosynthesis, and thus signify a bathymetric depth ranging from 10 to 100 m (Bosence & Wilson, 2003). Further, these authors suggested that warm water carbonates dominate the geological carbonate record as a result of their rapid rate of accumulation. These premises provide a basis to inductively reason that the presence of

carbonates indicate a shallow marine setting. Although an important caveat must be recognized in the symbiont conundrum for Paleozoic carbonates (James & Jones, 2016). This problem outlines the uncertain relationship between reef builders and photosynthetic symbionts during the Paleozoic. James & Wood (2010) clearly identify the dominant reef builders during the Ediacaran to Cambrian, including stromatolites, thrombolites, calcimicrobes, and archaeocyaths. Despite the confident identification of carbonate producers, James & Jones (2016) described an unknown relationship between these producers and photosynthetic symbionts in spite of exhaustive geochemical analysis. Therefore, it is possible that Ediacaran and Cambrian reef builders were not limited to the photic zone and they may not be indicative of strictly shallow marine paleoenvironments. The spatial distribution of carbonates during the late Precambrian is observed by Li et al. (2013), in which they primarily occur in equatorial to low paleolatitudes. These authors suggested that this provides support towards indicating warm water conditions through climatic and biological controls. Although the suggestion of primarily warm water production is only one part of the shallow marine warm water carbonate factory defined by Bosence & Wilson (2003), at the very least it does not discount the possibility of symbiotic relationships between the reef builders and photosynthetic organisms at this point in Earth history.

Differentiating shallow marine and terrestrial paleoenvironments relies heavily on the supplementary data compiled. The first criteria for defining terrestrial settings is the presence of continental crust, implied by the PalaeoPlates model. The compiled geologically significant data points are then utilized to distinguish these two paleoenvironments. These data points displayed spatial variability, with some regions containing many data points while other regions lacked data of any kind. This is not unexpected, as Blatt & Jones (1975) observed a lognormal relationship between sedimentary rock outcrop area and corresponding geologic age. These authors describe a decay curve for sedimentary rock outcrop with an approximate half-life of 130 Ma. Further, Holland (2016) also observed the inevitable decline in the probability of rock preservation with age and describes the strong influence of tectonic setting on the increase or decrease of preservation probability. The identification of terrestrial paleoenvironments is provided a slight advantage in that indicators are supplied by all rock types, not just sedimentary rocks as was the case for deep and shallow marine. In defining the paleocoast line, the differentiation between shallow marine and terrestrial paleoenvironments, an effort was made to

infer simplistic interpretations. This effort is backed by the observation that an important quality of a good hypothesis is simplicity (Copi, 1978). Finally, in coastal regions where data points were not present, there was a bias towards the interpretation of shallow marine over terrestrial paleoenvironments.

The last paleoenvironment defined were orogenic belts, based on metamorphic and cooling data points from the compiled dataset. As orogenic belts could also be referred to as terrestrial highlands, much of the same criteria for defining terrestrial paleoenvironments is applied to mountains. The presence of continental crust inferred by the PalaeoPlates model was required. This was followed by the presence of cooling data points, as exhumation is associated with mountain building processes. Finally, metamorphic data points were utilized as they indicate crustal thickening, associated with plate collisions.

With clearly defined criteria for determining paleoenvironments and the compilation of supplementary datasets, the process of digitizing new polygon geometries was conducted using GPlates. The large volume of data from disparate sources often presented data conflicts, typically in the form of inconsistent paleoenvironmental indicators from geological map polygons and geologically significant data points at a locality. Where these situations were encountered, preference was given to the data with a more certain temporal range. This was frequently the case between polygons defining shallow marine paleoenvironments and data points defining terrestrial paleoenvironments (e.g., cooling and crystallization points). In these instances, an effort was made to produce a coastline that was as simple as possible yet captured as much detail from the data points as possible. There were also situations where map polygon data resulted in complicated divisions between paleoenvironments. An example can be observed in Figure 6.4, where the present-day Arctic of North America during the Terreneuvian displays a complex array of shallow and deep marine settings. This is a result of the small scale in which the geological map was produced, ultimately leading to polygons representing wider geological times than is desired for this type of interpretation. To rectify this, additional definition queries were run in ArcGIS to further restrict map polygons to the time frame of interest (e.g., eliminating polygons that were defined as Cambrian – Devonian and rather only selecting polygons that were strictly defined as Cambrian). The restriction of the geological map data to

polygons with well-defined age limits produced results that were less confusing and easier to work with.

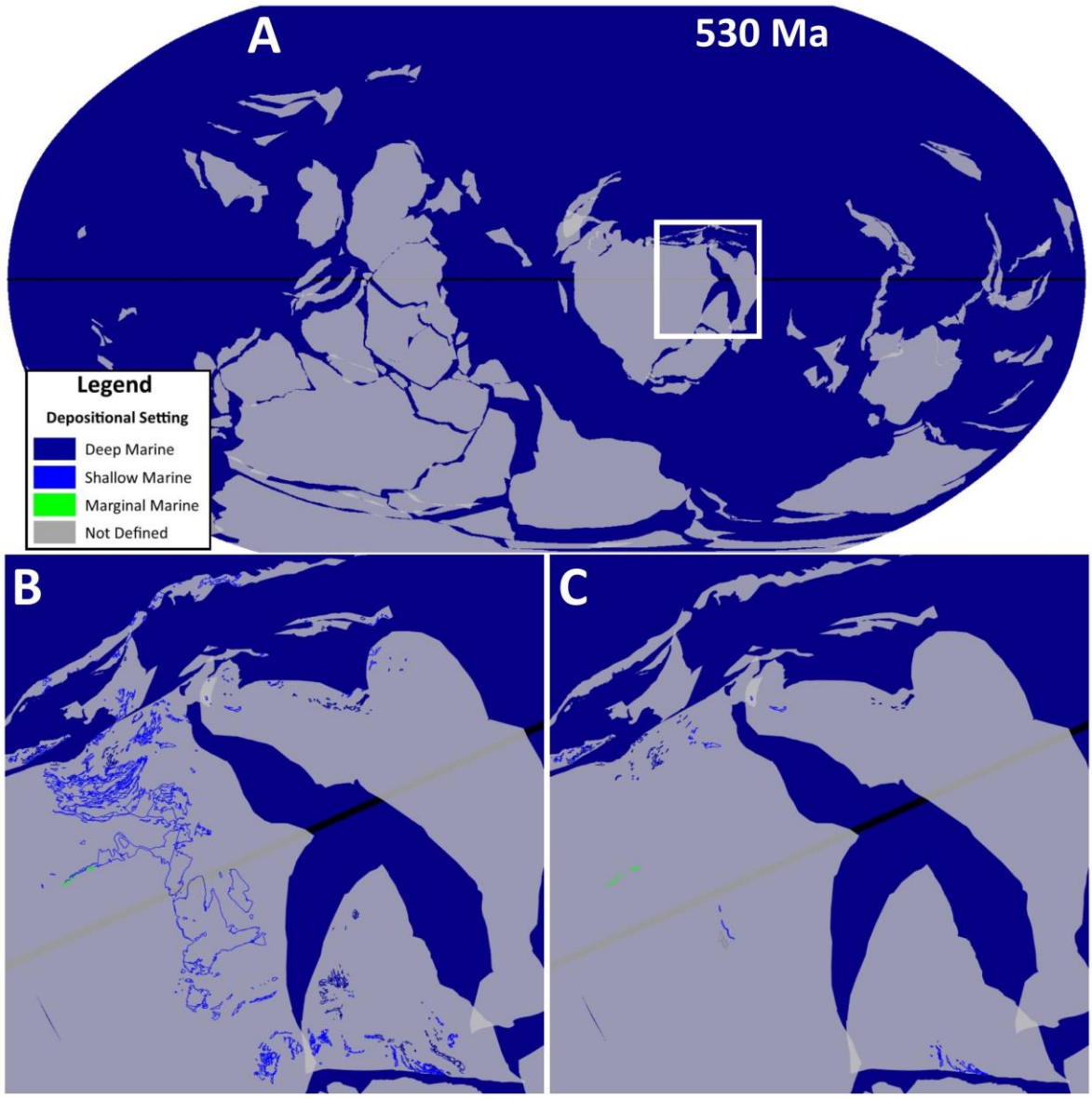


Figure 6.4 (A) Global paleogeographic reconstruction created within GPlates using the PalaeoPlates Model (Eglington, 2018c) at 530 Ma with a Robinson projection. The white box displays the region on Laurentia which the following images are zoomed in on. **(B)** Northern margin of Laurentia (in respect to the current geographic setting of North America) with the Arctic geological map (Harrison et al., 2011) displaying all polygons that have been defined with Cambrian as a maximum age. **(C)** Northern margin of Laurentia with the same geological map; however, polygons have been restricted to those that list Cambrian as a maximum and minimum age.

The final addition to the paleogeographic reconstructions were the climatic zones defined by Boucot et al. (2013). This work was limited to the Phanerozoic, ruling out the inclusion of climatic zones for the Ediacaran reconstruction. Although this study does not use the same plate tectonic model, comparing results at similar time intervals suggests that they are broadly similar. Figure 6.5 shows a comparison of the climatic reconstruction during the Early Cambrian created by Boucot et al. (2013), followed by the same data plotted with the PalaeoPlates Model. It should be observed that the time intervals are offset by 10 Ma, with the Boucot et al. (2013) reconstruction representing 540 Ma, while the PalaeoPlates reconstruction represents 530 Ma. The major difference between the two plate tectonic models is the placement of Laurentia and Siberia, with the PalaeoPlates model placing these continents in a more northerly direction. Despite these differences, the data points plot roughly in similar locations, and thus it is possible to digitize the climatic zones defined by Boucot et al. (2013) onto the PalaeoPlates Model. Boucot et al. (2013) observe that paleoclimate data during the early Paleozoic is limited, making their interpretations highly provisional.

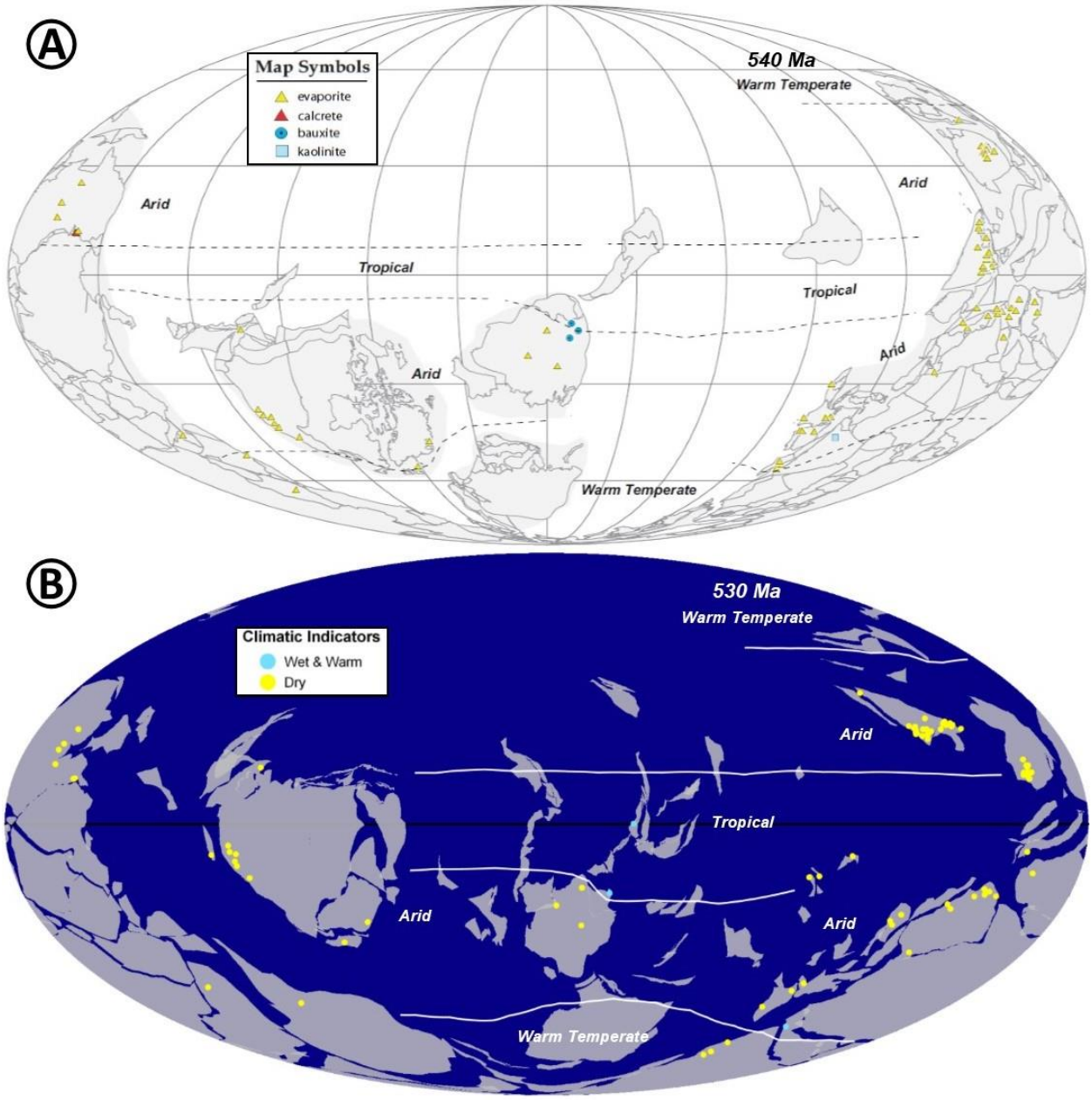


Figure 6.5 (A) Paleoclimatic interpretation during the early Cambrian (540 Ma), created by Boucot et al. (2013). **(B)** PalaeoPlates Model with the same data used by Boucot et al. (2013) during the early Cambrian (530 Ma). Although the plate tectonic models are offset by 10 Ma, it is possible to digitize the interpretations suggested by Boucot et al. (2013) for use in GPlates with the ichnology data.

6.1.3 Visual Analysis Results

6.1.3.1 Ediacaran

The Ediacaran marks the final stages in the assembly of the supercontinent Gondwanaland (Li et al., 2013). The continental configuration according to the PalaeoPlates Model at this time is displayed in Figure 6.6, illustrating the near complete assembly of Gondwanaland, with parts of present-day Australia, China, and Antarctica yet to collide with the rest of the supercontinent. Li et al. (2013) indicated that the final assembly of Gondwanaland does not occur until approximately 540 Ma, an additional 10 Ma from the time of the paleogeographic model displayed in Figure 6.6 B. In addition to the final assembly of this supercontinent, the PalaeoPlates Model places much of the continental crust within the southern hemisphere at 550 Ma. This includes Gondwanaland, composed of present-day Africa, South America, parts of Asia, and Antarctica. North America was not a part of Gondwanaland, although the model used suggests that it too is located in the southern hemisphere, in close proximity to the supercontinent.

Following the methods described in the previous section of this chapter, the compiled supplemental data was used to infer paleoenvironments across the globe during the Ediacaran (Figure 6.7). The Ediacaran has been observed as a period of low global sea level (Haq & Schutter, 2008), which was taken into account when inferring paleoenvironments. Despite this observation, the paleogeographic reconstruction was still biased towards shallow marine paleoenvironments in areas that lacked data points (e.g., Siberia). During the formation of Gondwanaland it has been observed that large orogenic belts formed across much of present-day Africa, commonly referred to as the "Pan-African Orogens" (Li et al., 2013). This observation is confirmed with igneous, metamorphic, and cooling data extracted from DateView. Finally, the abundance of intrusive igneous and metamorphic geochronology data long the eastern margin of the Australian, South China, and Eastern Antarctica paleocontinent is used to infer the existence of a subduction zone. This inference is supported by Cawood (2005), who described the initiation of the Terra Australis Orogen along these plates. Additional subduction zones are inferred from the supplementary data along the western margin of Siberia, and the northern margin of Gondwanaland (Figure 6.8).

The Ediacaran paleogeographic reconstructions lack paleoclimate and paleocurrent data, information that is provided in the Cambrian reconstructions of this section. As previously described, the lack of paleocurrent data during this time is due to the overly broad age category applied to information collected by Brand et al. (2015) during the Precambrian. The paleoclimate data extracted from Boucot et al. (2013) is not included in this reconstruction, as the interpretations provided by these authors are limited to the Phanerozoic. It is observed that the interpretations made by Boucot et al. (2013) during the earliest Cambrian are based on the lack of evidence indicating glaciations, and the large regions of siliciclastic sediments containing unweathered detrital mica flakes. With the last Ediacaran glaciation (i.e. the Gaskiers glaciation) recorded at roughly 580 Ma (Pu et al., 2016), it does not seem unreasonable to extrapolate the interpretations made by Boucot et al. (2013) into the late Ediacaran. Therefore, the climatic gradient is likely moderate with extensive temporal regions at this time.

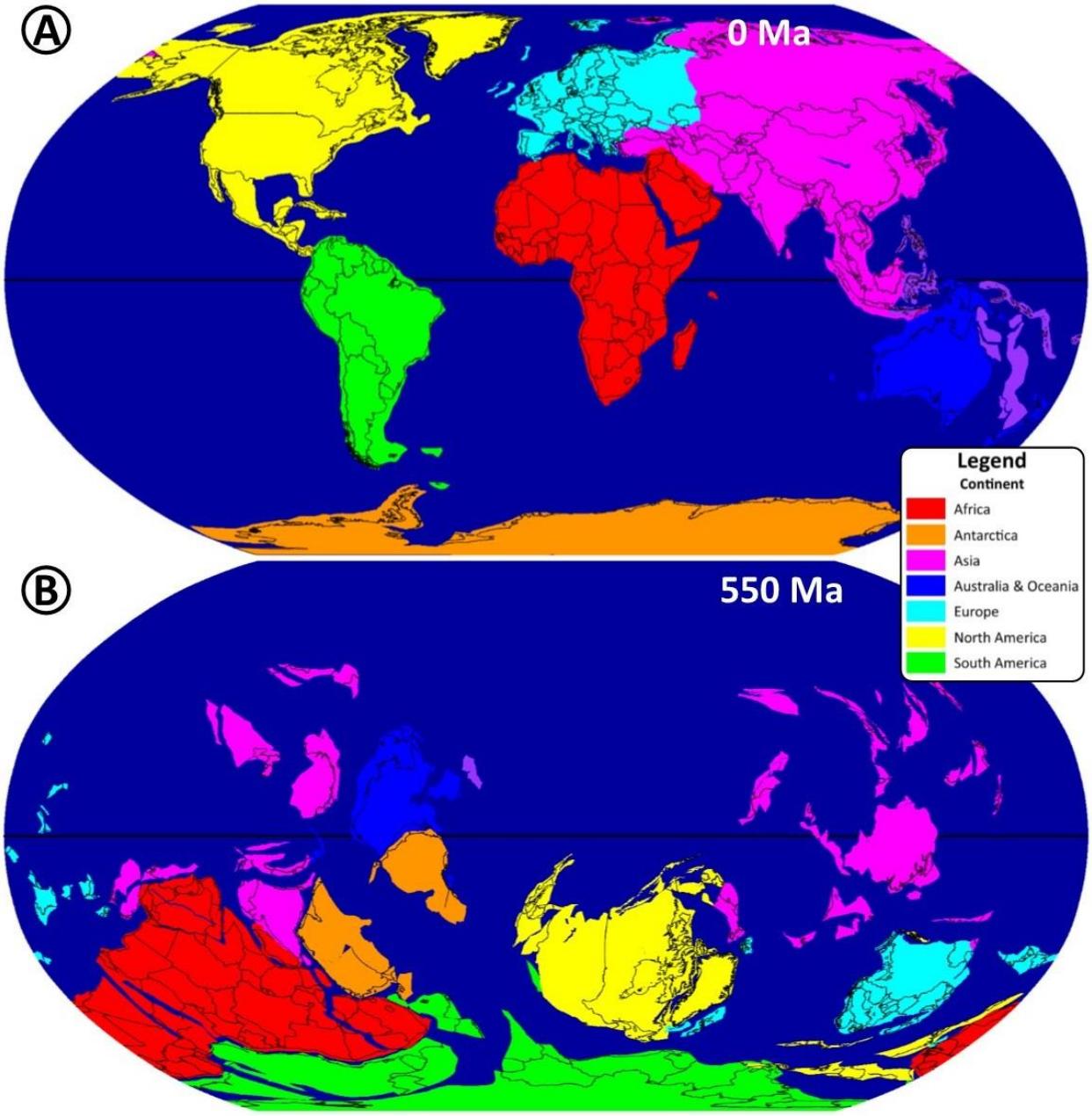


Figure 6.6 (A) Present-day geographic distribution of continental crust in the PalaeoPlates Model. Continents are colour coded to provide assistance in determining their placement in the paleogeographic reconstructions. **(B)** Paleogeographic reconstruction of the Ediacaran in GPlates using the PalaeoPlates Model.

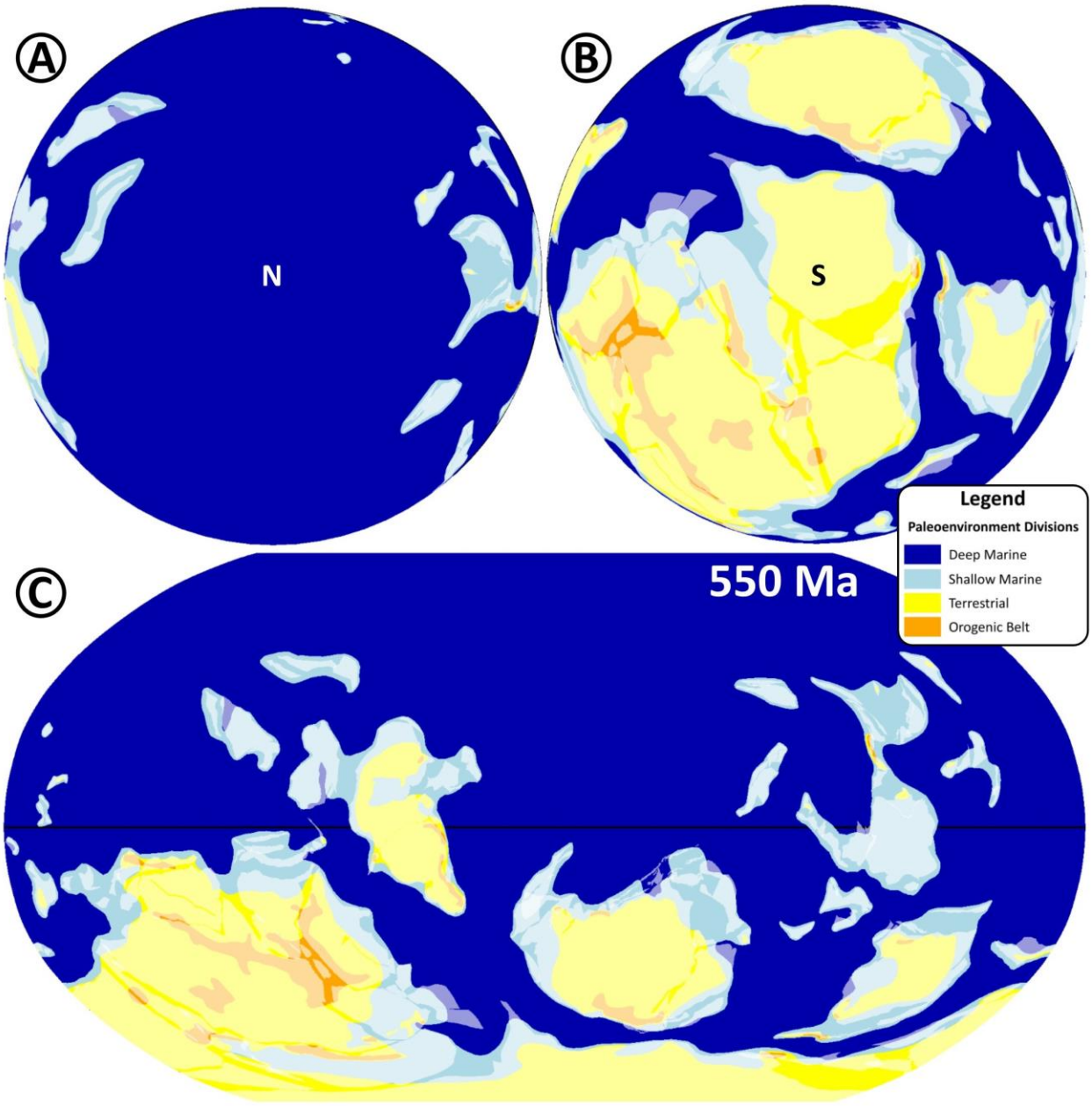


Figure 6.7 Ediacaran paleogeographic reconstruction using the PalaeoPlates Model in GPlates. The continental plates of the PalaeoPlates Model are faintly visible behind the inferred paleoenvironments. **(A)** North pole stereographic projection with an "N" indicating the approximate position of the pole. **(B)** South pole stereographic projection with an "S" indicating the approximate position of the pole. **(C)** Robinson global projection, roughly centered around Laurentia.

With the newly created paleogeographic reconstructions, the ichnology data extracted from IchnoDB was plotted for visual analysis (Figure 6.8). The first Ediacaran map analyzed (Figure 6.8 A) displays the recorded trace fossil occurrences and their generalized depositional environments, to which there is no obviously apparent pattern or trend. There is a relatively even distribution of depositional environments among the trace fossil occurrences, with the exception of marginal marine settings. The restricted level of ichnological activity in marginal marine environments during the Ediacaran could be accounted for by several factors. The first of which is the small number of guilds observed among Ediacaran biota, all of which were closely related to the vast distribution of microbial mats (Seilacher, 1999; Mángano & Buatois, 2017). In addition to the lifestyles present in the Ediacaran, trophic columns lacked the presence of predators (Seilacher, 1999; Xiao & Laflamme, 2009), resulting in simplified ecological interactions (Mángano & Buatois, 2017). Considering these two factors, perhaps the expansion into increasingly stressful ecospace was simply unnecessary. Alternatively, prior to the "Fortunian diversification event" (Mángano & Buatois, 2014), it is possible that the Ediacaran biota lacked the ability to occupy marginal marine environments.

Beyond a lack of observable patterns or trends between trace fossil occurrence and depositional environments, it is noted that there are very few trace fossil occurrences during the Ediacaran. This is likely a product of sessile organisms dominating the Ediacaran biota (Xiao & Laflamme, 2009), simple behavioral diversity prior to the Fortunian diversification event (Mángano & Buatois, 2014), and the restriction of Ediacaran biota to the water-sediment interface (Seilacher, 1999). Despite the limited number of trace fossil occurrences during the Ediacaran, their spatial distribution indicates that there were no climatic or paleolatitudinal restrictions. Trace fossils are observed throughout the southern hemisphere, including records in Avalonia (i.e. present-day Newfoundland) which is located relatively close to the South pole in the PalaeoPlates Model. The global distribution of trace fossils supports the observations of Xiao et al. (2016), where they observed that Ediacaran biota are reported from every continent except Antarctica.

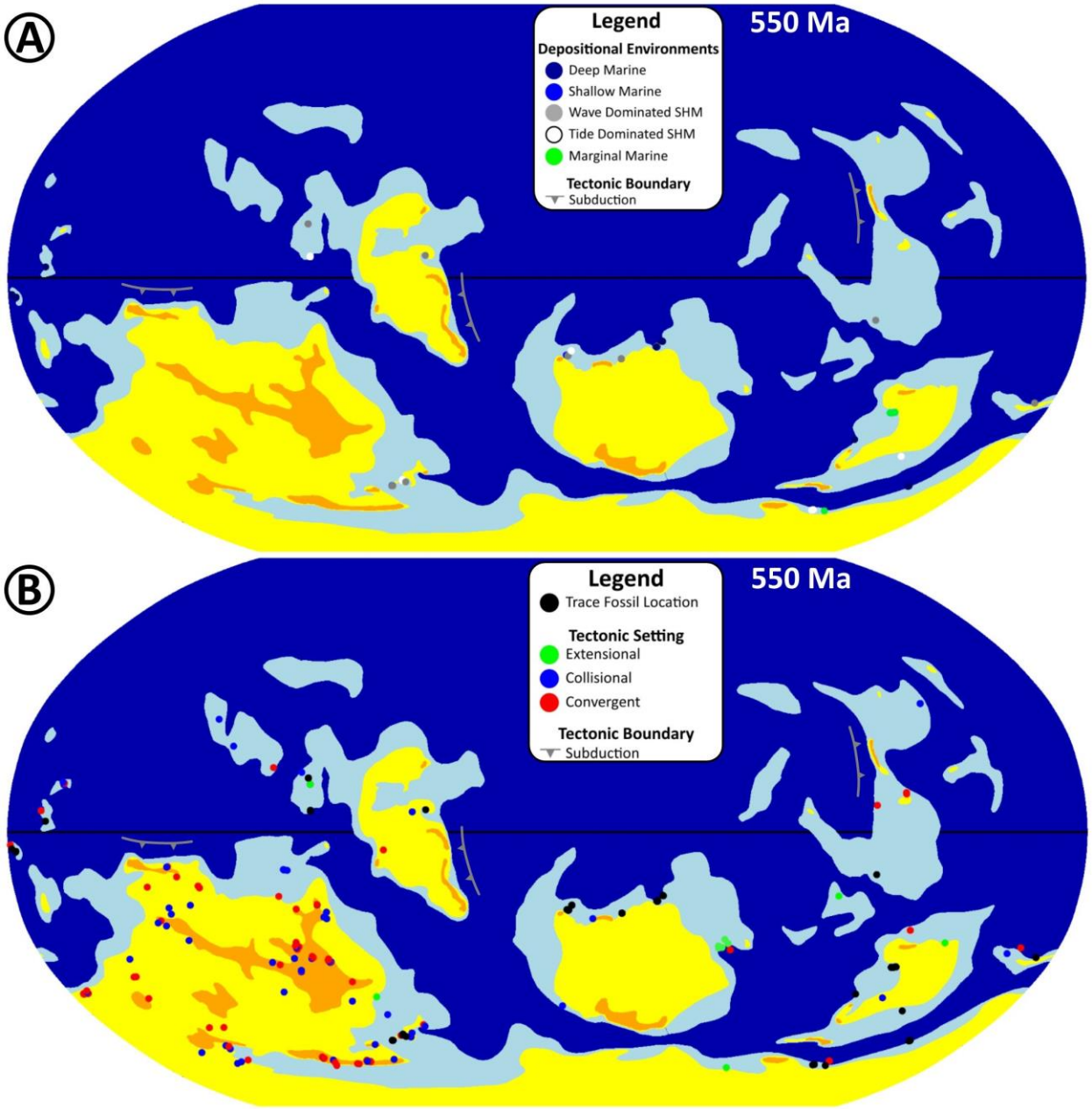


Figure 6.8 Ediacaran paleogeographic reconstruction using the PalaeoPlates Model in GPlates, with a Robinson global projection centered around Laurentia. Inferred plate tectonic boundaries are included where supplementary data implies their existence. Refer to Figure 6.7 for a legend of the paleoenvironments displayed. **(A)** All Ediacaran ichnology data extracted from IchnoDB is displayed, with colour symbology representing generalized depositional environments. **(B)** All Ediacaran ichnology data extracted from IchnoDB is displayed in conjunction with detrital zircon data. Inferred tectonic settings into which the detrital zircon grains were deposited are displayed to illustrate geodynamic settings.

The inclusion of detrital zircon data in conjunction with the ichnology data was plotted in Figure 6.8 B to observe global geodynamics during the Ediacaran. As previously described, fossil preservation is greatly influenced by the tectonic setting in which a sediment basin in created and as a result there exists a preservation bias (Holland, 2016). Considering this observation, trace fossil localities will occur more frequently near extensional tectonic settings; however, that is not substantiated in this figure. Rather, there is an inordinate number of inferred convergent and collisional tectonic settings associated with the trace fossil localities in the Ediacaran reconstruction. This unexpected result can be explained by poorly defined deposition ages applied to the detrital zircon data extracted from Puetz et al. (2018). Without accurately defined deposition ages, there is a two-fold issue in using this data. The first being a bias towards convergent and collisional settings when using the method described by Cawood et al. (2012) to infer geodynamic environments. The second problem is that GPlates requires a time frame in which to display data, in this case that is defined by these poorly estimated deposition ages. This leads to biased inferences that quite possibly should not be displayed during this specific time period of interest.

Considering this dilemma, the detrital zircon data extracted from Puetz et al. (2018) is misleading, and analysis based on these inferences would be highly unreliable. The remaining detrital zircon data extracted from a meticulously compiled dataset described in chapter 4 of this thesis does not exhibit the same issue. Although this dataset displays limited occurrences during this time period, precluding global analysis of trace fossil occurrence in association with tectonic settings. Apart from the tectonic settings inferred from the detrital zircon, the supplemental data extracted from DateView provided indicators of subduction zones (Figure 6.8 B). Where these zones are present there is a distinct lack of trace fossil occurrences, which provides support to the preservation bias as subduction zones produce convergent and collisional settings.

6.1.3.2 Terreneuvian

The Terreneuvian represents the complete assembly of Gondwanaland (Li et al., 2013), with the amalgamation of the Australian, South China, and East Antarctic paleocontinent onto the supercontinent (Figure 6.9). Another significant paleogeographic change suggested by the PalaeoPlates Model is the placement of Siberia much closer to Baltica (i.e. much of present-day Europe), suggesting a collision between the two paleocontinents. Finally, Laurentia (i.e. present-day North America) is placed in a more northerly equatorial position when compared to the Ediacaran paleogeographic configuration (Figure 6.6 B).

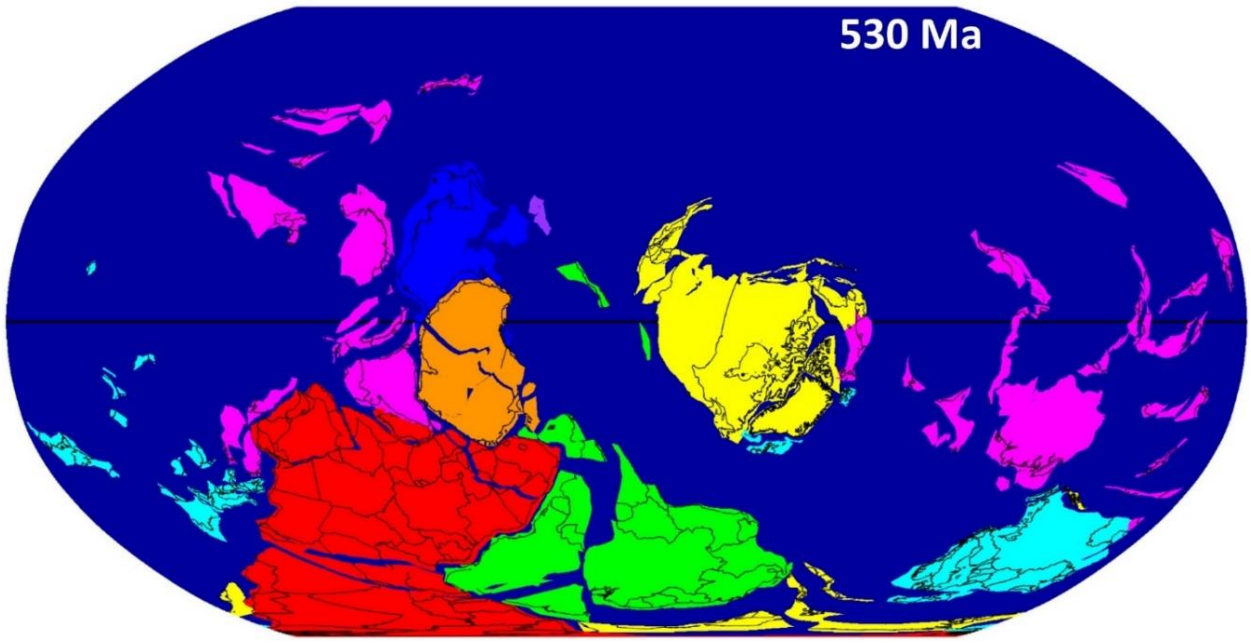


Figure 6.9 Global configuration of continental plates during the Terreneuvian according to the PalaeoPlates Model. Refer to Figure 6.6 for the legend regarding the colour of the continents.

The inference of paleoenvironments on the PalaeoPlates Model from the compiled supplementary material during the Terreneuvian is displayed in Figure 6.10. This figure displays the now fully formed Gondwanaland, and it is inferred that Siberia and Baltica are connected to the supercontinent through shallow marine paleoenvironments. It can also be observed that there is an increase in the occurrence of deep marine settings on continental plates (e.g., see Baltica, South China, and South Africa). This inference is supported by Haq & Schutter (2008), where they concluded that global sea level continued to rise through the Ediacaran into the Terreneuvian. In addition to sea-level rise, supplementary data suggests that the Pan-African Orogens created during the formation of Gondwanaland continued to expand upon the complete assembly of this continent. Finally, the subduction zone associated with the Terra Australis Orogen during the Ediacaran is inferred to have spread across the entire margin of Gondwanaland. Cawood (2005) proposed that this orogenic belt spanned 18,000 km along the entire northern margin of the supercontinent. The PalaeoPlates Model suggests that the expansion of this plate tectonic boundary is joined by sea floor spreading between Gondwanaland and Laurentia, producing the northerly drift of Laurentia away from Gondwanaland.

Visual analysis for the Terreneuvian was conducted with the paleogeographic reconstruction created and the ichnology data extracted from IchnoDB, (Figure 6.11). Supplementary paleoclimatic zones described by Boucot et al. (2013) and paleocurrent data extracted from Brand et al. (2015) were also included in this figure. As with the Ediacaran visual analysis, Figure 6.11 A does not show any clear patterns or trends observed between the spatial array of trace fossils and the depositional environments in which they were produced. There appears to be an even distribution of depositional environments among the trace fossil occurrences plotted, with an increased number of occurrences in marginal marine environments compared to the Ediacaran. This observation is in all likelihood attributed to the increased behavioral complexity and body plan diversification observed through the Terreneuvian (Mángano & Buatois, 2014), and the onset of the Agronomic Revolution (Seilacher & Pflüger, 1994; Seilacher, 1999; Mángano & Buatois, 2014, 2017). Mángano & Buatois (2014) observed that these factors led to trophic escalation and the expansion of ecospace utilization throughout depositional environments.

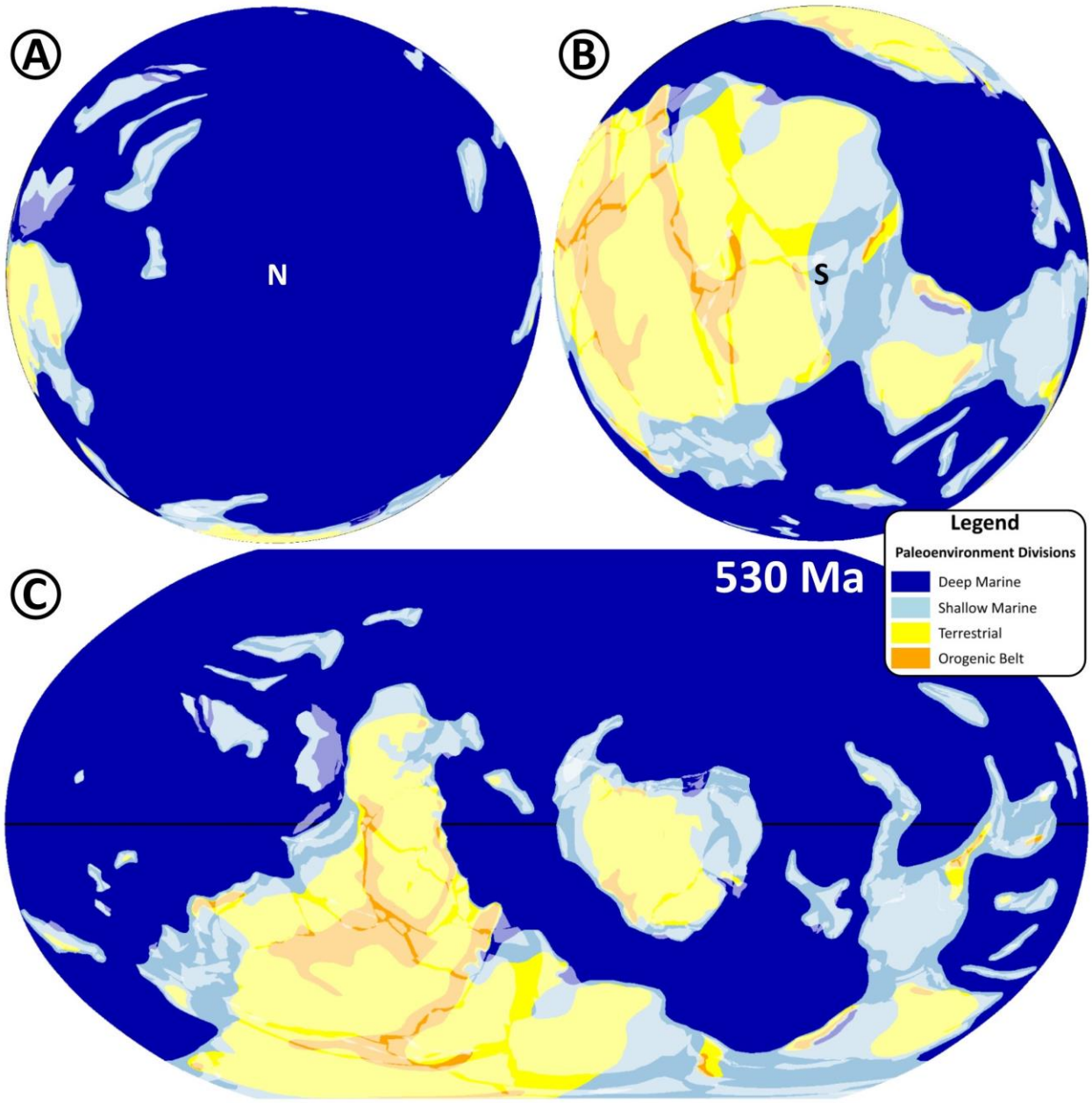


Figure 6.10 Terreneuvian paleogeographic reconstruction using the PalaeoPlates Model in GPlates. The continental plates of the PalaeoPlates Model are faintly visible behind the inferred paleoenvironments. **(A)** North pole stereographic projection with an "N" indicating the approximate position of the pole. **(B)** South pole stereographic projection with an "S" indicating the approximate position of the pole. **(C)** Robinson global projection, roughly centered around Laurentia.

In addition to environmental expansion, Figure 6.11 A displays a global distribution of trace fossil occurrences. There is no restriction to specific paleolatitudes, with localities spanning from equatorial to polar southern paleolatitudes. Further, no apparent paleoclimatic restrictions exist, with trace fossil occurrences in all of the paleoclimates identified by Boucot et al. (2013). The global distribution of trace fossils observed during the Terreneuvian is a continuation of the distribution observed during the Ediacaran, with recorded occurrences spread over greater paleocontinental areas and on more island arcs.

The inclusion of geodynamic settings in association with trace fossil occurrence (Figure 6.11 B) was included for the Terreneuvian visual analysis to maintain consistency, despite the questionable reliability of this data. There remains a significantly higher number of convergent tectonic settings than would be expected as a result of the preservation bias. Although there are a much more reasonable number of extensional settings, where significant groups of trace fossils are recorded (e.g., Laurentia, Baltica). As observed during the Ediacaran, there are no recorded trace fossil occurrences where subduction zones have been inferred. Possible exceptions to this observation are trace fossils located inland from the Terra Australis subduction zone along the eastern margin of Gondwanaland. In this example, perhaps they represent preservation in foreland or back arc basins that typically have low preservation potential. Finally, where paleocurrent data exists on Gondwanaland and Laurentia there is no obvious preference or exclusion of trace fossil occurrence. As trace fossils are located near major drainage systems (e.g. northern Laurentia) and absent from them (e.g., western Gondwanaland), it is difficult to draw any significant conclusions.

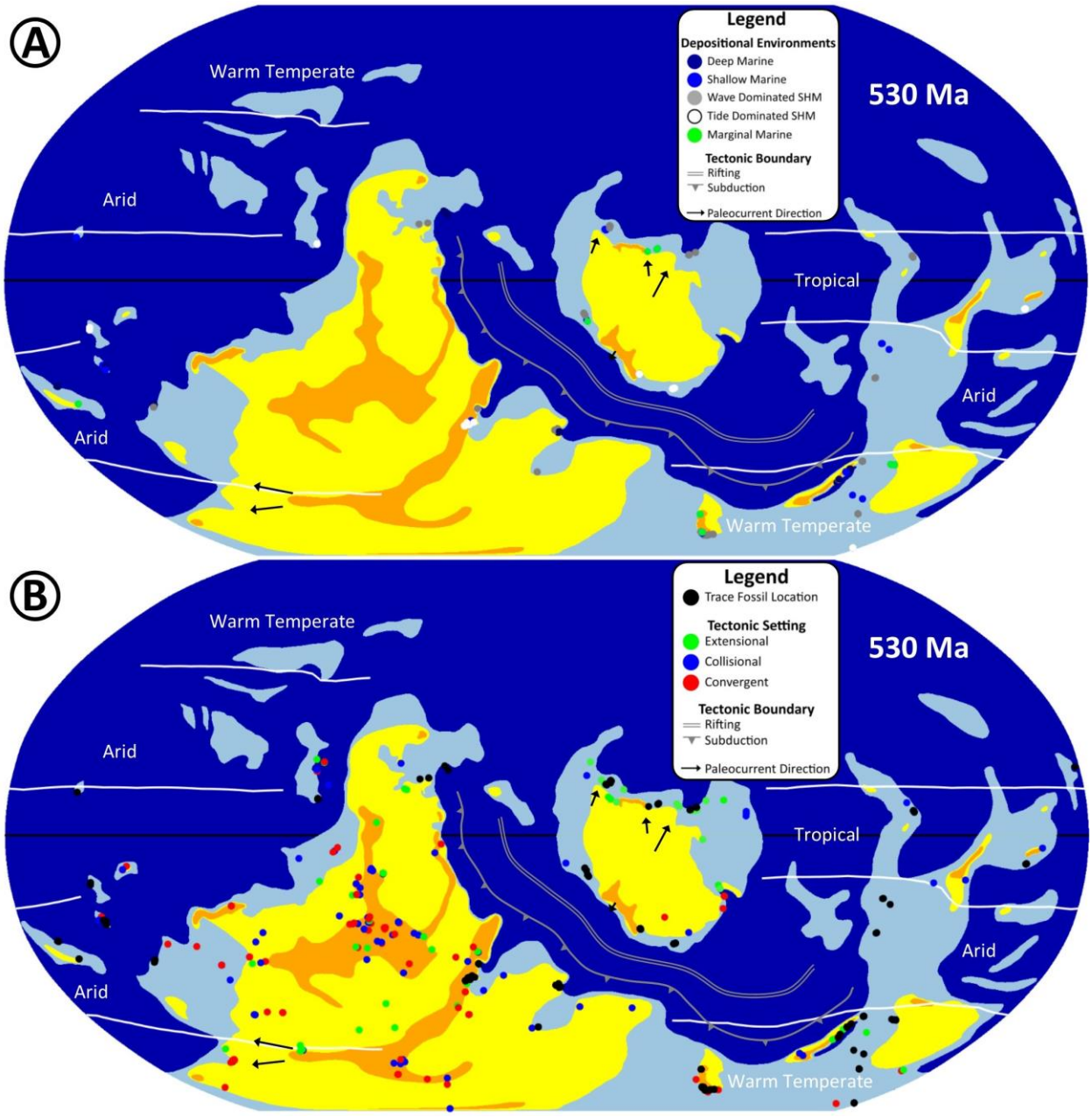


Figure 6.11 Terreneuvian paleogeographic reconstruction using the PalaeoPlates Model in GPlates, with a Robinson global projection centered around Laurentia. Paleoclimatic zones inferred by Boucot et al. (2013) are included. Where available, paleocurrent data from Brand et al. (2015) is generalized and plotted. Inferred plate tectonic boundaries are included where supplementary data implies their existence. Refer to Figure 6.10 for a legend of the paleoenvironments displayed. **(A)** All Terreneuvian ichnology data extracted from IchnoDB, with colour symbology representing generalized depositional environments. **(B)** All Terreneuvian ichnology data extracted from IchnoDB is displayed in conjunction with detrital zircon data. Inferred tectonic settings into which the detrital zircon grains were deposited are displayed to illustrate geodynamic settings.

6.1.3.3 Cambrian Epoch 2

The Cambrian Epoch 2 displays a very similar continental configuration to that of the Terreneuvian (Figure 6.12). Notable alterations to the placement of paleocontinents according to the PalaeoPlates Model includes an increase in the separation between the western margin of Laurentia and the northern portion of Gondwanaland compared to the Terreneuvian. In addition to this, the model suggests that all continental plates are generally drifting south, increasing the previously significant concentration of landmass in the southern hemisphere. Beyond these two observations, there appears to be relative stability in the position of the continents in relation to each other during this time.

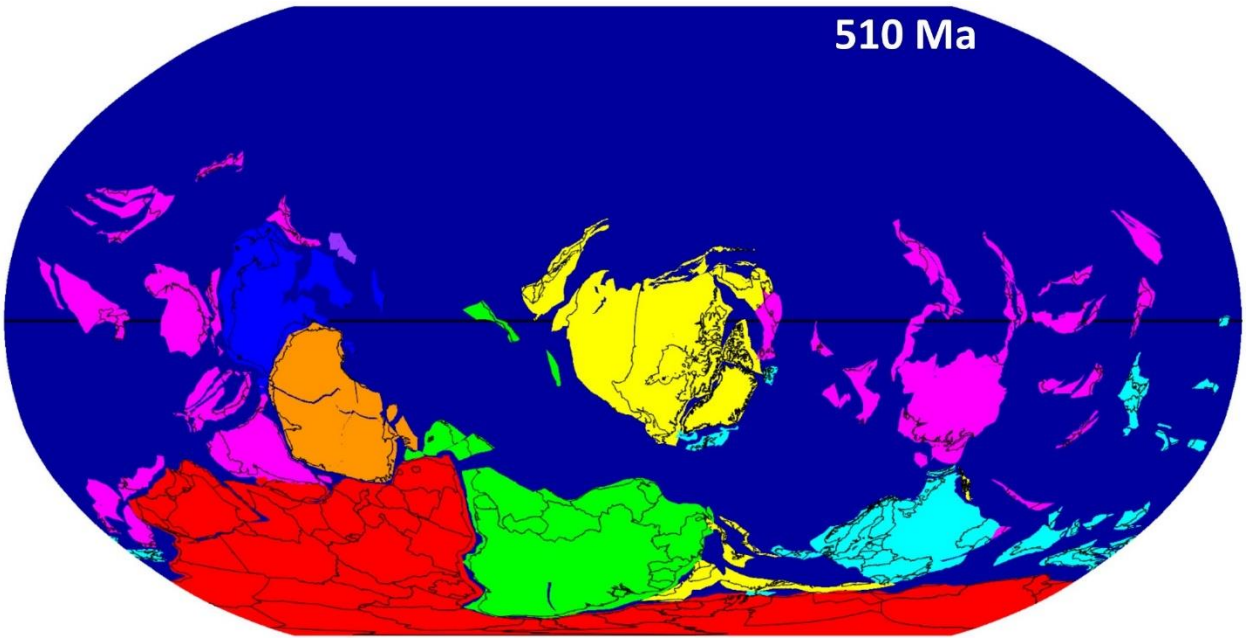


Figure 6.12 Global configuration of continental plates during the Cambrian Epoch 2 according to the PalaeoPlates Model. Refer to Figure 6.6 for the legend regarding the colour of the continents.

The determination of paleoenvironments on the PalaeoPlates Model using the compiled supplementary material during the Cambrian Epoch 2 is displayed in Figure 6.13. Compared to the previous paleogeographic reconstructions, the Cambrian Epoch 2 is inferred to have the largest amount of deep marine paleoenvironments within continental plate boundaries. Further, shallow marine paleoenvironments are much more widespread, specifically on Laurentia, and the northern and western margins of Gondwanaland. This observation is again supported by the global sea level curve reported by Haq & Schutter (2008), which shows an increase in sea level throughout the Cambrian. To further supplement the inference of increased shallow marine paleoenvironments, the PBDB dataset provides thousands of additional points as a result of the Cambrian Explosion. With the lack of terrestrial expansion at this time, these paleontologic data points provide ample support for rising sea level. Finally, the Terra Australis Orogen (Cawood, 2005) and its related subduction zone is supported by the data extracted from DateView, with new points suggesting orogenic activity along the margin of present-day South America. The PalaeoPlates Model suggests slightly more complex plate boundaries than the simple ocean floor spreading inferred during the Terreneuvian. During the Cambrian Epoch 2, rifting between Laurentia and Gondwanaland is combined with significant segments of transform boundaries, and the addition of another spreading center between Laurentia and Siberia.

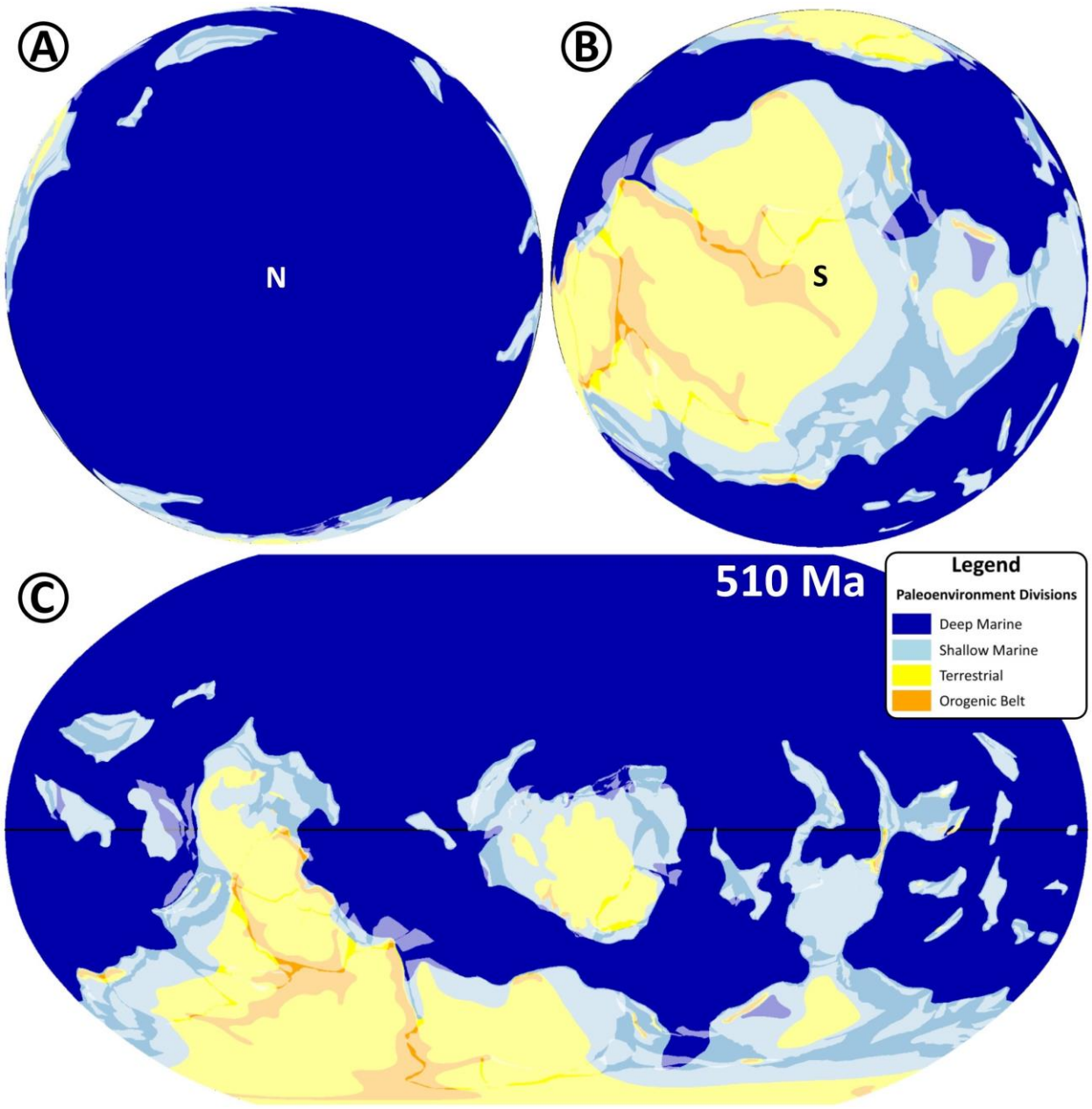


Figure 6.13 Cambrian Epoch 2 paleogeographic reconstruction using the PalaeoPlates Model in GPlates. The continental plates of the PalaeoPlates Model are faintly visible behind the inferred paleoenvironments. **(A)** North pole stereographic projection with an "N" indicating the approximate position of the pole. **(B)** South pole stereographic projection with an "S" indicating the approximate position of the pole. **(C)** Robinson global projection, roughly centered around Laurentia.

Visual analysis of Cambrian Epoch 2 ichnology data extracted from IchnoDB was conducted with the aid of the paleogeographic reconstruction created (Figure 6.14). As with the Terreneuvian visual analysis, paleoclimatic zones and paleocurrent data were included to supplement pattern recognition. Similar to both the Ediacaran and Terreneuvian, Figure 6.14 A does not display any clear patterns or trends between trace fossil occurrence and the depositional environments in which they were formed. The spatial distribution of trace fossils and variety of depositional environments during the Cambrian Epoch 2 are comparable to that observed during the Terreneuvian, with the only major difference being the increased number of occurrences. As this time period coincides with the Cambrian Explosion, this event may have some influence on the increased number of occurrences. Although, the Cambrian Explosion describes body fossil diversification and therefore does not entirely explain the increased number of trace fossil records. Rather, this increase in trace fossil occurrences is more likely accounted for by new behaviors in deposit feeders and additional infaunal ecospace utilization (Mángano & Buatois, 2014). These authors described the biodeposition of organic material by infaunal filter feeders as the probable cause for addition behavioral diversification during the Cambrian Epoch 2. Finally, it is observed that much like the previous two time periods analyzed, there are no paleolatitudinal or paleoclimatic restrictions of the spatial distribution of trace fossils.

Again, the inclusion of geodynamic settings in association with trace fossil occurrence (Figure 6.14 B) was included in the visual analysis of the Cambrian Epoch 2 to remain consistent. The reliability of this data is still suspect, as there are far more convergent tectonic settings inferred than expected. Laurentia is the exception, with the implied geodynamic settings dominated by extensional settings (e.g., passive margins) associated with the numerous trace fossil occurrences recorded on this continent. In addition, trace fossil occurrences are generally not associated with the inferred subduction zones. The Terra Australis subduction zone is an anomaly, in which case these trace fossil occurrences may represent atypical basin preservation. Finally, the included paleocurrent data does not provide meaningful indication of trace fossil occurrence or absence. Where this information is present there are examples of trace fossil occurrence (e.g., western Gondwanaland) and trace fossil absence (e.g., northern Laurentia), providing no basis for meaningful conclusions.

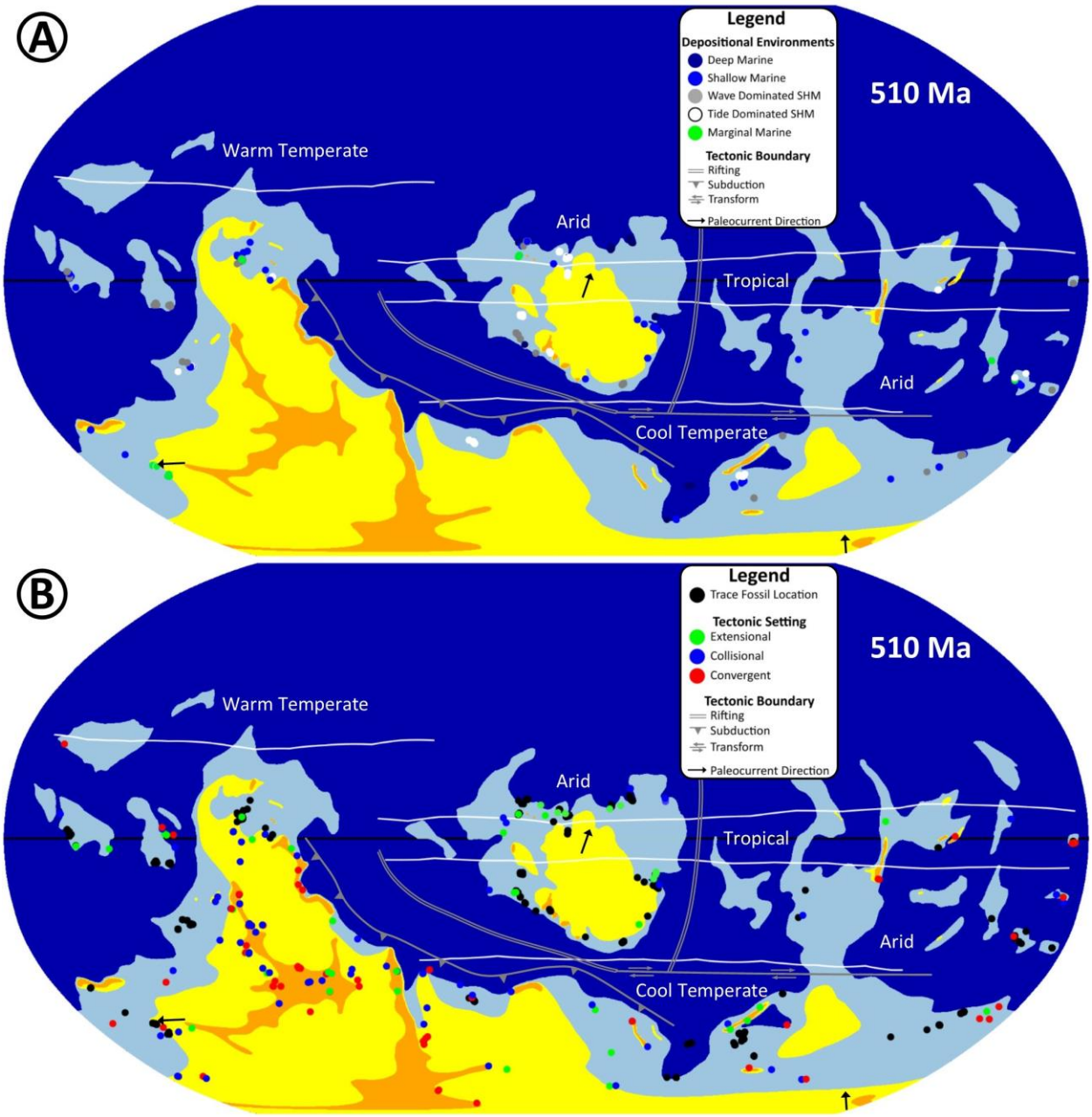


Figure 6.14 Cambrian Epoch 2 paleogeographic reconstruction using the PalaeoPlates Model in GPlates, with a Robinson global projection centered around Laurentia. Paleoclimatic zones inferred by Boucot et al. (2013) are included. Where available, paleocurrent data from Brand et al. (2015) is generalized and plotted. Inferred plate tectonic boundaries are included where supplementary data implies their existence. Refer to Figure 6.13 for a legend of the paleoenvironments displayed. **(A)** All Cambrian Epoch 2 ichnology data extracted from IchnoDB, with colour symbology representing generalized depositional environments. **(B)** All Cambrian Epoch 2 ichnology data extracted from IchnoDB is displayed in conjunction with detrital zircon data. Inferred tectonic settings into which the detrital zircon grains were deposited are displayed to illustrate geodynamic settings.

6.2 Statistical Analysis

The other data mining technique to be employed in the investigation between depositional environments and trace fossil occurrence is correspondence analysis. This statistical technique was included as it provides a more quantified approach to data mining than that of the previously investigated visual analysis. Correspondence analysis (CA) is a type of ordination, a mathematical technique used to describe underlying structure and patterns within multivariate data (Davis, 2002; Shaw, 2003; Legendre & Legendre, 2012). The selection of CA for statistical analysis of ichnotaxonomic data was based on this method's use of the chi-squared (χ^2) distance, which is capable of processing non-parametric data types such as nominal data (Davis, 2002). Another beneficial characteristic of the chi-squared distance is that it is not influenced by joint absences, an important requirement for species abundance analysis (Borcard et al., 2011).

To conduct correspondence analysis, a contingency table must be created from a dataset in which rows (*r*) represent observations and, columns (*c*) represent variables (Davis, 2002). This author observed that CA assumes $r > c$, a condition that must be accounted for with the creation of a contingency table from the ichnology data. In this study, rows (or 'observations') are represented by depositional environments and columns (or 'variables') are represented by trace fossils.

The use of CA is advantageous in that it is a R-mode statistical analysis measuring the dependence between variables based on the correlation coefficient, while also solving a Q-mode problem in the comparison between objects (Davis, 2002; Borcard et al., 2011). As a result of this characteristic, the ordination of both observations and variables can be plotted simultaneously, producing a biplot (Legendre & Legendre, 2012). The subsequent biplots may reveal structure in the data that would not have been evident by displaying patterns of correlations between variables or relationships between observations (Greenacre, 2010). The interpretation of CA biplots is relatively straightforward in that variables with large contributions to the position of the observations will be found in close proximity on the biplot (Quinn & Keough, 2002).

Legendre & Legendre (2012) described a significant issue in conducting CA where the chi-squared (χ^2) distance is heavily influenced by rare species. It is observed that these occurrences contribute little to the first few ordination axes; however, they are found on the

periphery of biplots which leads to the false sense of importance. Legendre & Legendre (2012) emphasize that CA is meant to display the main axis of variation in the dataset, and it does not deal with exceptions. These authors suggested that it is best to simply remove the rare species before proceeding with CA. An additional method to define and describe outliers created as a byproduct of ordination techniques is outlined by Hubert & Engelen (2004). The multiple ordinations in the following section only observe "good leverage points" according to the classification scheme created by these authors. Outliers of this type are described as occurring close to the axes defined by the ordination, but far from the regular 'cloud' of data.

As the purpose of conducting ordination is to display the major features of a dataset along a reduced number of axes (Borcard et al., 2011), it is often unclear how many of the resulting axes are significant. These authors described three methods in determining how many axes represent significant variation of the data and which are likely reflecting random variance. The first method described is only applicable to ecology, where eigenvalues over 0.6 are retained, as this value has been defined as the cut-off for indicating a very strong environmental gradient in the data. The second method is known as the broken stick model, where a 'stick' with a length equal to the total inertia is randomly divided into a quantity equal to the number of eigenvalues. These pieces are then arranged in decreasing length next to the corresponding eigenvalue and only axes with an eigenvalue larger than that of the piece of stick are kept. Finally, Borcard et al. (2011) described the Kaiser-Guttman criterion, where the mean eigenvalue is calculated from the resulting ordination and only the axes with eigenvalues greater than the mean are retained for interpretation.

In addition to CA, cluster analysis is included in this study, as there is potential for their results to serve as an additional line of support to any patterns observed in the biplots produced. Cluster analysis is a heuristic procedure in which objects or descriptors that are sufficiently similar are grouped together (Borcard et al., 2011; Legendre & Legendre, 2012). There are many types of clustering, and many methods in which to conduct cluster analysis. Legendre & Legendre (2012) underscored the importance of choosing and understanding both the clustering method and the association measure used to base the clustering on. The two types of clustering that are of interest to this study include hierarchical and non-hierarchical. Of the hierarchical methods, complete linkage agglomerative clustering and unweighted pair-group method using

arithmetic averages (UPGMA) appear to be appropriate for the data present in this study. Legendre & Legendre (2012) described complete linkage agglomerative clustering as being desirable in ecology because it helps identify clusters with clear discontinuities. While the second hierarchical method, UPGMA, is described as being effective in forming clusters that contain similar fauna (Wu et al., 2014). The only non-hierarchical method used in this study was K-means, which is described as clustering objects into a user defined number of groups that contain objects that are mathematically more similar to each other than objects of the other clusters (Legendre & Legendre, 2012).

6.2.1 Data Selection and Transformation

Having selected CA as the statistical method of choice due to the dataset characteristics, this analysis will look for potential patterns between trace fossils and depositional environments. In this situation trace fossils will act as the variables, with the assumption that they represent a paleoecological summary, assisting in the definition of distinct depositional environments throughout the Ediacaran to Cambrian transition. Typical paleoecological factors effecting trace fossil distribution include bathymetry, food supply, hydrodynamic energy, oxygen content, salinity, sedimentation rate, substrate coherence and stability, temperature, and turbidity (Frey & Seilacher, 1980; Pemberton et al., 1992a; Brett, 1998; Buatois & Mángano, 2011).

Before conducting CA on the datasets extracted from IchnoDB, two major issues were defined. The first being how trace fossils should be grouped, followed by what time periods should be examined. As was the case with visual analysis, time divisions for CA were based on three observable increases in the ichnodiversity spanning the Ediacaran to Cambrian periods described by Mángano & Buatois (2014). Based on this summary of the Ediacaran to Cambrian transition, age groupings for correspondence analysis were defined as Ediacaran, Terreneuvian, and Cambrian Epoch 2. With the age criteria defined, queries were constructed to extract the data that would be used. To ensure that the CA are representative of the ages defined, trace fossils observed in lithostratigraphic units that span more than one of the defined epochs were not included.

As previously stated, CA requires that $r > c$, which represents an issue with the large number of ichnotaxa recorded in IchnoDB. To address this problem, trace fossils must be

grouped in a meaningful way to reduce the number of variables found within the contingency tables used in CA. There are two metrics that appear to be appropriate for this investigation, ichnodisparity and ichnodiversity. Ichnodiversity is described as representing the quantity of ichnotaxa present (Buatois & Mángano, 2011). Conversely, ichnodisparity groups trace fossils into categories of architectural design, a method that is defined by the trace fossil overall morphological plan (Buatois & Mángano, 2011; Buatois et al., 2017). Ichnodiversity was selected for statistical analysis as it has been observed that if used properly, ichnodiversity may display trends along depositional environments (Ekdale, 1988; Buatois et al., 1997; Mángano & Buatois, 2004). Of the three types of ichnodiversity described by Buatois & Mángano (2011), beta ichnodiversity identifies ichnofauna differentiation between environmental gradients. Considering the primary objective of this thesis is the investigation of relationships between depositional environments and trace fossil occurrence, the selection of ichnodiversity over ichnodisparity is the most appropriate. With the selection of ichnodiversity trace fossils were further grouped into ichnogenera, a common practice described by Buatois & Mángano (2011), reducing the number of columns present; however, the columns still far out-number the rows. To further reduce the number of columns, ichnogenera that contain frequency counts of 1 are removed from the contingency table, followed by progressively higher frequency counts until the number of columns removed satisfies the $r > c$ criteria for CA.

Finally, it should be observed that the ichnology dataset used for CA is subject to bias, specifically the classification of ichnotaxonomy and depositional environments. Potential bias within the individual ichnotaxonomic classifications is addressed through the use of ichnogenera rather than ichnospecies, as ichnogenus taxonomic classification is more thoroughly established than at the ichnospecies level (Buatois & Mángano, 2011). In addition, a comprehensive review of Ediacaran and Cambrian published ichnology data by Dr. Gabriela Mángano and Dr. Luis Buatois provides a level of consistency in the ichnotaxonomic classifications recorded in IchnoDB. This review extended to additional details, including the determination of depositional environments within the published material. To further reduce any impact of bias in the classification of depositional environments IchnoDB requires entries to follow defined standards, creating a level of consistency within the digitized data.

6.2.2 Data Mining and Interpretation

With the selection of data and subsequent transformation through the criteria imposed, the final two steps of the knowledge discovery process, data mining and interpretation can be conducted. The multivariate statistical software Coran (Eglington, 2018a) was utilized to produce the CA ordination. This process was straightforward, simply importing the contingency tables and selecting the multivariate analysis desired from a list of options. With the results from this program, Microsoft Excel was used to produce graphs summarizing the analysis.

To conduct cluster analysis, the statistical software R (R Core Team, 2018) was used. The vegan package (Oksanen et al., 2018) describes a process in the decostand() function help page, in which values similar to those of CA can be produced. In this process a contingency table is standardized using the chi-squared method, followed by the creation of a similarity index using Euclidean distances. With these calculated values it is then possible to conduct the three types of Q-mode cluster analysis previously identified as appropriate. This will produce clusters of related depositional environments according to the method utilized. The code used in R studio to conduct this analysis is provided in the supplementary material.

6.2.2.1 Ediacaran Results and Interpretation

The raw data extracted from IchnoDB for correspondence analysis of the Ediacaran is compiled into a contingency table (Table 6.1). This is a small dataset compared to the following Cambrian epochs, owing to the generally infrequent occurrence of trace fossils during this time (Seilacher et al., 2005). This is also a somewhat contentious time period for trace fossils, as several reviews (e.g., Seilacher et al., 2005; Jensen et al., 2006; Mángano & Buatois, 2014) have categorized many previously identified trace fossils as body fossils or microbially induced sedimentary structures (Buatois & Mángano, 2016).

Table 6.1 Contingency table created from Ediacaran trace fossil data extracted from IchnoDB. Abbreviations are as follows: *Archaeonassa* (*Aa*), *Bergaueria* (*Be*), *Gordia* (*Go*), *Helminthoidichnites* (*He*), *Helminthopsis* (*Hl*), *Palaeophycus* (*Pa*), *Torrowangea* (*To*); Deep marine (DM), Deep marine turbidite system (DMTS), Marginal marine deltaic (MMD), Marginal marine deltaic – wave dominated (MMDWD), Marginal marine deltaic – wave dominated delta front (MMDWDDF), Shallow marine (SHM), Shallow marine tide dominated (SHMTD), Shallow marine tide dominated – Intertidal (SHMTDI), Shallow marine tide dominate – supratidal (SHMTDSPT), Shallow marine wave dominated – offshore (SHMWDOF), Shallow marine wave dominated – shelf (SHMWDSH).

	Aa	***Be***	***Go***	***He***	***Hl***	***Pa***	***To***	*row sum*
DM	0	0	0	4	3	8	2	*17*
DMTS	0	0	4	0	0	1	1	*6*
MMD	3	4	0	0	0	0	0	*7*
MMDWD	0	0	2	0	0	2	0	*4*
MMDWDDF	0	3	3	0	0	0	0	*6*
SHM	0	0	0	1	0	2	0	*3*
SHMTD	0	0	4	1	0	5	3	*13*
SHMTDI	0	0	0	0	0	2	1	*3*
SHMTDSPT	0	1	1	0	1	2	2	*7*
SHMWDOF	3	0	0	7	1	3	1	*15*
SHMWDSH	0	0	1	0	0	5	0	*6*
column sum	*6*	*8*	*15*	*13*	*5*	*30*	*10*	***87***

The results of the Ediacaran correspondence analysis performed with Coran are displayed in Figure 6.15. Not all data produced within Coran is retained for analysis, rather the Kaiser-Guttman criterion is applied to the scree plot of components (Figure 6.15 A). This method implies that components two and three represent significant variation within the data, while the other components are likely random variance. The components retained for analysis represent a large portion of the variance in the data, explaining 42.8% and 32.6% respectively. Observing the biplot produced (Figure 6.15 D), there does not appear to be any distinct patterns between depositional environments and ichnogenera during the Ediacaran.

Component two explains the largest amount of variance as a result of the ordination, accounting for 42.8%. Observing the biplot in Figure 6.15 D, negative values of component two contain strictly marginal marine environments; however, they are also assigned positive values in the ordination, nullifying any clear paleoecological significance. The variables influencing component two provide no assistance in the explanation of results, with *Bergaueria* and *Archaeonassa* acting largely against *Torrowangea*, *Palaeophycus*, *Helminthopsis*, and

Helminthoidichnites as displayed in Figure 6.15 B. This distinction of variables does not appear to be explained by architectural design, as simple horizontal trails (e.g., *Archaeonassa*, *Gordia*, *Helminthopsis*, and *Helminthoidichnites*) span nearly the entire spectrum of the axis. In addition to architectural design, ethology provides no support in explaining the distinction of variables. Buatois et al. (2017) observed that simple horizontal trails share the same ethological interpretation in that they are thought to represent simple grazing or locomotion.

Component three of the Ediacaran CA exhibits similar issues to that of component two. This component does not provide any clear distinction between depositional environments, nor does it show a logical separation of variables. The depositional environments display variations of marginal, shallow, and deep marine environments across the entire spectrum of the third component, with no clear groups present. The variables influencing the position of depositional environments in this component are equally as confusing. This is made clear in Figure 6.15 C, where simple horizontal trails are seen to be acting against each other (e.g., *Gordia* acting against *Archaeonassa*, *Helminthopsis*, and *Helminthoidichnites*).

The use of Q-mode cluster analysis (Figure 6.16) to assist with visual pattern recognition within the ordination biplot does not provide additional clarification to the CA results. Both non-hierarchical K-means clustering and hierarchical complete linkage agglomerative clustering produce the same result, while the UPGMA method creates different, but similar clusters. A clear explanation for the production of these clusters cannot be defined in terms of known scientific phenomena.

Figure 6.15 Ediacaran correspondence analysis results extracted from Coran. **(A)** Scree plot displaying the eigenvalues for each component (axis) created from the ordination, with a red line representing the Kaiser-Guttman criterion. **(B)** Component two variable loadings. **(C)** Component three variable loadings. **(D)** Biplot displaying the relationships between the observations (depositional environments) and variables (ichnogenera). See Table 6.1 for definitions of the depositional environments displayed in this plot.

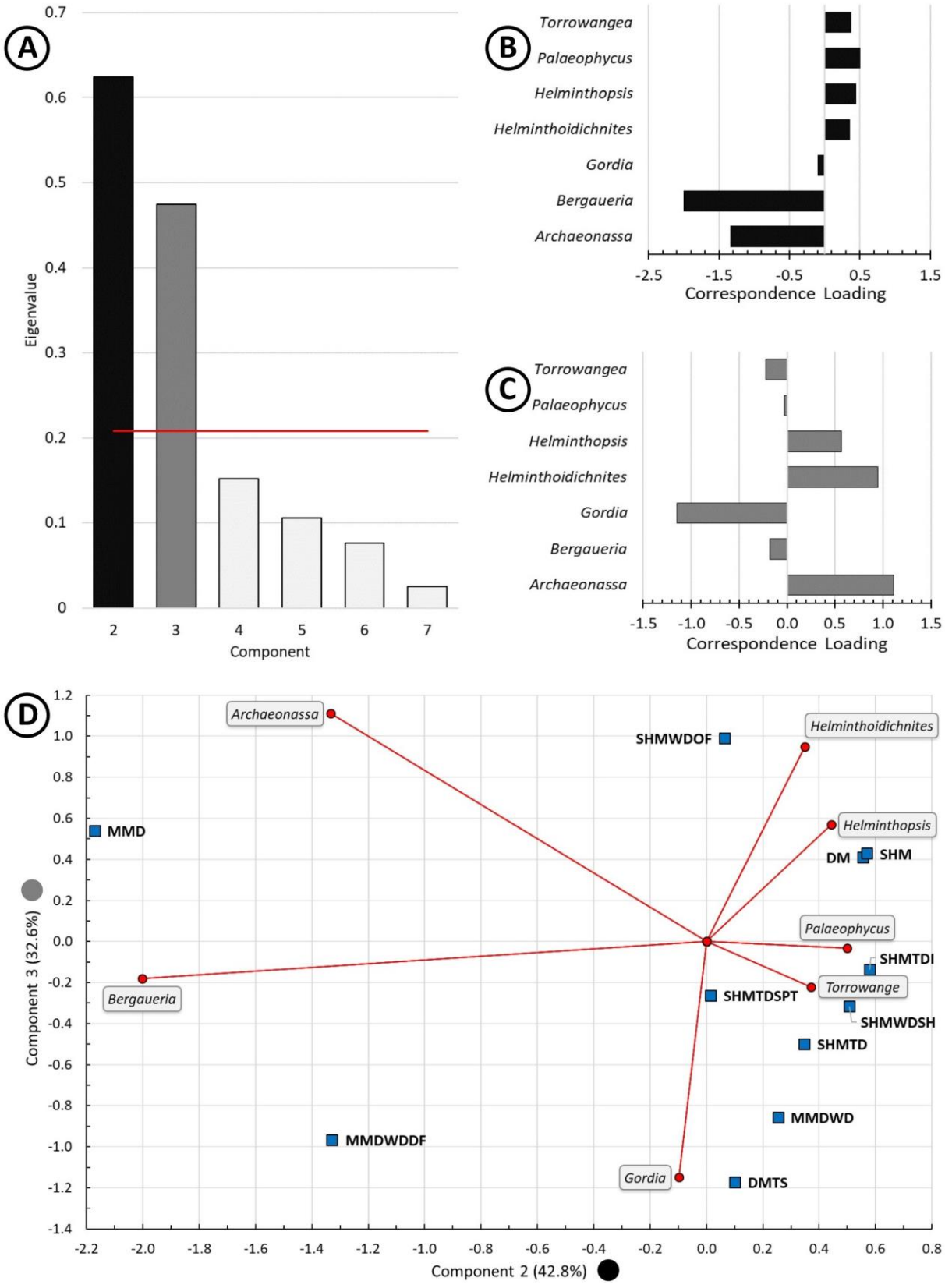


Figure 6.15 Caption on the previous page.

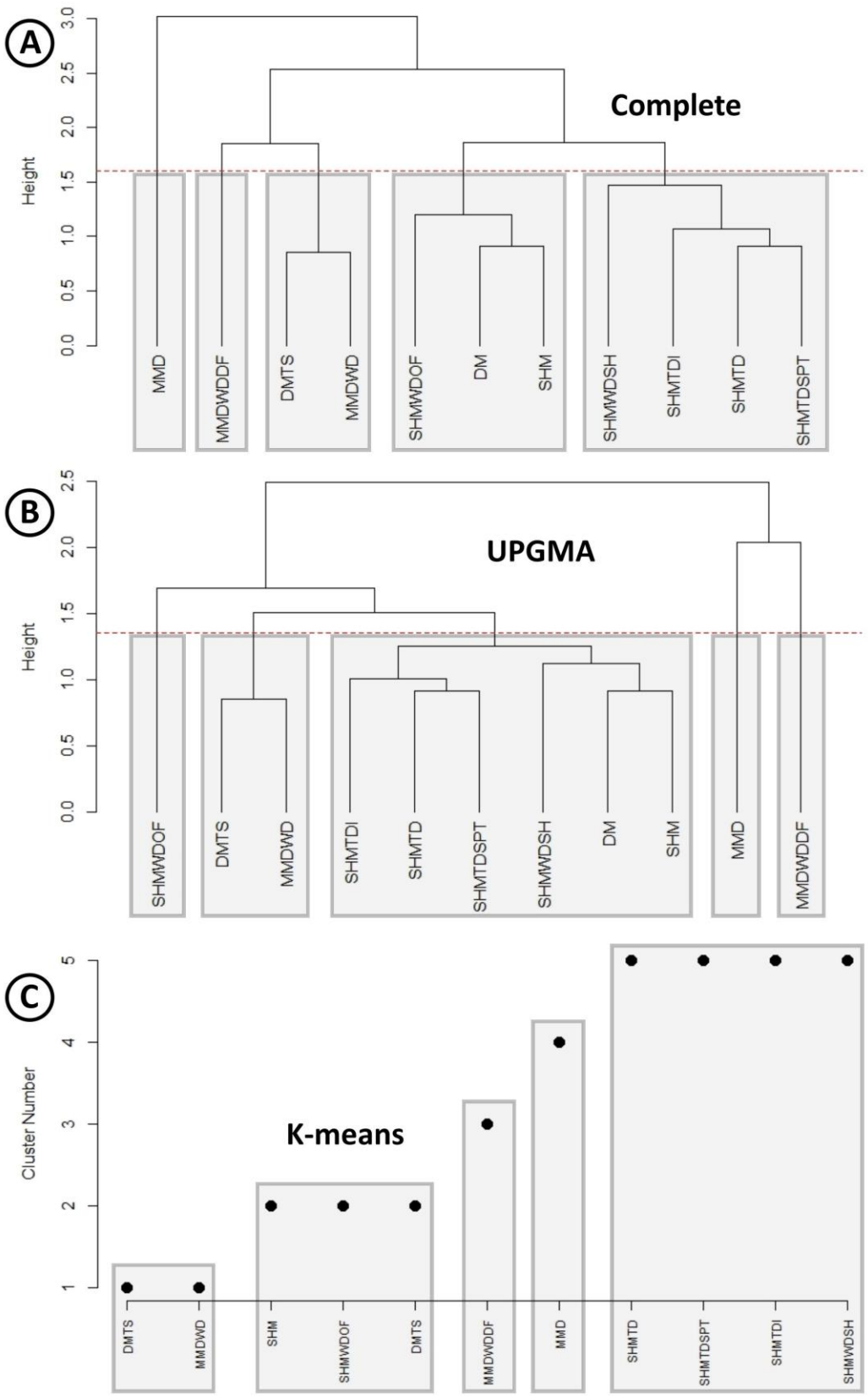


Figure 6.16 Caption on the following page.

Figure 6.16 Results of Q-mode cluster analysis conducted in R from the Ediacaran ordination. Refer to Table 6.1 for definitions of the depositional environments displayed in these plots. **(A)** Dendrogram of the hierarchical complete linkage agglomerative clustering. A dashed red line indicates the point at which each dendrogram produces five clusters, with the defined clusters highlighted in grey boxes. **(B)** Dendrogram of the hierarchical unweighted pair-group method using arithmetic averages clustering. Again, a dashed red line produces five clusters from these results, which are highlighted in grey boxes. **(C)** Plot displaying the non-hierarchical K-means clusters, with depositional environments grouped and highlighted according to the results produced in R.

There appears to be no paleobiological or paleoecological explanation for the results of the Ediacaran correspondence analysis. Without any observable patterns, it is likely that during the Ediacaran there was no relationship between depositional environments and trace fossil occurrence. Although this study is focused on identifying patterns in which scientific phenomena can be used to explain the observed relationships, in this situation the lack of patterns is of equal importance. The absence of relationships during the Ediacaran may be explained by the observation that trace fossils during this period were facies-crossing, leading to assemblages in shallow and deep marine environments that were nearly identical (Buatois & Mángano, 2016).

6.2.2.2 Terreneuvian Results and Interpretation

The data extracted for Terreneuvian correspondence analysis from IchnoDB has been compiled into a contingency table for analysis (Table 6.2). The sample size from this epoch is significantly larger than that of the Ediacaran, with 380% more trace fossil occurrences compiled. This is not unexpected, as Mángano & Buatois (2014) observed the largest increase in ichnodiversity and ichnodisparity during the Ediacaran to Cambrian transition (Figure 6.2). The resulting increase in trace fossil bauplan diversity, behavioral complexity, and improved infaunal ecospace utilization observed are the leading factors in the increased occurrence of ichnogenera.

Table 6.2 Contingency table created from Terreneuvian trace fossil data extracted from IchnoDB. Abbreviations are as follows: *Arenicolites* (*Ar*), *Didymaulichnus* (*Dd*), *Diplocraterion* (*Dp*), *Helminthopsis* (*Hl*), *Monomorphichnus* (*Mo*), *Oldhamia* (*Ol*), *Palaeophycus* (*Pa*), *Planolites* (*Pl*), *Rusophycus* (*Ru*), *Skolithos* (*Sk*), *Treptichnus* (*Tr*); Deep marine turbidite system (DMTS), Marginal marine (MM), Marginal marine deltaic (MMD), Shallow marine (SHM), Shallow marine tide dominated (SHMTD), Shallow marine tide dominated – intertidal (SHMTDI), Shallow marine tide dominated – intertidal tidal flat (SHMTDITF), Shallow marine tide dominate – subtidal (SHMTDSBT), Shallow marine wave dominated (SHMWD), Shallow marine wave dominated – shoreface (SHMWDSF), Shallow marine wave dominated – offshore transition (SHMWDOFT), Shallow marine wave dominated – offshore (SHMWDOF), Shallow marine wave dominated – offshore lower (SHMWDOFL), Shallow marine wave dominated – shelf (SHMWDSH).

	Ar	***Dd***	***Dp***	***Hl***	***Mo***	***Ol***	***Pa***	***Pl***	***Ru***	***Sk***	***Tr***	*row sum*
DMTS	0	2	0	4	0	11	2	0	0	0	0	*19*
MM	0	0	0	0	0	0	5	4	0	0	0	*9*
MMD	0	1	4	0	0	0	1	1	2	2	0	*11*
SHM	1	4	1	0	4	0	3	3	1	2	10	*29*
SHMTD	1	0	2	0	2	0	0	2	0	1	0	*8*
SHMTDI	1	1	0	0	0	0	1	1	1	0	2	*7*
SHMTDITF	2	0	1	0	0	0	2	0	0	3	5	*13*
SHMTDSBT	3	2	0	0	3	0	7	3	3	13	2	*36*
SHMWD	0	0	0	0	0	0	0	3	0	5	3	*11*
SHMWDSF	0	1	1	1	0	0	2	3	0	0	0	*8*
SHMWDOFT	3	5	3	1	4	0	4	9	3	13	4	*49*
SHMWDOF	3	7	2	10	17	9	8	20	6	5	10	*97*
SHMWDOFL	0	0	0	1	0	0	2	1	0	0	3	*7*
SHMWDSH	3	1	3	1	2	1	4	7	0	2	3	*27*
column sum	*17*	*24*	*17*	*18*	*32*	*21*	*41*	*57*	*16*	*46*	*42*	***331***

The determination of components that represent significant variance via the Kaiser-Guttman criterion was conducted on the Terreneuvian CA results (Figure 6.17). This criterion implies that the first four components represent significant variation in the ordination, resulting in their retention for further analysis. Analyzing the placement of both depositional environments and ichnogenera within the biplot produced from this ordination (Figure 6.18), there appears to be a relationship along component two. This component of this ordination represents the largest amount of variance, explaining 41.2%, and it is the only component in which explainable relationships are identified. This axis of the biplot displays the ichnogenera *Oldhamia* and *Helminthopsis* influencing the placement of deep marine turbidite systems (DMTS) and shallow marine wave dominated offshore (SHMWDOF) depositional environments, both of which can be explained by the "Agronomic Revolution" (Seilacher & Pflüger, 1994; Seilacher, 1999; Mángano & Buatois, 2014, 2017).

Figure 6.17 Terreneuvian correspondence analysis results extracted from Coran. **(A)** Scree plot displaying the eigenvalues for each component (axis) created from the ordination, with a red line presenting the Kaiser-Guttman criterion. **(B)** Component two variable loadings. **(C)** Component three variable loadings. **(D)** Component four variable loadings. **(E)** Component five variable loadings.

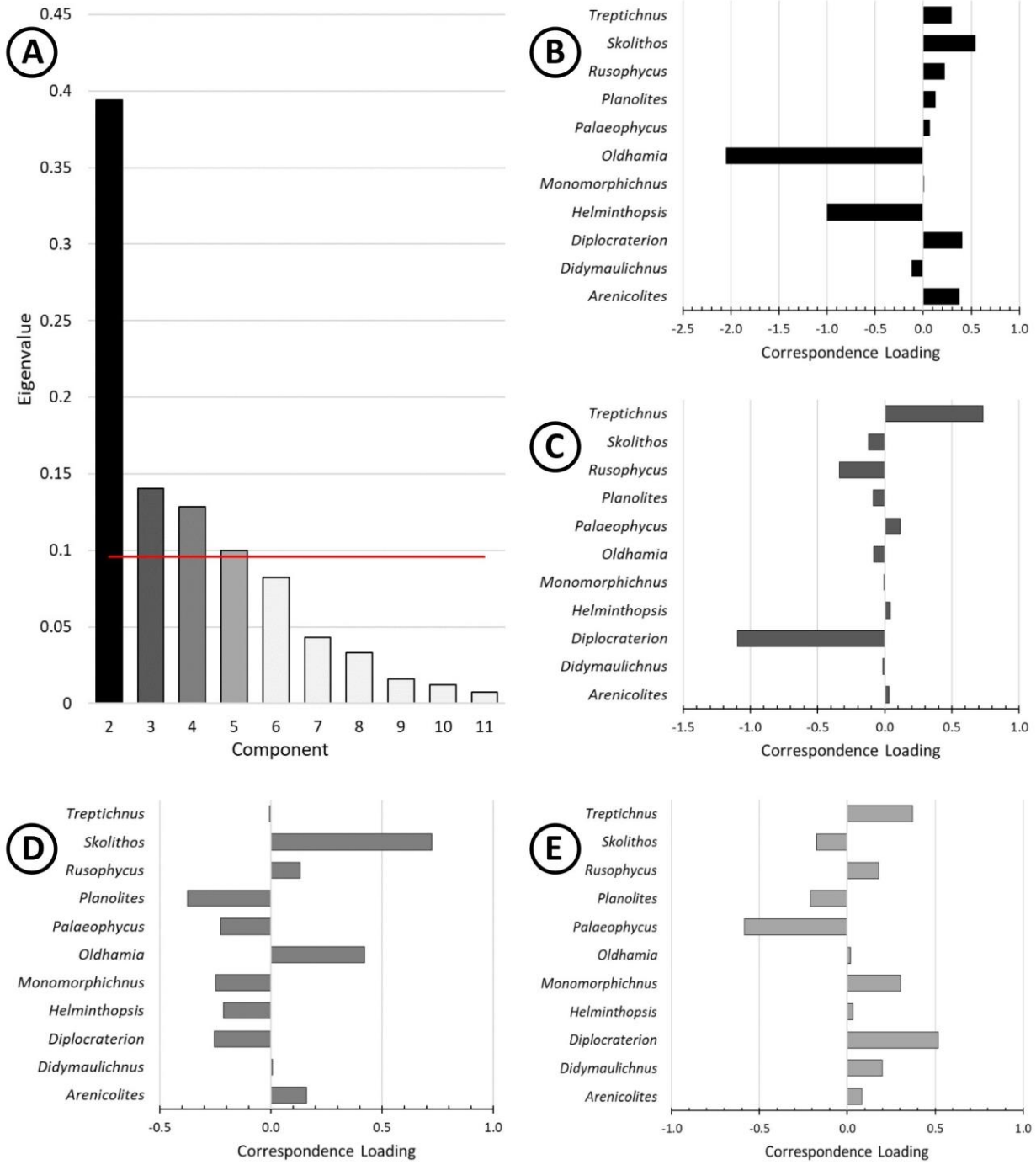


Figure 6.17 Caption on the previous page.

The Agronomic Revolution is described as the change in seafloor ecology from a dominantly matground environment during the Precambrian, to mixgrounds in the Phanerozoic (Seilacher & Pflüger, 1994; Seilacher, 1999; Mángano & Buatois, 2014, 2017). Seilacher (1999) described the Precambrian microbial mats supporting four distinct lifestyles, mat stickers, mat encrusters, mat scratchers, and undermat miners, with a later addition a fifth lifestyle, mat digesters by Buatois & Mángano (2016a). Of interest to this study are the undermat miners, represented by *Oldhamia*, and to a lesser extent, mat grazers, represented by *Helminthopsis* in the Terreneuvian CA. These matground related trace fossils run horizontal to sediment bedding planes, rather than crossing them (Seilacher, 1999). The shift from globally extensive matgrounds to mixgrounds was the result of penetrative bioturbation, which is observed in a very limited sense during the Fortunian (e.g., Buatois et al., 2014; Gougeon et al., 2018; Laing et al., 2018), becoming firmly established globally during Cambrian Age 2 (Mángano & Buatois, 2017). These authors cited multiple sources in which geochemical proxies, observations of limited primary fabric disturbance, and the lack of deep-tier dwelling structures throughout the Fortunian support this conclusion.

The ecological change described as the Agronomic Revolution and the subsequent effect on benthic organisms (i.e. Cambrian substrate revolution) appears to explain the distribution of depositional environments and ichnogenera observed in component two of the Terreneuvian biplot (Figure 6.18). The restriction of matground environments to deeper or more stressed settings (e.g., Hagadorn & Bottjer, 1999; Seilacher, 1999; Mángano & Buatois, 2017) is displayed in the biplot. Despite this epoch containing the Cambrian ages both before and after the onset of the Agronomic Revolution (i.e. Fortunian and Cambrian Age 2), deep marine turbidite systems (DMTS) settings are strongly influenced by the presence of *Oldhamia* and *Helminthopsis*, while shallow marine wave dominated offshore (SHMWDOF) settings are influenced by these matground trace fossils to a much lesser extent. Cluster analysis (Figure 6.19) supports the observation of a distinct separation of deep marine settings from those of shallow or marginal marine. All three methods create a distinct cluster containing only DMTS, with the two hierarchical methods showing this depositional environment to be the most dissimilar to any of the other settings.

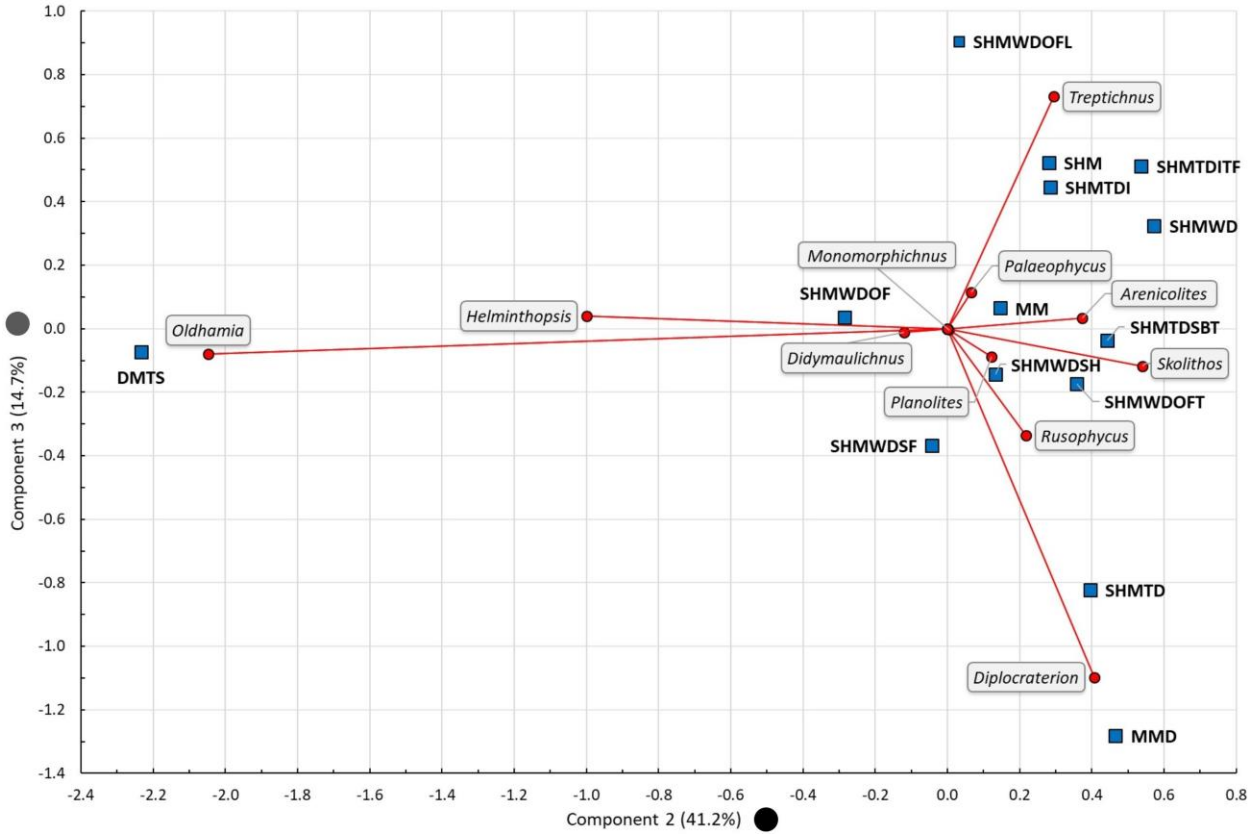


Figure 6.18 Terreneuvian biplot produced from the correspondence analysis results, displaying the relationships between the observations (depositional environments) and variables (ichnogenera). See Table 6.2 for definitions of the depositional environments displayed in this plot.

The analysis of component three reveals no patterns between depositional environments and ichnogenera, nor are there relevant clusters within the observations or variables. Cluster analysis appears to support this lack of significance, with individual clusters containing environments representative of markedly different paleoecological conditions. Further, the variable loadings across this component do not assist in providing an explanation for the placement of depositional environments as a result of ordination. Large negative values of component three are largely influenced by *Diplocraterion*, while large positive values are influenced by *Treptichnus*. No meaningful explanation can be provided in terms of ethology or ichnofacies to explain the influence of these ichnogenera on the placement of depositional environments in the biplot.

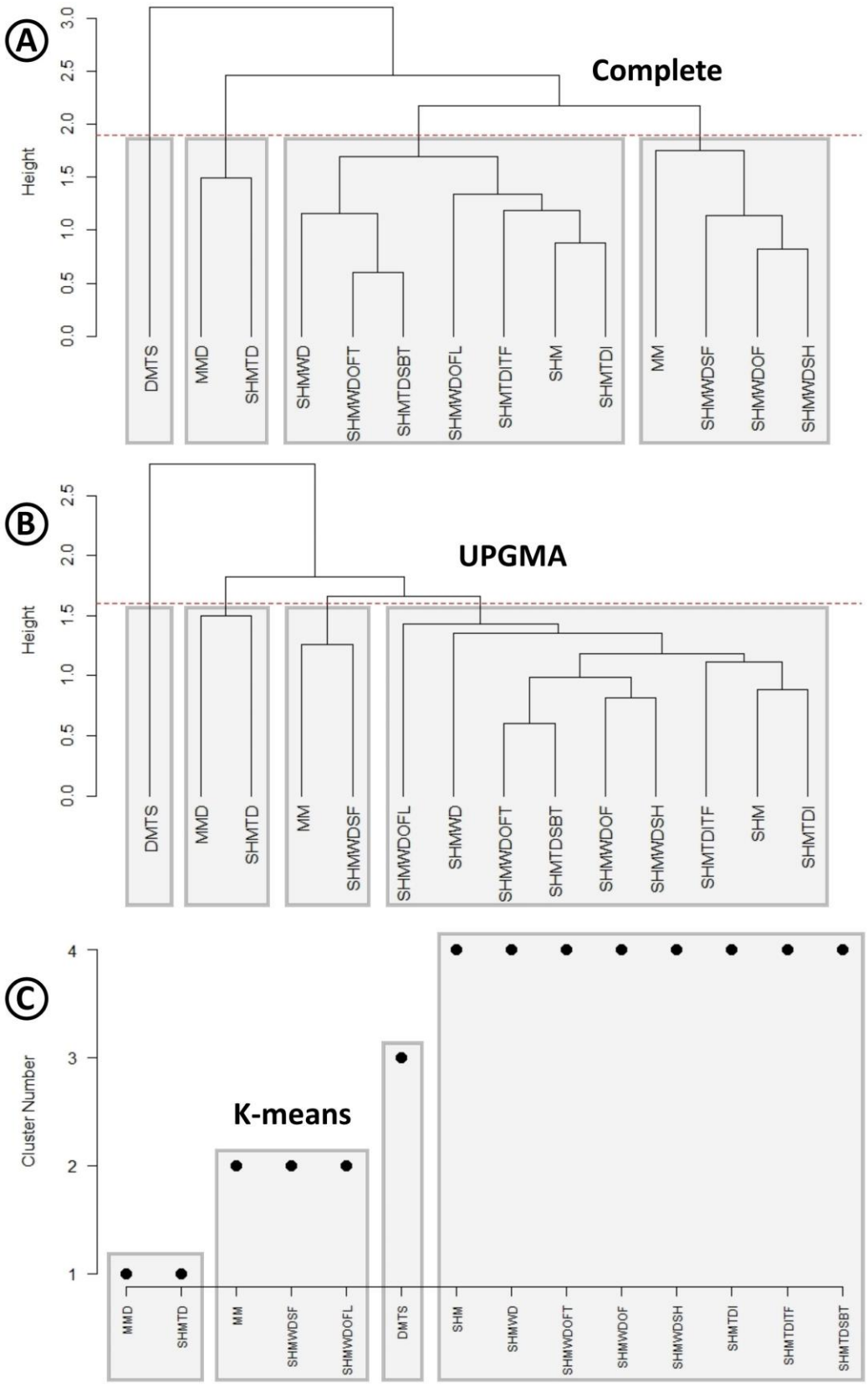


Figure 6.19 Caption on the following page.

Figure 6.19 Results of Q-mode cluster analysis conducted in R from the Terreneuvian ordination. Refer to Table 6.2 for definitions of the depositional environments displayed in these plots. **(A)** Dendrogram of the hierarchical complete linkage agglomerative clustering. A dashed red line indicates the point at which this dendrogram produces four clusters, with the defined clusters highlighted in grey boxes. **(B)** Dendrogram of the hierarchical unweighted pair-group method using arithmetic averages clustering. Again, a dashed red line indicates the point at which the dendrogram produces four clusters, with the defined clusters highlighted in grey boxes. **(C)** Plot displaying the non-hierarchical K-means clusters, with depositional environments grouped and highlighted according to the results produced in R.

The remaining components retained for analysis appear to offer no aid in the interpretation of the ordination, much like component three. Component four represents 14.5% of the variance in the data, while component five represents 10.4% of the variance. Variable loadings for each component are included in Figure 6.17; however, the variable groups cannot be explained ethologically or otherwise. Component four loadings show *Skolithos* and *Oldhamia* opposing *Planolites*, *Palaeophycus*, *Monomorphichnus*, *Helminthopsis*, and *Diplocraterion*. The separation of the ichnogenera in this component cannot be explained ethologically, as *Skolithos* and *Planolites* are both categorized as domichnia. Further, ichnofacies divisions do not provide an explanation, as both *Skolithos* and *Diplocraterion* are typical of the *Skolithos* ichnofacies, representing suspension feeders which require higher hydrodynamic energy and less turbid waters (Seilacher, 1963; Buatois & Mángano, 2011). Alternatively, the combination of *Skolithos* and *Oldhamia* influencing positive component values is confusing, as these ichnotaxa have little in common. They represent different ethologies (i.e. domichnia and fodinichnia respectively) and a preference for different ecological settings. Component five displays similar issues, with variable loadings displaying *Planolites, Palaeophycus,* and *Skolithos* acting against *Treptichnus, Monomorphichnus,* and *Diplocraterion*. These two ichnogenera groups present the similar ethological and ichnofacies issues present in component four.

6.2.2.3 Cambrian Epoch 2 Results and Interpretation

The data extracted for Cambrian Epoch 2 correspondence analysis from IchnoDB has been compiled into a contingency table for analysis (Table 6.3). The sample size is again significantly larger than the previous epochs analyzed, showing a 284% increase in trace fossil occurrences compared to the Terreneuvian dataset. The increase in sample size between the two

Cambrian datasets can be explained by an increased number of recorded localities, and a slightly higher ichnogenera density at individual localities during the Cambrian Epoch 2. As described in Section 6.1.3.3, the increased number of Cambrian Epoch 2 trace fossil occurrences is likely attributable to continued behavioral development and ecospace exploitation (Mángano & Buatois, 2014).

The results of the correspondence analysis conducted on the Cambrian Epoch 2 dataset with Coran are displayed in Figure 6.20. Two components have been retained for analysis, using the Kaiser-Guttman criterion as a guide for component retention. Component two of this ordination summarizes a distinctly larger portion of the variance within the original data, explaining 65.0%. The other component retained explains a smaller 13.2% of the variance. Observing the biplot created from the ordination (Figure 6.21), component two displays a dramatic distinction between deep marine turbidite systems (DMTS) and the remaining depositional environments. This is very similar to the relationship observed in the Terreneuvian CA, but much more obvious. As with the Terreneuvian results, the relationship between DMTS and *Oldhamia* can be explained by the Agronomic Revolution.

Table 6.3 Contingency table created from the Cambrian Epoch 2 trace fossil data extracted from IchnoDB. Abbreviations are as follows: *Bergaueria* (*Be*), *Cruziana* (*Cr*), *Dimorphichnus* (*Dm*), *Diplichnites* (*Di*), *Monomorphichnus* (*Mo*), *Oldhamia* (*Ol*), *Palaeophycus* (*Pa*), *Phycodes* (*Ph*), *Planolites* (*Pl*), *Rusophycus* (*Ru*), *Skolithos* (*Sk*), *Teichichnus* (*Te*); Deep marine – turbidite system (DMTS), Marginal marine (MM), Marginal marine delta – tide dominated (MMDTD), Shallow marine (SHM), Shallow marine platform (SHMPT), Shallow marine tide dominated (SHMTD), Shallow marine tide dominated – intertidal (SHMTDI), Shallow marine tide dominated – subtidal (SHMTDSBT), Shallow marine wave dominated (SHMWD), Shallow marine wave dominated – shoreface (SHMWDSF), Shallow marine wave dominated – lower shoreface (SHMWDSFL), Shallow marine wave dominated – offshore transition (SHMWDOFT), Shallow marine wave dominated – offshore (SHMWDOF), Shallow marine wave dominated – shelf (SHMWDSH).

	Be	***Cr***	***Dm***	***Di***	***Mo***	***Ol***	***Pa***	***Ph***	***Pl***	***Ru***	***Sk***	***Te***	*Row sum*
DMTS	1	0	0	0	1	63	0	0	0	0	0	0	*65*
MM	0	0	1	0	2	0	21	3	3	4	12	0	*46*
MMDTD	24	8	0	0	0	0	0	0	8	8	8	0	*56*
SHM	3	6	7	3	9	0	14	4	14	8	20	2	*90*
SHMPT	1	0	3	2	0	0	3	0	4	0	1	0	*14*
SHMTD	1	1	1	1	2	0	1	3	4	5	3	5	*27*
SHMTDI	0	0	7	7	0	0	0	0	0	7	7	0	*28*
SHMTDSBT	16	27	15	7	4	0	34	25	43	33	41	21	*266*
SHMWD	1	5	1	6	4	0	14	2	11	1	12	4	*61*
SHMWDSF	4	5	1	3	4	0	5	5	20	7	13	0	*67*
SHMWDSFL	0	0	0	0	0	0	0	0	3	0	1	1	*5*
SHMWDOFT	9	21	5	14	7	0	8	9	35	28	9	12	*157*
SHMWDOF	1	5	4	2	8	0	8	0	16	3	3	0	*50*
SHMWDSH	4	0	0	0	0	0	0	1	0	0	2	0	*7*
column sum	*65*	*78*	*45*	*45*	*41*	*63*	*108*	*52*	*161*	*104*	*132*	*45*	***939***

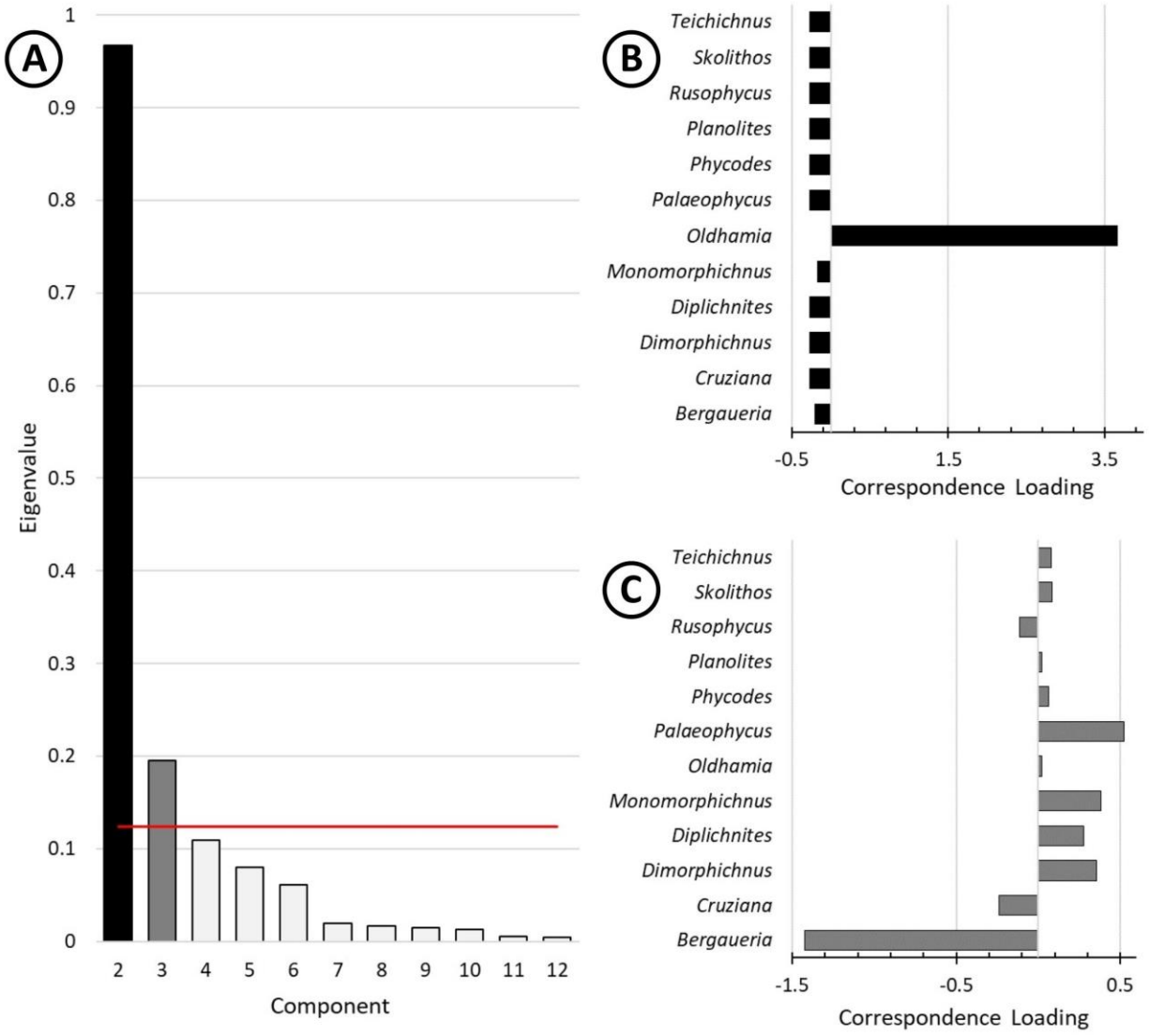


Figure 6.20 Cambrian Epoch 2 correspondence analysis results extracted from Coran. **(A)** Scree plot displaying the eigenvalues for each component (axis) created from the ordination, with a red line representing the Kaiser-Guttman criterion. **(B)** Component two variable loadings. **(C)** Component three variable loadings.

As previously described, the Agronomic Revolution details the change in seafloor ecology from a matground environment to a mixgrounds, taking place during Cambrian Age 2 (Seilacher & Pflüger, 1994; Seilacher, 1999; Mángano & Buatois, 2014, 2017). Similar to component two of the Terreneuvian, component two of the Cambrian Epoch 2 displays a dramatic influence from *Oldhamia* on the placement of deep marine turbidite system (DMTS) depositional environment. The presence of *Oldhamia*, an ichnogenera recognized as an undermat miner, indicates the presence and persistence of matground settings in DMTS environments beyond the initiation of the Agronomic Revolution (Orr, 2001; Buatois & Mángano, 2003; Seilacher et al., 2005; MacNaughton et al., 2016; Mángano & Buatois, 2017). The Terreneuvian data (Table 6.2) showed that *Oldhamia*, and therefore matgrounds, occurred in shallow marine wave dominated offshore and shelf depositional environments, in addition to the primary depositional environment of deep marine turbidite systems. This was interpreted as a shift to more stressed paleoecological settings as the agronomic revolution was beginning. The Cambrian Epoch 2 data (Table 6.3) shows that *Oldhamia* strictly occurs in DMTS depositional settings. This provides support for the agronomic revolution acting as a protracted event (Mángano & Buatois, 2017), with the expansion of the agronomic revolution during the Cambrian Epoch 2 into increasingly stressed environments and restricting matgrounds to the most extreme of these settings. Finally, all three clustering methods support the distinction of DMTS depositional environments as being entirely distinct from the remaining environments (Figure 6.22).

The final component retained for analysis, was component three, which represents 13.2% of the variation within the data. This component displays a rather distinct variable loading plot in Figure 6.20 C, to which no explanation can be provided. *Bergaueria* dominates the negative aspect of component three, while *Palaeophycus*, *Monomorphichnus*, and *Dimorphichnus* all influence the positive aspect. An ethological distinction cannot be made between these two groups, as both *Bergaueria* and *Palaeophycus* are classified as domichnia. Typical ichnofacies groupings offer no explanation for this distinction of ichnogenera either. *Bergaueria* are often associated with the *Skolithos* ichnofacies (e.g., Buatois & Mángano, 2011; Knaust, 2017); however, the *Skolithos* ichnogenera act in the opposite direction of *Bergaueria* for this component. Further, *Bergaueria* can also be associated with the *Cruziana* ichnofacies, but *Dimorphichnus* and *Monomorphichnus* are also typical of this ichnofacies. Despite the unclear

division of variables in this component, the dominance of *Bergaueria* is curious. It is observed that *Bergaueria* is likely produced by actinians (Alpert, 1973), a type of sea anemone, which may prefer distinct paleoecological conditions. It is unclear what specific conditions those might be, as the two depositional environments largely influenced by this variable (i.e. shallow marine wave dominated shelf and marginal marine tide dominated delta) have little in common.

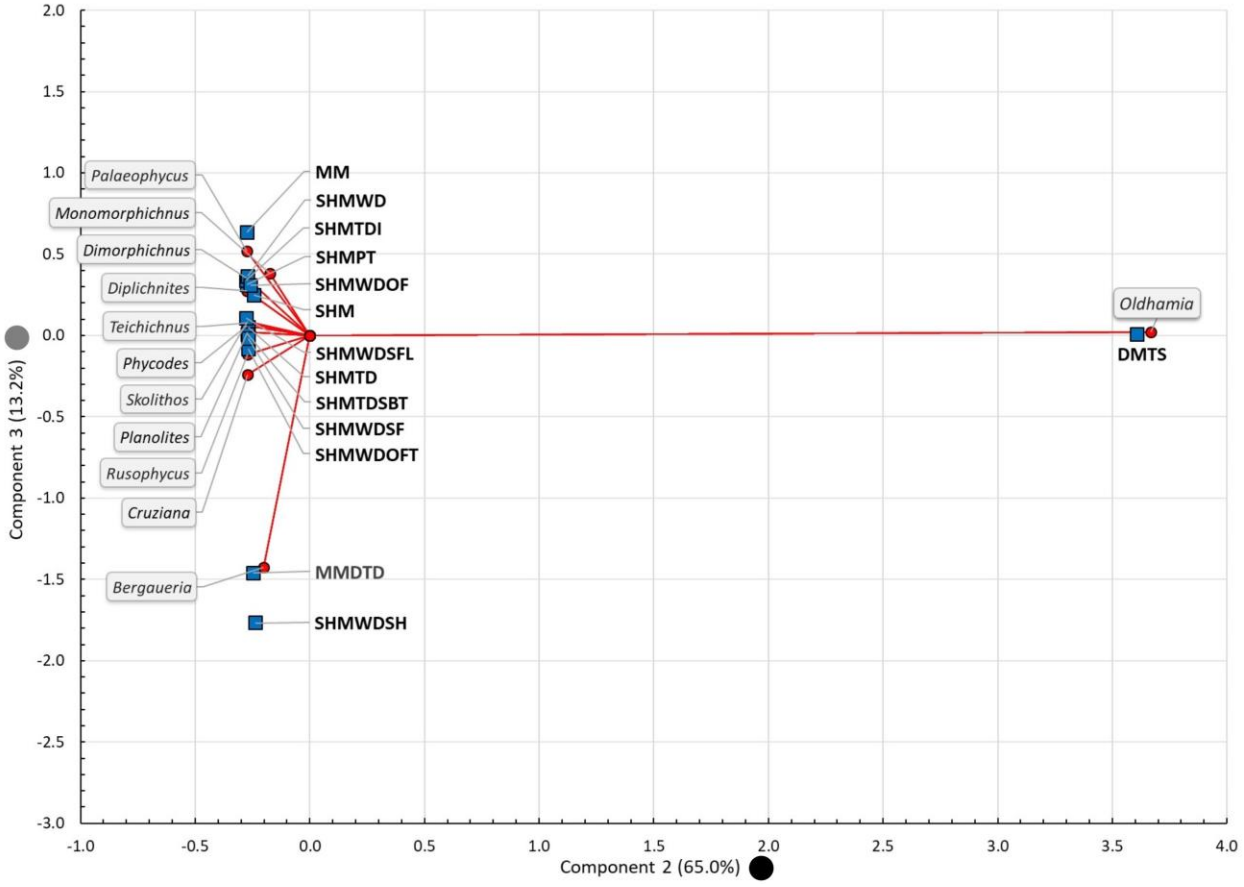


Figure 6.21 Cambrian Epoch 2 biplot produced from the correspondence analysis results, displaying the relationships between the observations (depositional environments) and variables (ichnogenera). See Table 6.3 for definitions of the depositional environments displayed in this plot.

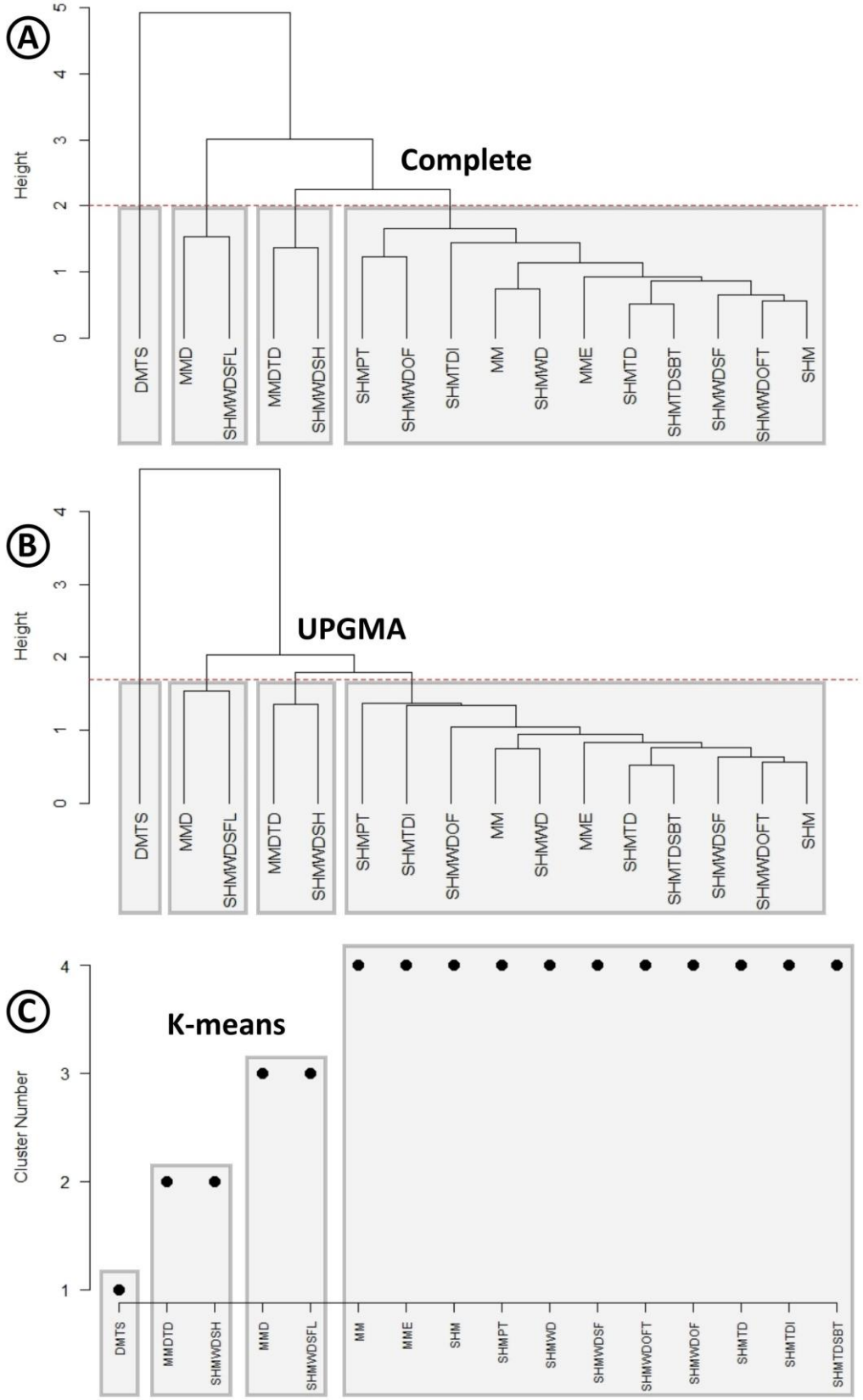


Figure 6.22 Caption on the following page.

Figure 6.22 Results of Q-mode cluster analysis conducted in R from the Cambrian Epoch 2 ordination. Refer to Table 6.3 for definitions of the depositional environments displayed in these plots. **(A)** Dendrogram of the hierarchical complete linkage agglomerative clustering. A dashed red line indicates the point at which this dendrogram produces four clusters, with the defined clusters highlighted in grey boxes. **(B)** Dendrogram of the hierarchical unweighted pair-group method using arithmetic averages clustering. Again, a dashed red line indicates the point at which this dendrogram produces four clusters, with the defined clusters highlighted in grey boxes. **(C)** Plot displaying the non-hierarchical K-means clusters, with depositional environments grouped and highlighted according to the results produced in R.

7. Conclusions

The purpose of this study was to use multiple analytical methods in the investigation of depositional environments and trace fossil occurrence. Through the creation of an ichnology database and the compilation of multiple other geological and paleontological datasets, both visual and statistical analytical methods were employed. Visualization techniques have produced underwhelming results towards the analysis of trace fossil and depositional environments, although they have proven useful in other aspects of paleontology related research. Alternatively, statistical analysis provided useful results in the investigation of possible relationships.

The requirement of detailed ichnology data was an essential aspect of this study. This prerequisite led to the identification of a near absence of ichnology data within online paleontology databases (e.g., PBDB and GBDB). With a clear gap in the data collection and storage capacity of current open access databases, this study proposes efficient and simple alterations to current database designs enabling detailed ichnological data storage. The inclusion of a field in which a record can indicate body or trace fossil storage would acknowledge philosophical differences between these branches of paleontology and support the storage of equivalent concepts between them. In the absence of these alterations and a lack of trace fossil data, the structure and logic behind the creation of an ichnology database, IchnoDB, is included here. With this database the utility of detailed ichnological data is displayed, with the trace fossil data extracted for visual and statistical analysis. In presenting this database structure and content it is intended to provide a basis for database improvement and support for additional research related to ichnology, paleontology, sequence stratigraphy, sedimentology, and paleogeography.

To support visual analysis of trace fossils and their depositional environments, several supplementary datasets were compiled. The most significant of these was the detrital zircon data, which was compiled to infer basic tectonic environments (Cawood et al., 2012). The inclusion of detrital zircon provides an indication of the geodynamic setting, which is observed to influence sedimentary basin creation and preservation, ultimately controlling what remains of the fossil record (Holland, 2016). A large detrital zircon dataset was compiled from previously published research across North America to test the utility of detrital zircon as a tectonic indicator throughout the Precambrian and Phanerozoic eons. As a result, it was concluded that fossil preservation across the northern and western margins of Laurentia throughout the

Ediacaran and Cambrian were dominated by extensional tectonic settings and few collisional settings. This not only suggests favorable sediment and subsequent paleontological preservation across these regions, but it also displays the geodynamic history during this time.

In addition to the utility of detrital zircon data as an indicator for tectonic settings, this data was used to conduct visual provenance analysis as a type of validation for the plate tectonic model to be used. Two examples were provided, where the PalaeoPlates Model (Eglington, 2018c) was shown to agree with previously published provenance theories. This included the large fluvial drainage system across Laurentia during the Neoproterozoic as described by Rainbird et al. (1997, 1992) and the provenance of zircons of a distinct North American magmatic gap signature likely sourced from Australia (Ross et al., 1992). Although visual provenance analysis was shown to be an effective tool in plate tectonic model validation, this method is not meant to displace traditional provenance analysis methods. Rather, this technique is instead considered a good starting point for provenance analysis and the identification of general trends within the detrital zircon data.

With the compilation of a significant detrital zircon dataset it was recognized that statistical data mining techniques can also be employed on this supplementary information. MDS has been an effective statistical technique in the investigation of provenance analysis and geodynamic evolution. With a thorough understanding of the peculiarities of MDS (e.g., the dissimilarity measure selected, the MDS algorithm employed, the number of dimensions in which to produce the ordination, and the sample size), this technique was tested on a reduced version of the compiled Laurentian detrital zircon dataset. Utilizing roughly one third of the original data compilation, this dataset represented a much wider geographic region and the number of samples contained was an order of magnitude larger than any previously published examples. Despite the identification of interesting grain age trends and general geodynamic groupings, the results were not applicable to the scope of this thesis. Two issues were identified for future consideration, with the primary issue being the complexity of the dataset not being properly represented by the selection of two dimensions. In addition to this, the number and disparity (e.g., dissimilarity in age, duration, and geographic extent) of basins from which the dataset was comprised precluded effective summarization of the geodynamic evolution. The use

of MDS is undoubtedly effective when analyzing detrital zircon; however, the number and geological history of sedimentary basins from which the dataset is composed must be considered.

Having compiled a significant amount of ichnology data and supplemental detrital zircon data, multiple open access datasets were gathered to further supplement visual analysis through the creation of paleogeographic reconstructions. This included data that was used to infer simple paleogeographic settings such as deep and shallow marine, terrestrial, and orogenic belts. With these datasets three paleogeographic reconstructions were composed, which were then used to conduct visual analysis of possible relationships between depositional environments and trace fossil occurrence. Although these reconstructions are the most accurate depiction of the paleogeography through the late Ediacaran to the Cambrian Epoch 2 given the data compiled, the provisional nature of these models must be recognized. With new or improved datasets, updated plate tectonic models, or improved criteria in the definition of the paleogeographic settings identified these reconstructions will require alterations.

Visual analysis of trace fossil occurrence and the depositional environments in which they were produced was conducted using the newly created paleogeographic reconstructions and the supplementary datasets. As a result, there was no observed relationship between trace fossil occurrence and their depositional environments. Although there were several important conclusions stemming from this analysis, including the global distribution of trace fossils from the late Ediacaran to the Cambrian Epoch 2 with no spatial restrictions in terms of paleolatitude or paleoclimate. It has been interpreted that this is a clear display of the utilization of available ecospace throughout the Ediacaran to the Cambrian Epoch 2. Visual analysis also concluded that there was a gradual adaption to marginal marine environments by trace fossils, which was interpreted as the transition from matgrounds dependent Ediacaran biota to the mixgrounds Phanerozoic biota through the Agronomic Revolution (Seilacher & Pflüger, 1994; Seilacher, 1999; Mángano & Buatois, 2014, 2017). This observation was also suggested to be related to the Fortunian diversification event, which recorded trace fossil bauplan diversification (Mángano & Buatois, 2014). Finally, through visual analysis it was clear that trace fossil occurrences increased significantly from the Ediacaran to the Cambrian Epoch 2. The Agronomic Revolution was again interpreted to be a major factor for this observation, as the Phanerozoic biota diversified and utilized infaunal ecospace.

The inclusion of detrital zircon data as a summary of geodynamic settings for visual analysis did not work as intended. Although the detrital zircon dataset compiled as part for this thesis was helpful, the data was restricted to North America. To address this limitation, an additional global detrital zircon dataset was obtained (Puetz et al., 2018) to provide geodynamic settings globally. Despite the level of detail contained within this dataset compared to other open access detrital zircon compilations, the deposition ages of the lithostratigraphic units from which the zircon grains were extracted was not included. This lack of detail not only affects the visualization of temporal data, but it is also required to provide reliable tectonic interpretations. Therefore, the inclusion of this dataset is appropriate in theory, but the missing information produces results that are not reliable. Future application of this data type will require accurate deposition ages of the samples analyzed.

Finally, the application of a statistical data mining method, correspondence analysis (CA), was applied to the data extracted from IchnoDB as a tool towards uncovering possible relationships between trace fossils and depositional environments. From this method three biplots were produced for the intervals of time investigated, exposing significant relationships (or lack thereof) in all three intervals. The Ediacaran dataset did not display any relationships to which an explanation could be provided; however, the lack of a relationship between trace fossils and depositional environments was significant. It is suggested that this provides evidence for Ediacaran trace fossils displaying facies crossing characteristics (Buatois & Mángano, 2016a), contrary to their representation of a narrow facies range during the Phanerozoic (Pemberton et al., 1992b). The Terreneuvian and Cambrian Epoch 2 biplots display a clear relationship between *Oldhamia* and deep marine depositional environments, identifying progressive habitat segregation. This correlation is interpreted to represent the paleoecological effect on undermat miners (i.e. *Oldhamia*) through the Agronomic Revolution. Further support for this interpretation is provided by the minor influence of *Oldhamia* on additional depositional environments (i.e. offshore and shelf) during the Terreneuvian, followed by this ichnogenus occurring in strictly deep marine depositional environments during the Cambrian Epoch 2. It is suggested that this displays the protracted influence of the Agronomic Revolution through the early Cambrian, with the distribution of matgrounds restricted to increasingly harsh and stressed depositional environments. The use of CA to uncover progressive ichnogenus differentiation as

a result of the Agronomic Revolution provides a quantified description of increasing beta ichnodiversity during the early Cambrian.

This study demonstrates the mixed results of visual and statistical data mining techniques in the investigation of any significant relationships between trace fossil occurrence and their associated depositional environments. Although the visual analysis techniques employed in this thesis displayed several important trends, it is difficult to represent two variables (e.g., ichnogenus and related depositional environment) with one spatial locality. Therefore, visual analysis of ichnological data is best employed for spatial analysis of single variables, or multiple variables with a very limited number of possibilities. Future visual analysis might consider the use of smaller intervals of time, or a continuation of the paleogeographic analysis conducted by Jensen et al. (2013) with a very similar dataset. Statistical data mining techniques have been shown to provide quantified results that can be interpreted using area specific knowledge and scientific phenomena. Additional statistical categorization could be applied to any combination of data fields associated with the trace fossil records of IchnoDB. The many possibilities for exploratory data analysis exemplify the statement put forward by Morgenthaler (2009), where the creativity of the analyst is the only limitation.

8. Appendix

A IchnoDB Supplementary Material

Supplementary material for Chapter 2 is included as a compressed file, and contains the .mdb file of IchnoDB, as well as a .docx file providing additional details about the main tables within IchnoDB. The .docx file also contains a guide for extracting data from publications to support efficient data entry in the database.

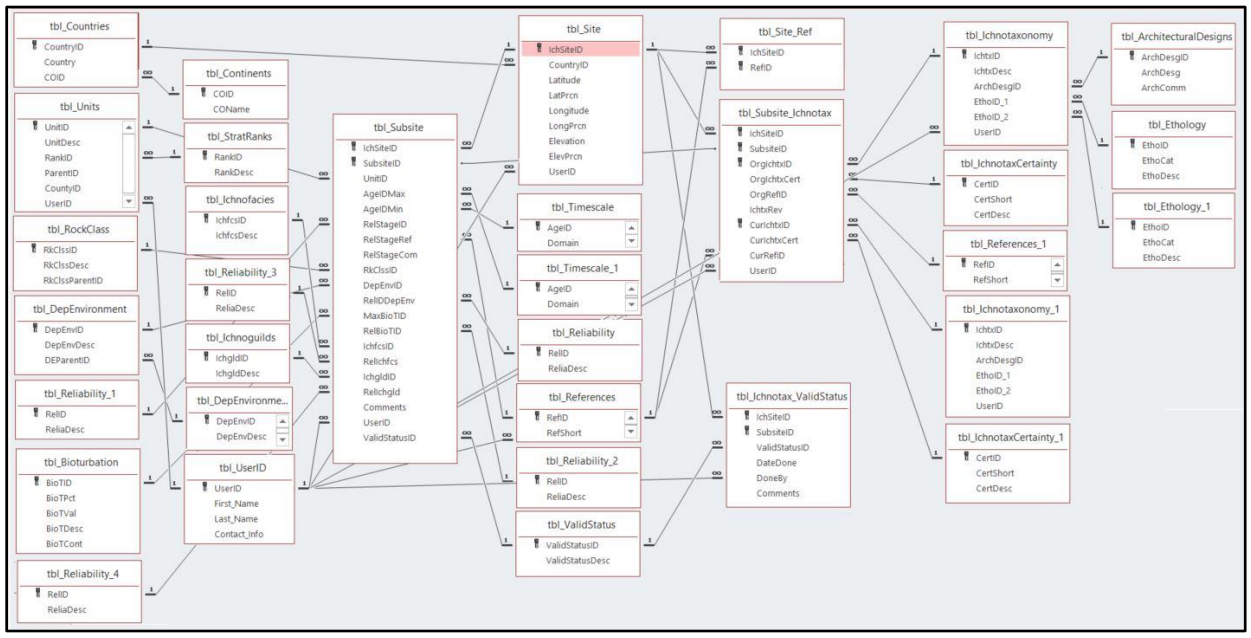


Figure A.1 Relationship diagram between all tables of IchnoDB defined in Microsoft Access. Primary keys within each table are represented by a small key symbol, and the nature of each relationship is defined as either one (1) or many (∞) with the exception of the SubSiteID between tbl_Subsite and tbl_Subsite_Ichnotax which is a many to many relationship.

B Detrital Zircon Supplementary Material

Supplementary material for Chapter 4 and 5 is included as a compressed file due to the large size of these datasets. The additional information included for Chapter 4 includes the U-Pb and Lu-Hf datasets as .xlsx files, which were compiled from detrital zircon publications across North America.

The additional information included for Chapter 5 contains all of the .csv files, and the .r file used to conduct MDS ordination. This supplementary file also includes a .docx file describing the data pre-treatment required to conduct MDS with the parametric Bray-Curtis statistic.

The following section has been designated to the appendix as this work has become less relevant to the direction of study with the progression of thesis research. The results and ideas captured here are considered significant, and it is hoped that the work flow and ideas will instigate future work.

B.1 Review of The Lu-Hf Isotope System in Zircon

Zircon grains contain a large number of radioactive isotopes and their stable decay products, including uranium, thorium, rare earth elements (REEs), and hafnium (Kinny & Maas, 2003). This characteristic of zircon makes it a great candidate for geochemistry analysis. The Lu-Hf isotope system is described as particularly useful, as zircon preserves the initial $^{176}Hf/^{177}Hf$ ratio at the time of crystallization (Figure B.1). Kinny & Maas (2003) observe that this ratio can be used to calculate Hf model ages, it can be used to create a Lu-Hf isochron, or it can be used to determine an epsilon value (ε_{Hf}) that can serve as a geochemical tracer. The utility of the Lu-Hf isotopic system to study the evolution of crust and mantle through time was first observed by Patchett & Tatsumoto (1981). Epsilon values calculated for a given age can be used to differentiate juvenile, mantle derived crust (e.g., positive ε_{Hf} values) from crust that has been extracted through the re-melting of older crust (e.g., negative ε_{Hf} values) (Kinny & Maas, 2003). This same idea holds true for model ages derived from Hf isotopes, where model ages close to the age of crystallization represent juvenile, mantle derived crust and model ages that are

much older than the crystallization age represents old reworked crust. Kinny & Maas (2003) conclude that analysis containing both U-Pb and Lu-Hf isotopic data are the most useful, as the U-Pb isotopes serve to accurately determine age, a requirement for sound petrogenic information to be obtained from the Lu-Hf isotopic system. In terms of provenance analysis, identifying ε_{Hf} values or model ages from the detrital zircon grains can add additional criteria when looking for source regions, resulting in more robust suggestions.

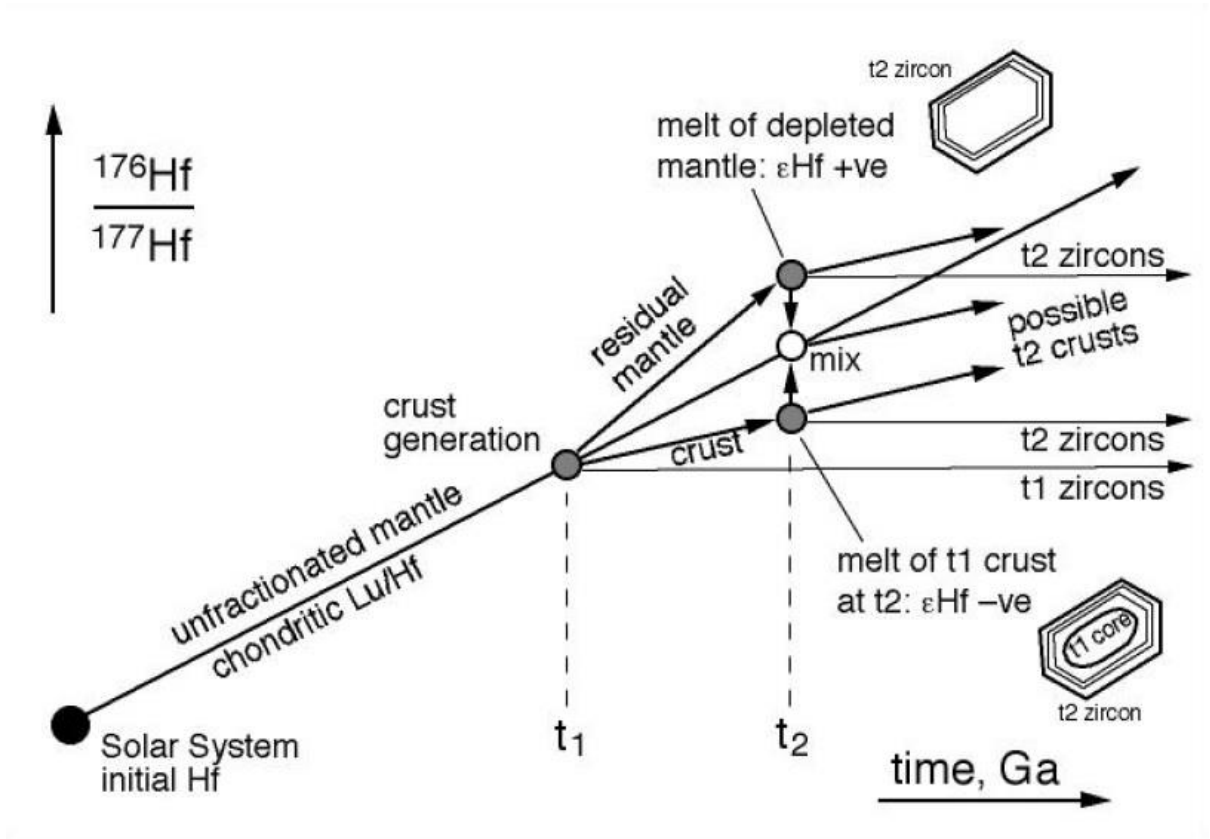


Figure B.1 Schematic Hf isotope evolution diagram (after Kinny & Maas, 2003). This shows an episode of crustal generation at time t_1, and the resulting divergent Hf isotope evolution in the generated crust and the remaining mantle. At time t_2 a range of possible Hf ratios may form from any combination of crustal and/or mantle extraction.

The Sm-Nd isotopic system is also used for geochronology and isotopic tracing purposes, with zircon again being a suitable mineral as it favors the concentration of heavy REEs, resulting in high Sm/Nd ratios that are favorable for high precision dating (Kinny & Maas, 2003). These authors observed that this isotopic system is not comparable to the accuracy or precision

provided by the U-Pb system. Rather, the Sm-Nd system is more useful as an isotopic tracer, with the recognition that the Lu-Hf and Sm-Nd systems behave in a comparable fashion during most magmatic processes, a theory that is supported by the correlation between ε_{Hf} and ε_{Nd} in crustal rocks (Vervoort et al., 1999). This correlation can be used to convert Nd isotopic data into equivalent Hf isotopic data when looking for provenance sources.

B.2 Using Detrital Hf Data

Of the detrital zircon data compiled, a small portion (~10%) included Lu-Hf isotope information. The utility of the Lu-Hf isotopic system was identified in the early 1980's as a tool to analyze crustal evolution (Patchett & Tatsumoto, 1981). Due to the small quantity of Hf isotopes found within zircon grains, this system was not routinely analyzed in detrital zircon studies until improvements in mass spectrometer techniques were made (e.g., Amelin et al. 1999). As described by Amelin et al. (1999), the Lu-Hf isotope system can be used to infer juvenile crust or old reworked crustal sources in the form of ε_{Hf} values or model ages. The calculation of either value from detrital zircon grains can be used in conjunction with magmatic Lu-Hf data to further restrict their provenance sources.

Unfortunately, a global dataset of igneous Lu-Hf analyses does not currently exist to compare with detrital zircon Lu-Hf analyses. A majority of the igneous isotope data contained within the DateView geochronology database (Eglington, 2004) is related to the Sm-Nd isotope system, with very limited amounts of igneous Lu-Hf data. Utilizing the linear relationship between ε_{Hf} and ε_{Nd} described by Vervoort et al. (1999), available ε_{Nd} values can be transformed into ε_{Hf} values. This takes advantage of all available data in the comparison of igneous values against those found within the detrital zircon grains.

Using initial isotopic ratios and their related measures (e.g., ε values) is an effective tool when comparing rocks of roughly similar ages; however, with rocks of disparate ages the time factor used in equations calculating these ratios makes them unreliable for comparative purposes (Champion & Huston, 2016). This issue arises from the constant evolution of the depleted mantle. These authors observe that one of the easiest ways to compare isotopic data from rocks or minerals of a widely different age is through the calculation of model ages. A two-stage model (T_{2DM}) was chosen over a single-stage model (T_{DM}), as the single-stage models become

increasingly unreliable as parent-daughter ratios increase (Champion & Huston, 2016). From the calculated two-stage depleted mantle model ages, a crustal residence age (T_{CRes}) can be calculated by simply subtracting the crystallization age of the rock or mineral from the two-stage model age of that same sample (e.g., Champion & Huston, 2016; Lancaster et al., 2011). Admittedly, calculating crustal residence age is not done without the major assumption that the magma has remained a closed system since extraction from the mantle (Champion & Huston, 2016; Roberts & Spencer, 2015). With the understanding that these calculations are strictly models and not isochrons, this technique can help distinguish juvenile crustal sources (T_{CRes} < 200 Ma) from old crustal reworking (T_{CRes} > 1500 Ma), which can then be used for comparative purposes.

FitPDF (Eglington, 2018b) has been used to plot the resulting crustal residence times calculated from the detrital zircon data (Figure B.2). Using this software, colour symbology is applied to assist in the identification of juvenile crustal and old crustal reworking signatures within samples. An additional feature has been included in FitPDF, which limits the T_{CRes} displayed to the high probability peaks (<80%) within the PDPs of each sample. This feature was included to address issues with age uncertainty and initial Hf values as defined by Fisher et al. (2014). These authors demonstrate that ancient Pb loss can result in large age uncertainties, while the zircon grain retains its original Lu-Hf isotopic composition from crystallization. This leads to calculating initial Hf values using an incorrect age, and therefore incorrect initial Hf values and all values derived based on the initial Hf. By restricting the T_{CRes} to high probability peaks within the PDPs, only ages that would be considered reliable are used and the problem addressed by Fisher et al. (2014) is avoided.

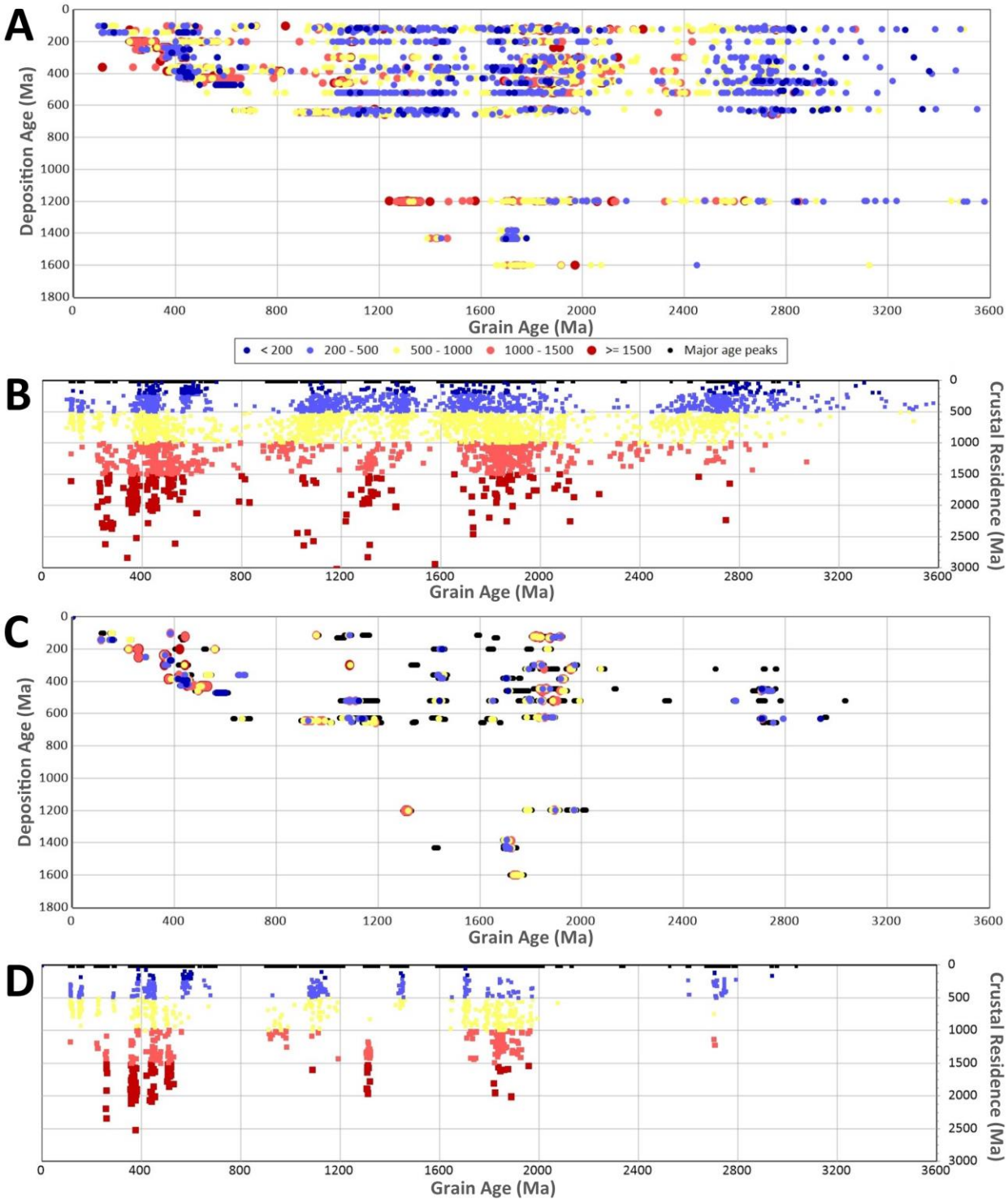


Figure B.2 Caption on the following page.

Figure B.2 Multiple probability density plots showing crustal residence times in association with significant grain age peaks defined by U-Pb analysis on the same zircon grains. The colour symbology in these plots is displayed at the bottom of the first plot, where dark blue points represent juvenile crustal sources (<200 Ma model ages), light blue (200 Ma to 500 Ma model ages) yellow (500 Ma to 1000 Ma model ages) and light red (1000 Ma to 1500 Ma model ages) points represent intermediate values, and finally dark red points represent old crustal reworking (>1500 Ma model ages). **(A)** Viewing all T_{CRes} (Ma) ages from the detrital zircon Hf analysis with approximate deposition age (Ma) of the sample displayed on the left axis, and the grain age (Ma) of the zircon on the bottom axis. **(B)** Viewing all T_{CRes} ages from the detrital zircon Hf analysis from "above," with the T_{CRes} age (Ma) displayed on the right axis and grain age (Ma) displayed on the bottom axis. **(C)** View of only T_{CRes} ages (Ma) from the detrital zircon Hf analysis that occur on high probability grain age peaks to avoid analysis of zircon grains with large age uncertainty caused by possible Pb loss as described by Fisher et al. (2014). Deposition age (Ma) is displayed on the left axis, while the grain age (Ma) is displayed on the bottom axis. **(D)** View of T_{CRes} ages (Ma) restricted to high probability grain age peaks from "above," with the T_{CRes} age (Ma) displayed on the right axis and grain age (Ma) displayed on the bottom axis.

For comparison and possible restriction of provenance sources for the detrital zircon calculated Hf T_{CRes} model ages the igneous isotopic dataset from DateView was used. This dataset contains 1,466 records of igneous Lu-Hf isotope analysis, with the remaining 6,319 records containing igneous Sm-Nd isotopic data. With the observation that ε_{Nd} and ε_{Hf} are correlated (Vervoort et al., 1999), a secondary objective of this study was to see if a similar correlation existed between Nd T_{2DM} and Hf T_{2DM}. To investigate this possibility, a dataset of compiled igneous whole rock Lu-Hf and Nd-Sm isotope analyses (Mao, 2018) was used. This dataset was then imported into Geodate (Eglington & Harmer, 1999), software that facilitates isotope modelling and regression. A weighted unconstrained regression was conducted on the calculated T_{2DM} model ages for both isotope systems, with the results displayed in Figure B.3. Using all 48 data points, the regression results in an errorchron with a mean sum of the weighted deviates (MSWD) equal to 9.421. Comparing this value relative to the critical F value of 1.37, defined by the number of samples used, the uncertainties do not explain the scatter about the best fit line, and therefore the regression should not be considered trustworthy. Despite the unreliable fit of the regression between these isotopic model ages, Sm-Nd isotopic data from DateView was converted to Lu-Hf values to determine if this method holds any value for provenance source restriction should a correlation be demonstrated with a different or larger dataset.

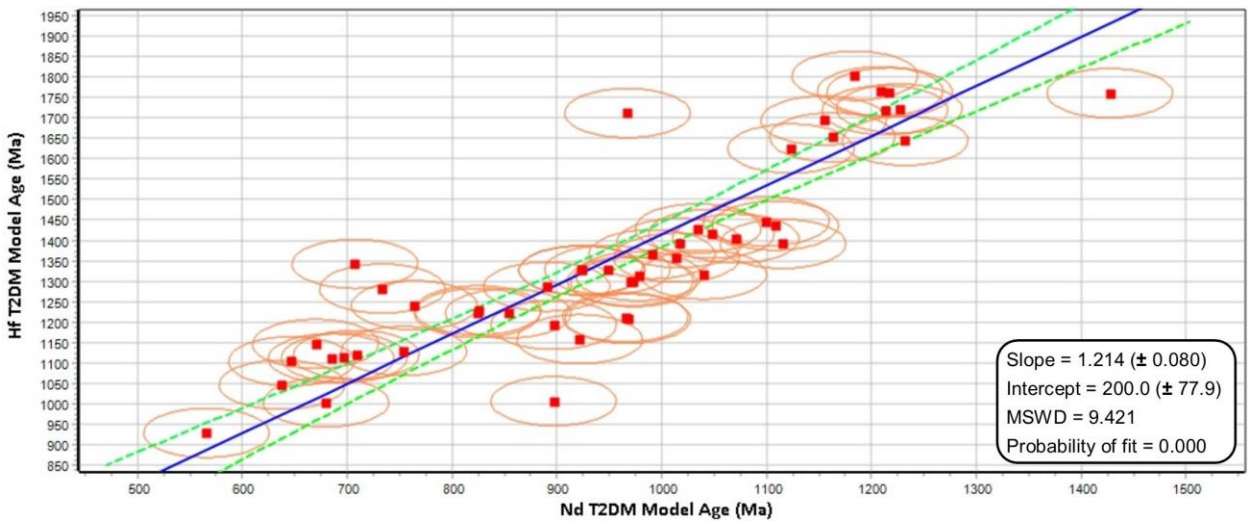


Figure B.3 Weighted unconstrained regression results from Geodate between Hafnium two-stage depleted mantle model ages and Neodymium two-stage depleted mantle model ages. All data points are plotted in red, with orange error ellipses. The best fit line of the regression is plotted in blue, with a dashed green error envelope.

With the linear relationship defined by the regression in Geodate, all DateView Sm-Nd values were converted to Hf T_{2DM} model ages, followed by the calculation of Hf T_{CRes} model ages. All data was integrated for visual analysis in ArcGIS to identify possible provenance sources (Figure B.4). As displayed in Figure B.4, this process is not effective as there is a distinct lack of juvenile or old crustal reworking signatures in which to restrict provenance sources. Rather, most calculated Hf T_{Cres} model ages are values that lie between young crustal sources and old crustal reworking which indicates that crustal sources are mixed in various proportions. It is the inability to determine what proportion of each end member is responsible for the resulting intermediate values that makes these Hf T_{CRes} model ages ineffective in restricting provenance sources. Of the limited juvenile crustal signatures, none are located on North America, and there are no old crustal reworking signatures. Although these model ages were produced from an errorchron, the efficacy of this method is questionable. Perhaps a larger igneous Nd and Hf isotope database would produce more encouraging results from which provenance sources of detrital zircon grains could be restricted.

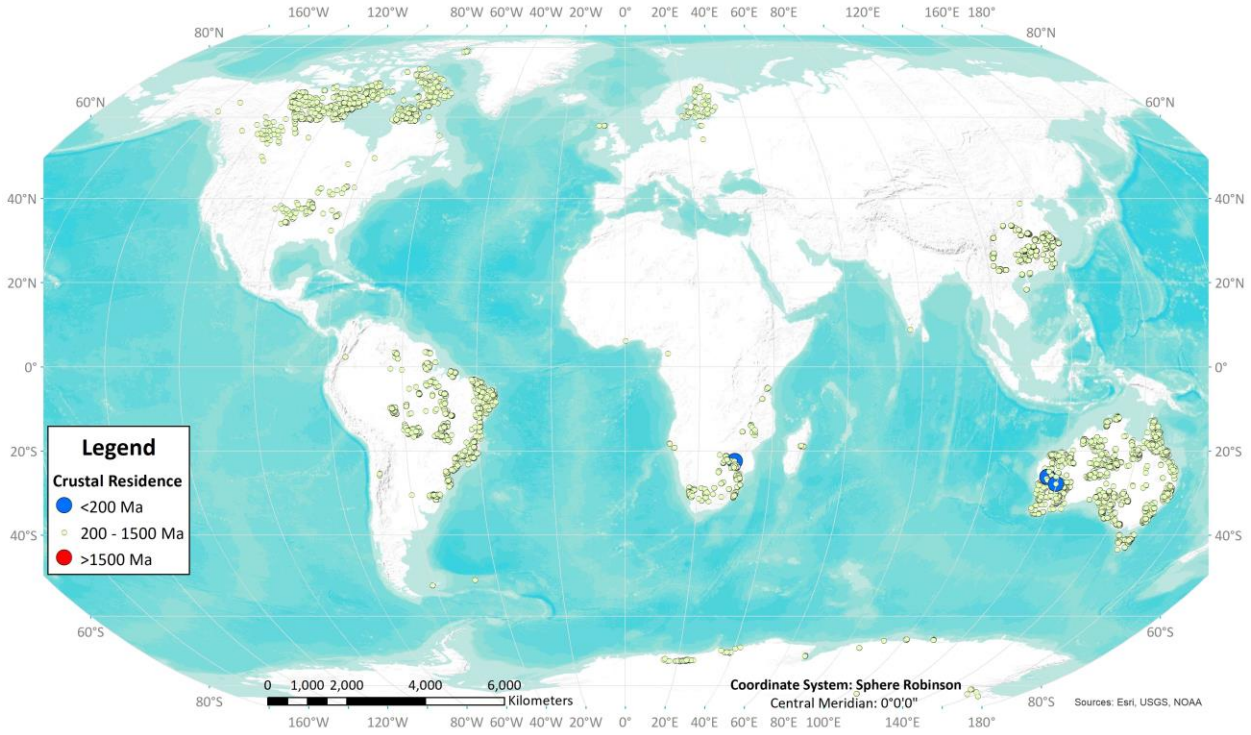


Figure B.4 Nd data extracted from DateView, transformed to display Hf crustal residence time (Ma). Symbology emphasizes juvenile crustal residence ages (<200 Ma), and old crustal residence ages (>1500 Ma), although there are no occurrences of the crustal residence ages over 1500 Ma.

C Supplementary Material to Visual and Statistical Analysis of Depositional Environments and Trace Fossil Occurrence

Supplementary material for Chapter 6 includes figures displaying the data used in the creation of paleoenvironmental interpretations and a compressed electronic file. The additional information included for Chapter 6 has been split into files used for visual analysis and files used for statistical analysis. Supplementary material for visual analysis is the paleontology data that was downloaded from the Paleobiology Database (PBDB) on 14 October 2018, using the following parameter: time intervals = 650 Ma to 486 Ma. The supplementary files for statistical analysis include the .xlsx files with the raw trace fossil data extracted from IchnoDB. In addition to the raw data, these spreadsheets include the contingency tables and the results from Coran (Eglington, 2018a) that were used to create the biplots and variable loading plots. Finally, the .csv files and code used in `R` for conducting cluster analysis is included.

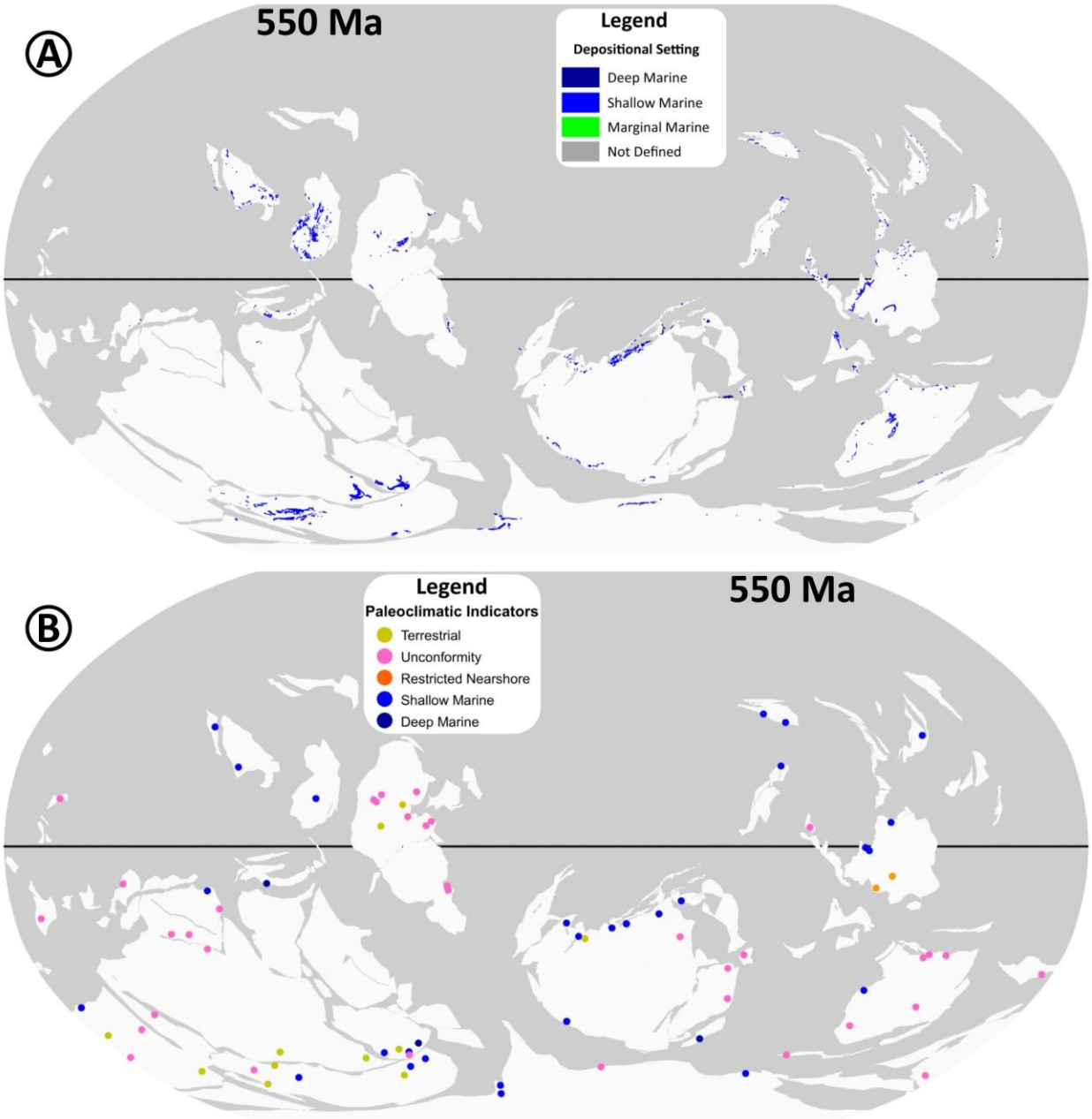


Figure C.1 Supplementary datasets utilized in the creation of paleogeographic reconstructions for the Ediacaran. **(A)** Geological map data from multiple sources. **(B)** Paleoclimatic indicators from Li et al. (2013).

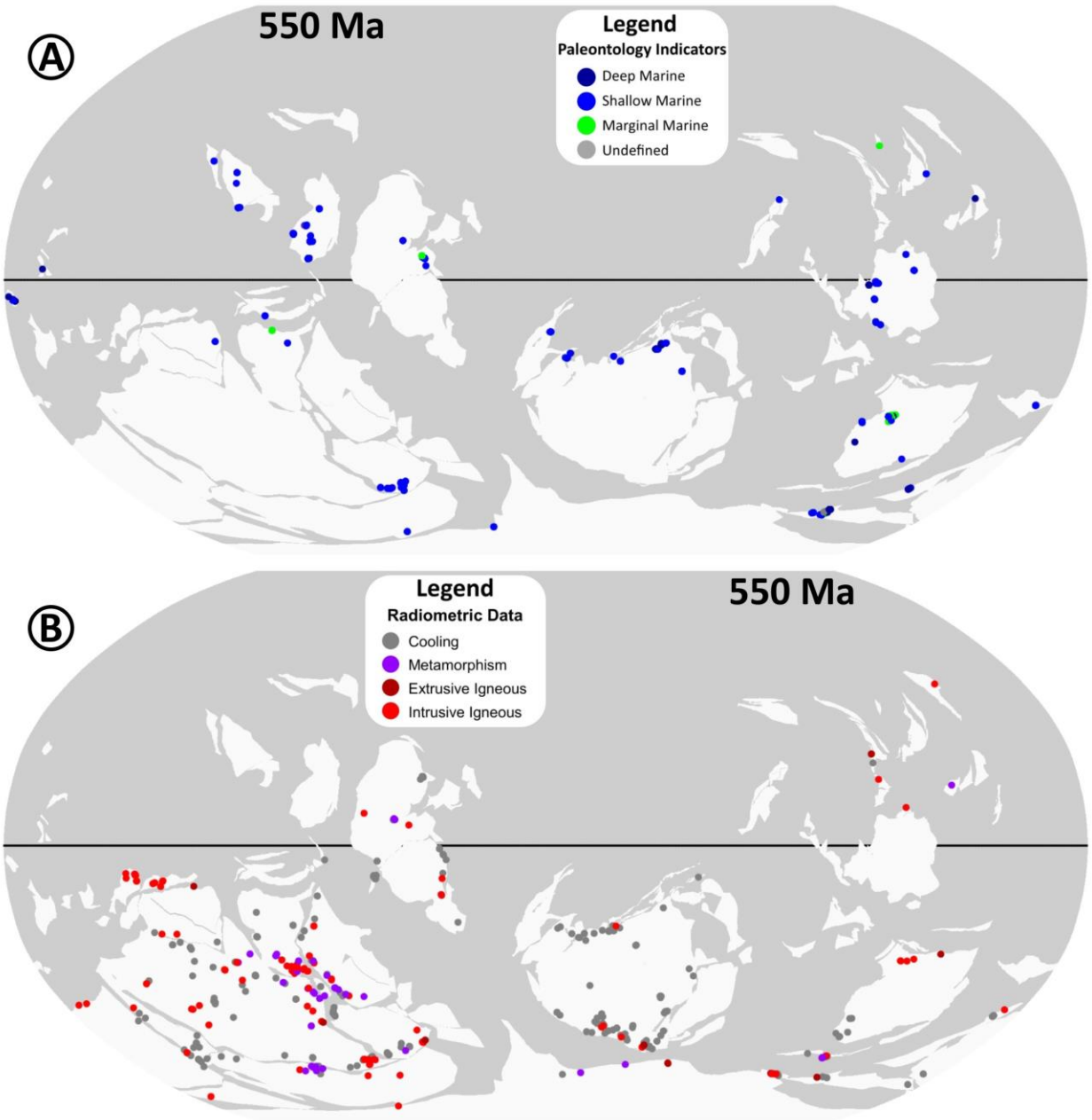


Figure C.2 Supplementary datasets utilized in the creation of paleogeographic reconstructions for the Ediacaran. **(A)** Paleontology data points displaying the depositional metadata from the PBDB and IchnoDB datasets. **(B)** Radiometric data points extracted from DateView (https://sil.usask.ca/Databases.htm).

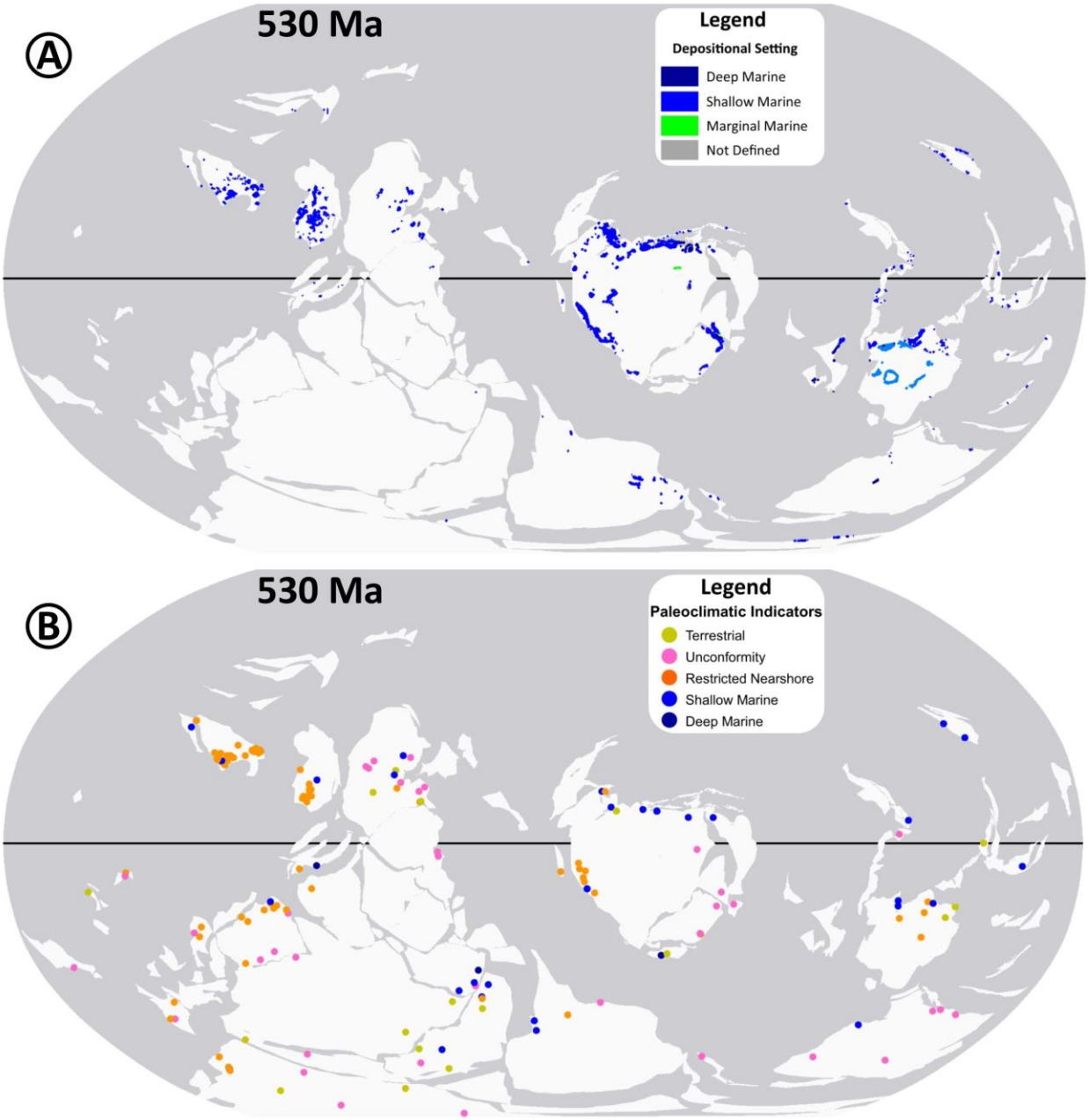


Figure C.3 Supplementary datasets utilized in the creation of paleogeographic reconstructions for the Terreneuvian. **(A)** Geological map data from multiple sources. **(B)** Paleoclimatic indicators from Li et al. (2013) and Boucot et al. (2013).

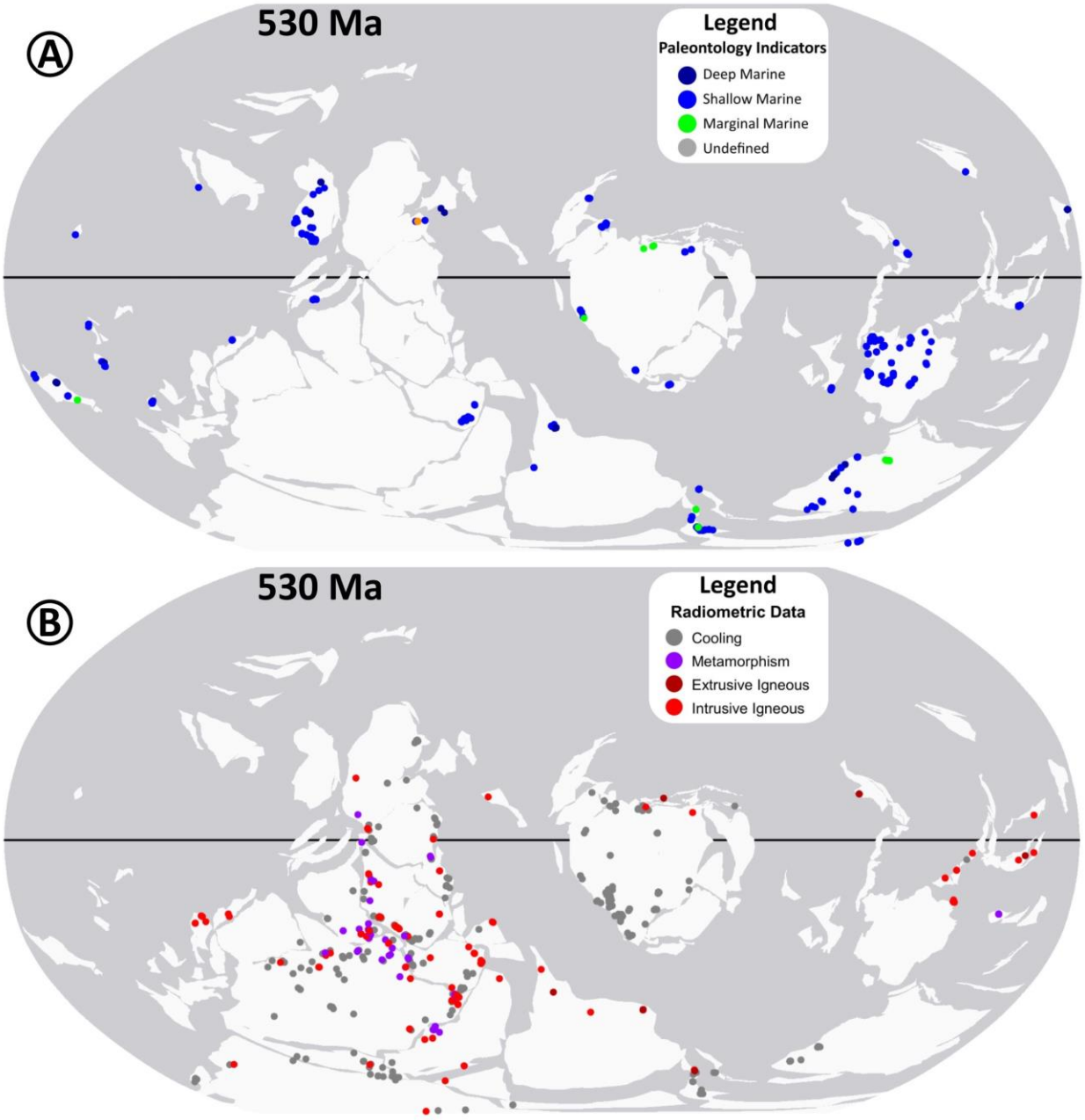


Figure C.4 Supplementary datasets utilized in the creation of paleogeographic reconstructions for the Terreneuvian. **(A)** Paleontology data points displaying the depositional metadata from the PBDB and IchnoDB datasets. **(B)** Radiometric data points extracted from DateView (https://sil.usask.ca/Databases.htm).

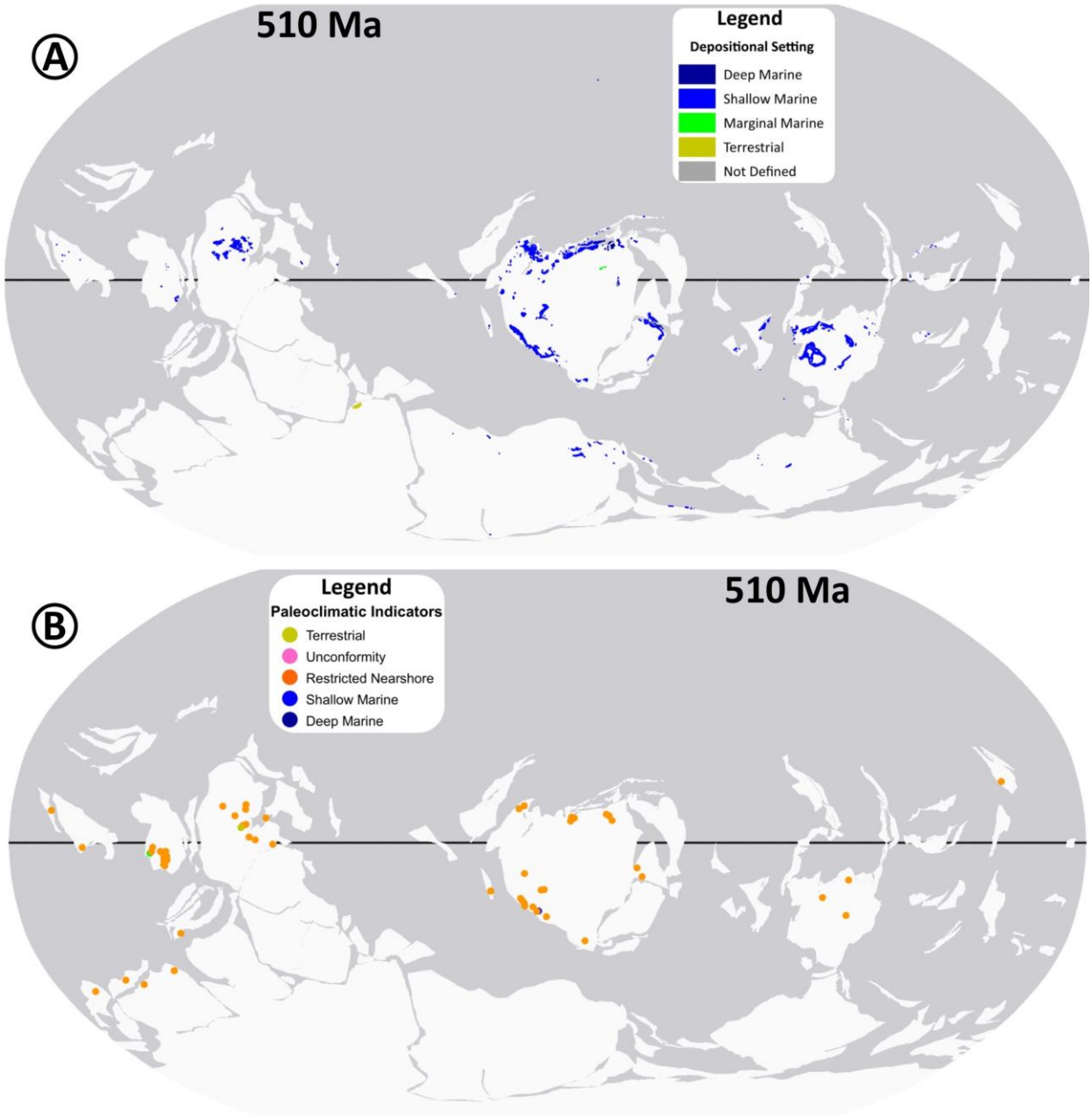


Figure C.5 Supplementary datasets utilized in the creation of paleogeographic reconstructions for the Cambrian Epoch 2. **(A)** Geological map data from multiple sources. **(B)** Paleoclimatic indicators from Boucot et al. (2013).

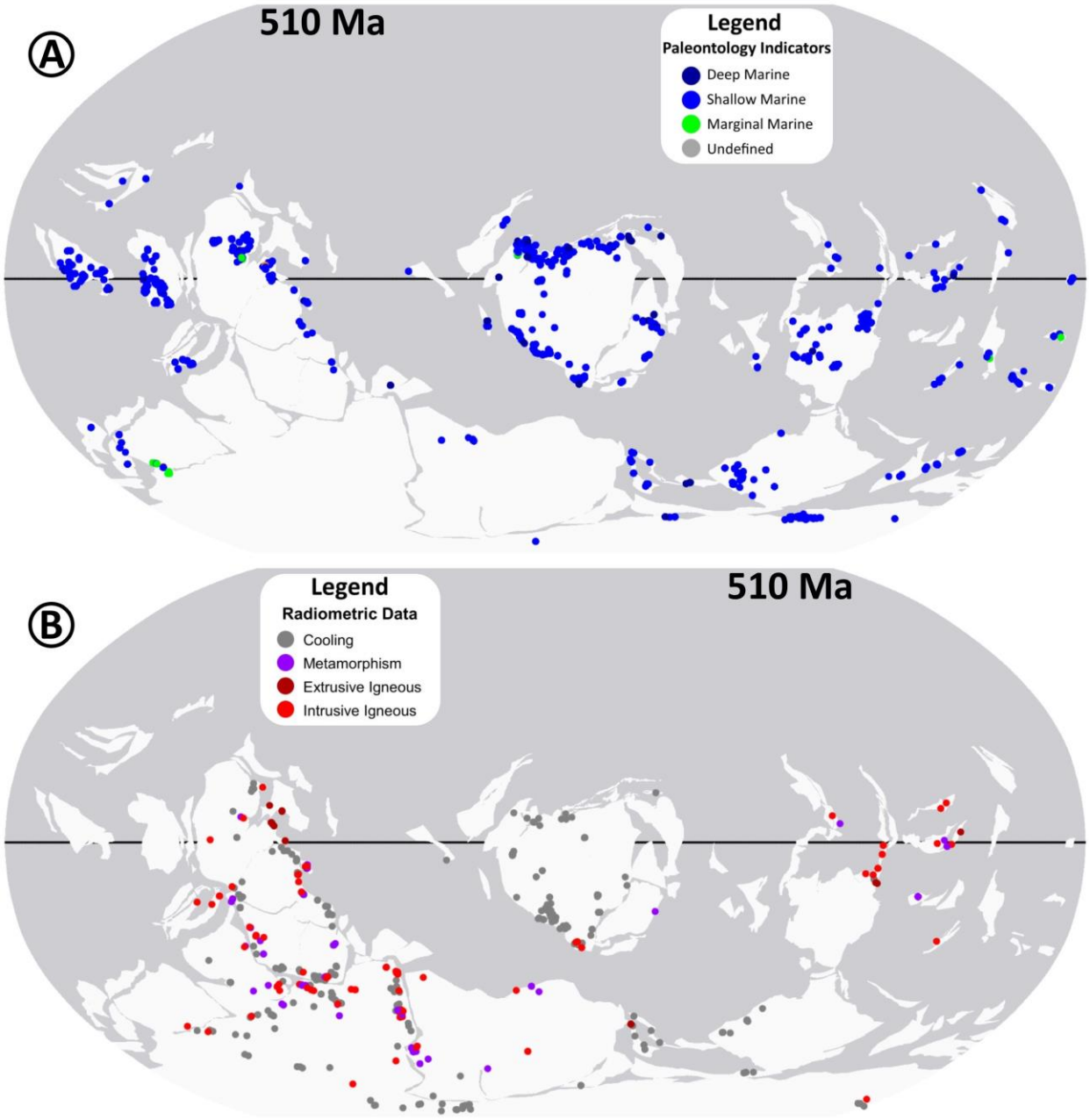


Figure C.6 Supplementary datasets utilized in the creation of paleogeographic reconstructions for the Cambrian Epoch 2. **(A)** Paleontology data points displaying the depositional metadata from the PBDB and IchnoDB datasets. **(B)** Radiometric data points extracted from DateView (https://sil.usask.ca/Databases.htm).

9. References

Alpert, S. P. (1973). *Bergaueria Prantl* (Cambrian and Ordovician), a Probable Actinian Trace Fossil. *Journal of Paleontology, 47*(5), 919–924.

Alroy, J. (2008). Dynamics of origination and extinction in the marine fossil record. *Proceedings of the National Academy of Sciences, 105*, 11536–11542. https://doi.org/10.1073/pnas.0802597105

Alroy, J., Marshall, C. R., Bambach, R. K., Bezusko, K., Foote, M., Fursich, F. T., … Webber, A. (2001). Effects of sampling standardization on estimates of Phanerozoic marine diversification. *Proceedings of the National Academy of Sciences, 98*(11), 6261–6266. https://doi.org/10.1073/pnas.111144698

Amelin, Y., Lee, D. C., Halliday, A. N., & Pidgeon, R. T. (1999). Nature of the Earth's earliest crust from hafnium isotopes in single detrital zircons. *Nature, 399*, 252–255. https://doi.org/10.1038/20426

Andersen, T., Kristoffersen, M., & Elburg, M. A. (2016). How far can we trust provenance and crustal evolution information from detrital zircons? A South African case study. *Gondwana Research, 34*, 129–148. https://doi.org/10.1016/j.gr.2016.03.003

Arnott, R. W. C. (2010). Deep-Marine Sediments and Sedimentary Systems. In N. P. James & R. W. Dalrymple (Eds.), *Facies Models 4* (pp. 295–322). Geological Association of Canada.

Bambach, R. K. (1983). Ecospace utilization and guilds in marine communities through the Phanerozoic. In M. J. S. Tevesz & P. L. McCall (Eds.), *Biotic Interactions in Recent and Fossil Benthic Communities* (pp. 719–746). New York: Plenus Press.

Batini, C., Ceri, S., & Navathe, S. B. (1992). *Conceptual Database Design: An Entity-Relationship Approach*. Benjamin Cummings.

Belousova, E. A., Kostitsyn, Y. A., Griffin, W. L., Begg, G. C., O'Reilly, S. Y., & Pearson, N. J. (2010). The growth of the continental crust: Constraints from zircon Hf-isotope data. *Lithos, 119*, 457–466. https://doi.org/10.1016/j.lithos.2010.07.024

Belvedere, M., Baucon, A., Turin, S., Mietto, P., Felletti, F., & Muttoni, G. (2011). The impact of the digital trend on ichnology: ICHNOBASE. In *Dinosaur Track Symposium Abstract Book*.

Bertling, M., Braddy, S. J., Bromley, R. G., Demathieu, G. R., Genise, J., Mikuláš, R., … Uchman, A. (2006). Names for trace fossils: A uniform approach. *Lethaia, 39*, 265–286. https://doi.org/10.1080/00241160600787890

Birks, H. J. B. (1987). Multivariate analysis in geology and geochemistry: An introduction. *Chemometrics and Intelligent Laboratory Systems, 2*, 15–28.

Blatt, H., & Jones, R. L. (1975). Proportions of exposed igneous, metamorphic, and sedimentary rocks. *Geological Society of America Bulletin*, *86*, 1085–1088. https://doi.org/10.1130/0016-7606(1975)86<1085:POEIMA>2.0.CO;2

Bonham-Carter, G. F. (1994). *Geographic Information Systems for Geoscientists*. Elsevier.

Bonvalot, S., Balmino, G., Briais, A., Kuhn, M., Peyrefitte, A., Vales, … Sarrailh, M. (2012). World Gravity Map, 1:5,000,000. Paris: BGI-CGMW-CNES-IRD.

Borcard, D., Gillet, F., & Legendre, P. (2011). *Numerical Ecology with R*. (R. Gentleman, K. Hornik, & G. G. Parmigiani, Eds.), *Use R!* New York: Springer. https://doi.org/10.1007/978-1-4419-7976-6

Bosence, D. W., & Wilson, R. C. (2003). Carbonate depositional systems. In A. L. Coe (Ed.), *The Sedimentary Record of Sea-Level Change* (pp. 209–233). Cambridge University Press.

Boucot, A. J., Xu, C., & Scotese, C. R. (2013). *Phanerozoic paleoclimate: An atlas of lithologic indicators of climate*. *Phanerozoic Paleoclimate*. SEPM (Society for Sedimentary Geology). https://doi.org/10.2110/sepmcsp.11

Bowring, S. A., Grotzinger, J. P., Isachsen, C. E., Knoll, A. H., Pelechaty, S. M., & Kolosov, P. (1993). Calibrating rates of early Cambrian evolution. *Science*, *261*(5126), 1293–1298.

Boyden, J., Muller, R., Gurnis, M., Torsvik, T., Clark, J., Turner, M., … Cannon, J. (2011). Next-generation plate-tectonic reconstructions using GPlates. In G. Keller & C. Baru (Eds.), *Geoinformatics: Cyberinfrastructure for the Solid Earth Sciences* (pp. 95–113). Cambridge University Press.

Brand, L., Wang, M., & Chadwick, A. (2015). Global database of paleocurrent trends through the Phanerozoic and Precambrian. *Scientific Data*, *2*(150025). https://doi.org/10.1038/sdata.2015.25

Brett, C. E. (1998). Sequence Stratigraphy, Paleoecology, and Evolution: Biotic Clues and Responses to Sea-Level Fluctuations. *Palaios*, *13*, 241–262.

Bromley, R. G. (1990). *Trace Fossils. Biology and Taphonomy.* London: Unwin Hyman.

Bromley, R. G. (1996). *Trace Fossils. Biology, Taphonomy, and Applications*. London: Chapman & Hall.

Buatois, L. A., & Mángano, M. G. (1993). Ecospace utilization, paleoenvironmental trends, and the evolution of early nonmarine biotas. *Geology*, *21*, 595–598.

Buatois, L. A., & Mángano, M. G. (2003). Early Colonization of the Deep Sea: Ichnologic Evidence of Deep-marine Benthic Ecology from the Early Cambrian of Northwest Argentina. *PALAIOS*, *18*(6), 572–581.

Buatois, L. A., & Mángano, M. G. (2011). *Ichnology: Organism-Substrate Interactions in Space and Time*. Cambridge: Cambridge University Press.

Buatois, L. A., & Mángano, M. G. (2013). Ichnodiversity And Ichnodisparity: Significance And Caveats. *Lethaia*, *46*, 281–292.

Buatois, L. A., & Mángano, M. G. (2016a). Ediacaran Ecosystems and the Dawn of Animals. In M. Gabriela Mángano & L. A. Buatois (Eds.), *The Trace-Fossil Record of Major Evolutionary Events Volume 1: Precambrian and Paleozoic* (pp. 27–72). Springer. https://doi.org/10.1007/978-94-017-9600-2

Buatois, L. A., & Mángano, M. G. (2016b). Recurrent Patterns and Processes: The Significance of Ichnology in Evolutionary Paleoecology. In M. G. Mángano & L. A. Buatois (Eds.), *The Trace-Fossil Record of Major Evolutionary Events Volume 2: Mesozoic and Cenozoic* (pp. 449–474). Springer. https://doi.org/10.1007/978-94-017-9597-5

Buatois, L. A., & Mángano, M. G. (2018). The other biodiversity record: Innovations in animal-substrate interactions through geologic time. *GSA Today*, *28*, 1–7. https://doi.org/10.1130/GSATG371A.1

Buatois, L. A., Mángano, M. G., Maples, C. G., & Lanier, W. P. (1997). The Paradox of Nonmarine Ichnofaunas in Tidal Rhythmites: Integrating Sedimentologic and Ichnologic Data from the Late Carboniferous of Eastern Kansas, USA. *Palaios*, *12*, 467–481. https://doi.org/10.2307/3515384

Buatois, L. A., Mángano, M. G., Olea, R. A., & Wilson, M. A. (2016). Decoupled evolution of soft and hard substrate communities during the Cambrian Explosion and Great Ordovician Biodiversification Event. *Proceedings of the National Academy of Sciences*, *113*(25), 6945–6948. https://doi.org/10.1073/pnas.1523087113

Buatois, L. A., Narbonne, G. M., Mángano, M. G., Carmona, N. B., & Myrow, P. (2014). Ediacaran matground ecology persisted into the earliest Cambrian. *Nature Communications*, *5*(3544). https://doi.org/10.1038/ncomms4544

Buatois, L. A., Wisshak, M., Wilson, M. A., & Mángano, M. G. (2017). Categories of architectural designs in trace fossils: A measure of ichnodisparity. *Earth-Science Reviews*, *164*, 102–181. https://doi.org/10.1016/j.earscirev.2016.08.009

Cai, Y., Hua, H., Schiffbauer, J. D., Sun, B., & Yuan, X. (2014). Tube growth patterns and microbial mat-related lifestyles in the Ediacaran fossil *Cloudina*, Gaojiashan Lagerstätte, South China. *Gondwana Research*, *25*, 1008–1018. https://doi.org/10.1016/j.gr.2012.12.027

Cao, W., Zahirovic, S., Flament, N., Williams, S., Golonka, J., & Dietmar Müller, R. (2017). Improving global paleogeography since the late Paleozoic using paleobiology. *Biogeosciences*, *14*, 5425–5439. https://doi.org/10.5194/bg-14-5425-2017

Carrasco, M., Barnosky, A., Kraatz, B., & Davis, E. (2007). The Miocene MammaL Mapping Project (Miomap): An Online Database of Arikareean Through Hemphillian Fossil Mammals. *Bulletin of Carnegie Museum of Natural History*, 183–188.

Cawood, P. A. (2005). Terra Australis Orogen: Rodinia breakup and development of the Pacific and Iapetus margins of Gondwana during the Neoproterozoic and Paleozoic. *Earth-Science Reviews*, *69*, 249–279. https://doi.org/10.1016/j.earscirev.2004.09.001

Cawood, P. A., Hawkesworth, C. J., & Dhuime, B. (2012). Detrital zircon record and tectonic setting. *Geology*, *40*(10), 875–878. https://doi.org/10.1130/G32945.1

CGMW. (2000a). Geological Map of Namibia, 1:1,000,000.

CGMW. (2000b). Geological Map of South Africa, 1:1,000,000.

CGMW. (2001). Geological Map of South America, 1:5,000,000.

CGMW. (2017). Geological Map of Asia, 1:1,000,000.

Chamberlin, D. D., & Boyce, R. (1974). SEQUEL: A structured English query language. *Proceedings of the 1974 ACM SIGFIDET*, 249–264.

Champion, D. C., & Huston, D. L. (2016). Radiogenic isotopes, ore deposits and metallogenic terranes: Novel approaches based on regional isotopic maps and the mineral systems concept. *Ore Geology Reviews*, *76*, 229–256. https://doi.org/10.1016/j.oregeorev.2015.09.025

Cherniak, D. J., & Watson, E. B. (2000). Pb diffusion in zircon. *Chemical Geology*, *172*, 5–24.

Claramunt, C., & Thériault, M. (1995). Managing Time in GIS An Event-Oriented Approach. In *Recent Advances in Temporal Databases* (pp. 23–42).

Codd, E. F. (1970). A relational model of data for large shared data banks. *Commun. ACM*, *13*(6), 377–387.

Codd, E. F. (1979). Extending the database relational model to capture more meaning. *ACM Transactions on Database Systems*, *4*(4), 397–434.

Cohen, K. M., Finney, S. C., Gibbard, P. L., & Fan, J.-X. (2014). The ICS International Chronostratigraphic Chart. *Episodes*, *36*(3), 199–204. https://doi.org/10.1111/j.1502-3931.1980.tb01026.x

Collette, J. H., Hagadorn, J. W., & Lacelle, M. A. (2010). Dead in their tracks - Cambrian arthropods and their traces from intertidal sandstones of Quebec and Wisconsin. *PALAIOS*, *25*, 475–486. https://doi.org/10.2110/palo.2009.p09-134r

Condie, K. C. (1997). *Plate Tectonics and Crustal Evolution*. Elsevier.

Copi, I. M. (1978). *Introduction to Logic* (Fifth). New York: MacMillan Publishing Co.

Crimes, T. (1994). The Period of Early Evolutionary Failure and the Dawn of Evolutionary Success: the Record of Biotic Changes Across the Precambrian-Cambrian Boundary. In S. Donovan (Ed.), *The Palaeobiology of Trace Fossils* (p. 308). Baltimore: The John Hopkins University Press.

David, M., Campiglio, C., & Darling, R. (1974). Progresses in R- and Q-Mode Analysis: Correspondence Analysis and its Application to the Study of Geological Processes. *Canadian Journal of Earth Sciences*, *11*(1), 131–146.

Davis, J. C. (2002). *Statistics and Data Analysis in Geology* (Third Edit). New York: John Wiley & Sons.

Dhuime, B., Hawkesworth, C. J., Cawood, P. A., & Storey, C. D. (2012). A change in the geodynamics of continental growth 3 billion years ago. *Science*, *335*, 1334–1336. https://doi.org/10.1126/science.1216066

Dodson, M., Compston, W., Williams, I., & Wilson, J. (1988). A search for ancient detrital zircons in Zimbabwean sediments. *Journal of the Geological Society*, *145*, 977–983.

Doe, M. F., Jones, J. V., Karlstrom, K. E., Thrane, K., Frei, D., Gehrels, G., & Pecha, M. (2012). Basin formation near the end of the 1.60-1.45 Ga tectonic gap in southern Laurentia: Mesoproterozoic Hess Canyon Group of Arizona and implications for ca. 1.5 Ga supercontinent configurations. *Lithosphere*, *4*(1), 77–88. https://doi.org/10.1130/L160.1

Eglington, B. M. (2004). DateView: A windows geochronology database. *Computers and Geosciences*, *30*, 847–858.

Eglington, B. M. (2018a). Coran - Multivariate Statistical Analysis Software. Unpublished. Retrieved from http://sil.usask.ca/Software.htm

Eglington, B. M. (2018b). FitPDF. Unpublished. Retrieved from http://sil.usask.ca/Software.htm

Eglington, B. M. (2018c). PalaeoPlates Model. Unpublished.

Eglington, B. M., & Armstrong, R. A. (2004). The Kaapvaal Craton and adjacent orogens, southern Africa: A geochronological database and overview of the geological development of the craton. *South African Journal of Geology*, *107*, 13–32. https://doi.org/10.2113/107.1-2.13

Eglington, B. M., & Harmer, R. E. (1999). *GEODATE for Windows version 1: Isotope regression and modelling software*.

Ekdale, A. A. (1988). Pitfalls of Paleobathymetric Interpretations Based on Trace Fossil Assemblages. *Palaios*, *3*, 464–472. https://doi.org/10.2307/3514720

Ekdale, A. A., Bromley, R. G., & Pemberton, S. G. (1984). *Ichnology: the use of trace fossils in sedimentology and stratigraphy*. Society of Economic Paleontologists and Mineralogists.

England, P. C., & Thompson, A. B. (1984). Pressure - Temperature - Time Paths of Regional Metamorphism I. Heat Transfer during the Evolution of Regions of Thickened Continental Crust. *Journal of Petrology*, *25*, 894–928. https://doi.org/10.1093/petrology/25.4.894

Fan, J., Chen, Q., Hou, X., Miller, A. I., Melchin, M. J., Shen, S., … Zhang, L. (2013). Geobiodiversity Database: a comprehensive section-based integration of stratigraphic and paleontological data. *Newsletters on Stratigraphy*, *46*(2), 111–136. https://doi.org/10.1127/0078-0421/2013/0033

Fedo, C. M., Sircombe, K. N., & Rainbird, R. H. (2003). Detrital Zircon Analysis of the Sedimentary Record. *Reviews in Mineralogy and Geochemistry*, *53*(1), 277–303. https://doi.org/10.2113/0530277

Fisher, C. M., Vervoort, J. D., & Hanchar, J. M. (2014). Guidelines for reporting zircon Hf isotopic data by LA-MC-ICPMS and potential pitfalls in the interpretation of these data. *Chemical Geology, 363*, 125–133. https://doi.org/10.1016/j.chemgeo.2013.10.019

Frey, R. W. (1975). The realm of ichnology, its strengths and limitations. In R. W. Frey (Ed.), *The Study of Trace Fossils. A Synthesis of Principles, Problems, and Procedures in Ichnology* (pp. 13–38). New York: Springer-Verlag.

Frey, R. W., & Pemberton, S. G. (1984). Trace Fossil Facies Models. In *Facies Models* (Second, pp. 189–207). Geological Association of Canada.

Frey, R. W., & Seilacher, A. (1980). Uniformity in marine invertebrate ichnology. *Lethaia*, *13*(3), 183–207. https://doi.org/10.1111/j.1502-3931.1980.tb00632.x

Frey, R. W., & Wheatcroft, R. A. (1989). Organism-substrate relations and their impact on sedimentary petrology. *Journal of Geological Education*, *37*, 261–279.

Frost, B. R., & Frost, C. D. (2014). *Essentials of Igneous and Metamorphic Petrology*. Cambridge University Press.

Frost, C. D., & Winston, D. (1987). Nd Isotope Systematics of Coarse- and Fine-Grained Sediments : Examples from the Middle Proterozoic Belt-Purcell. *The Journal of Geology*, *95*(3), 309–327. https://doi.org/10.1086/521238

GA. (2012). Australian Geological Map, 1:1,000,000. Retrieved from https://ecat.ga.gov.au/geonetwork/srv/eng/catalog.search?node=srv#/metadata/c8856c41-0d5b-2b1d-e044-00144fdd4fa6

Gehrels, G. (2000). Introduction to detrital zircon studies of Paleozoic and Triassic strata in western Nevada and northern California. In M. Soreghan & G. Gehrels (Eds.), *Paleozoic and Triassic paleogeography and tectonics of western Nevada and northern California* (pp. 1–17). Boulder: Geological Society of America Special Paper 347.

Gehrels, G., & Pecha, M. (2014). Detrital zircon U-Pb geochronology and Hf isotope geochemistry of Paleozoic and Triassic passive margin strata of western North America. *Geosphere*, *10*(1), 49–65. https://doi.org/10.1130/GES00889.1

Goldstein, D., McKinney, M., Pacaud, J.-M., & Merle, D. (2010). The Ichnology Database. Facilitating Research and Educational Access with an Integrated Database. In *2010 GSA Denver Annual Meeting*.

Golonka, J. (2012). *Paleozoic paleoenvironment and paleolithofacies maps of Gondwana*. Krakow: AGH University of Science and Technology Press. https://doi.org/10.13140/2.1.3481.4403

Golonka, J., Ross, M. I., & Scotese, C. R. (1994). Phanerozoic Paleogeographic and Paleoclimatic Modeling Maps. *Canadian Society of Petroleum Geologists, Memoir 17*, 1–47.

Gougeon, R. C., Mángano, M. G., Buatois, L. A., Narbonne, G. M., & Laing, B. A. (2018). Early Cambrian origin of the shelf sediment mixed layer. *Nature Communications*, *9*, 1–7. https://doi.org/10.1038/s41467-018-04311-8

Greenacre, M. (2010). *Biplots in Practice*. Fundacion BBVA.

GSA. (2005). Geologic Map of North America, 1:5,000,000.

Hadlari, T., Davis, W. J., & Dewing, K. (2014). A pericratonic model for the Pearya terrane as an extension of the Franklinian margin of Laurentia, Canadian Arctic. *Bulletin of the Geological Society of America*, *126*, 182–200. https://doi.org/10.1130/B30843.1

Hadlari, T., Swindles, G. T., Galloway, J. M., Bell, K. M., Sulphur, K. C., Heaman, L. M., … Fallas, K. M. (2015). 1.8 Billion years of detrital zircon recycling calibrates a refractory part of earth's sedimentary cycle. *PLoS ONE*, *10*(12). https://doi.org/10.1371/journal.pone.0144727

Hagadorn, J. W., & Bottjer, D. J. (1999). Restriction of a Late Neoproterozoic Biotope: Suspect-Microbial Structures and Trace Fossils at the Vendian-Cambrian Transition. *Palaios*, *14*, 73–85. https://doi.org/10.2307/3515362

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (Third Edition). Elsevier. https://doi.org/10.1016/B978-0-12-381479-1.00001-0

Haq, B. U., & Schutter, S. R. (2008). A chronology of paleozoic sea-level changes. *Science*, *322*(5898), 64–68. https://doi.org/10.1126/science.1161648

Harrington, J. (2016). *Relational Database Design and Implementation* (Fourth Edition). San Francisco: Elsevier.

Harrison, J. C., St-Onge, M. R., Petrov, O. V., Strelnikov, S. I., Lopatin, B. G., Wilson, F. H., … Solli, A. (2011). Geological map of the Arctic 2159A. Natural Resources Canada. https://doi.org/10.4095/287868

Harrison, J. E., Griggs, A. B., & Wells, J. D. (1974). Tectonic features of the Precambrian Belt Basin and their influence on post-Belt structures. *U.S. Geological Survey Professional Paper*, *866*, 15.

Hartmann, J., & Moosdorf, N. (2012). The new global lithological map database GLiM: A representation of rock properties at the Earth surface. *Geochemistry, Geophysics, Geosystems*, *13*(12), 1–37. https://doi.org/10.1029/2012GC004370

Hendricks, J. R., Lieberman, B. S., & Stigall, A. L. (2008). Using GIS to study palaeobiogeographic and macroevolutionary patterns in soft-bodied Cambrian arthropods. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *264*, 163–175. https://doi.org/10.1016/j.palaeo.2008.04.014

Holland, S. M. (2008). Non-metric Multidimensional Scaling (MDS). University of Georgia, Athens.

Holland, S. M. (2016). The non-uniformity of fossil preservation. *Philosophical Transactions of the Royal Society B*, *371*(20150130), 1–11. https://doi.org/10.1098/rstb.2015.0130

Hubert, M., & Engelen, S. (2004). Robust PCA and classification in biosciences. *Bioinformatics*, *20*(11), 1728–1736. https://doi.org/10.1093/bioinformatics/bth158

Ireland, T. R., & Williams, I. S. (2003). Considerations in Zircon Geochronology by SIMS. *Reviews in Mineralogy and Geochemistry*, *53*(1), 215–241. https://doi.org/10.2113/0530215

James, N. P., & Jones, B. (2016). *Origin of Carbonate Sedimentary Rocks*. American Geophysical Union.

James, N. P., Kendall, A. C., & Pufahl, P. K. (2010). Introduction to biological and chemical sedimentary facies models. In *Facies Models 4*. https://doi.org/http://dx.doi.org/10.1097/JCP.0b013e3182196e64

James, N. P., & Wood, R. (2010). Reefs. In N. P. James & R. W. Dalrymple (Eds.), *Facies Models 4* (pp. 421–448). Geological Association of Canada.

Jensen, S., Buatois, L., & Mangano, M. (2013). Testing for palaeogeographical patterns in the distribution of Cambrian trace fossils. In D. Harper & T. Servais (Eds.), *Early Palaeozoic Biogeography and Palaeogeography* (pp. 45–58). London: The Geological Society of London, Memoirs.

Jensen, S., Droser, M. L., & Gehling, J. G. (2006). A Critical Look at the Ediacaran Trace Fossil Record. In S. Xiao & A.J. Kaufman (Eds.), *Neoproterozoic Geobiology and Paleobiology* (pp. 115-157). Springer.

Jensen, S., Palacios, T., & Martí-Mus, M. (2007). A brief review of the fossil record of the Ediacaran-Cambrian transition in the area of Montes de Toledo-Guadalupe, Spain. In P. Vickers-Rich & P. Komarower (Eds.), *The Rise and Fall of the Ediacaran Biota* (pp. 223–235). Geological Society of London. https://doi.org/10.1144/SP286.16

Jeram, A. J., Selden, P. A., & Edwards, D. (1990). Land animals in the Silurian: Arachnids and myriapods from Shropshire, England. *Science*, *250*, 658–661. https://doi.org/10.1126/science.250.4981.658

Kinny, P. D., & Maas, R. (2003). Lu-Hf and Sm-Nd isotope systems in zircon. *Reviews in Mineralogy and Geochemistry*, *53*(1), 327–341. https://doi.org/10.2113/0530327

Knaust, D. (2017). *Atlas of Trace Fossils in Well Core: Appearance, Taxonomy and Interpretation*. Springer.

Košler, J., Fonneland, H., Sylvester, P., Tubrett, M., & Pedersen, R. (2002). U – Pb dating of detrital zircons for sediment provenance studies - a comparison of laser ablation ICPMS and SIMS techniques. *Chemical Geology*, *182*, 605–618. https://doi.org/10.1016/S0009-2541(01)00341-2

Košler, J., & Sylvester, P. J. (2003). Present Trends and the Future of Zircon in Geochronology: Laser Ablation ICPMS. *Reviews in Mineralogy and Geochemistry*, *53*(1), 243–275. https://doi.org/10.2113/0530243

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, *29*, 1–27. https://doi.org/10.1007/BF02289565

Laing, B. A., Buatois, L. A., Mángano, M. G., Narbonne, G. M., & Gougeon, R. C. (2018). *Gyrolithes* from the Ediacaran-Cambrian boundary section in Fortune Head, Newfoundland, Canada: Exploring the onset of complex burrowing. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *495*, 171–185. https://doi.org/10.1016/j.palaeo.2018.01.010

Lam, A. R., Stigall, A. L., & Matzke, N. J. (2018). Dispersal in the Ordovician: Speciation patterns and paleobiogeographic analyses of brachiopods and trilobites. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *489*, 147–165. https://doi.org/10.1016/j.palaeo.2017.10.006

Lancaster, P. J., Storey, C. D., Hawkesworth, C. J., & Dhuime, B. (2011). Understanding the roles of crustal growth and preservation in the detrital zircon record. *Earth and Planetary Science Letters*, *305*, 405–412. https://doi.org/10.1016/j.epsl.2011.03.022

Legendre, P., & Legendre, L. (2012). *Numerical Ecology*. *Developments in Environmental Modelling* (Third Edition). Amsterdam: Elsevier. https://doi.org/10.1016/B978-0-444-53868-0.50008-3

Lewis, R. S., Vervoort, J. D., Burmester, R. F., & Oswald, P. J. (2010). Detrital zircon analysis of Mesoproterozoic and Neoproterozoic metasedimentary rocks of north-central Idaho: implications for development of the Belt–Purcell basin. *Canadian Journal of Earth Sciences*, *47*, 1383–1404. https://doi.org/10.1139/E10-049

Li, Z.-X., Evans, D. A. D., & Halverson, G. P. (2013). Neoproterozoic glaciations in a revised global palaeogeography from the breakup of Rodinia to the assembly of Gondwanaland. *Sedimentary Geology*, *294*, 219–232. https://doi.org/10.1016/j.sedgeo.2013.05.016

Lloyd-Williams, M. (1999). Empirical studies of the knowledge discovery approach to health-information analysis. In M. Bramer (Ed.), *Knowledge Discovery and Data Mining* (pp. 139–159). The Institution of Electrical Engineers.

Logan, M. (2011). Multivariate analyses. In *Tutorials on R and statistics*. Retrieved from http://www.flutterbys.com.au/stats/tut/

MacNaughton, R. B., Cole, J. M., Dalrymple, R. W., Braddy, S. J., Briggs, D. E. G., & Lukie, T. D. (2002). First steps on land: Arthropod trackways in Cambrian-Ordovician eolian sandstone, southeastern Ontario, Canada. *Geology*, *30*(5), 391–394.

MacNaughton, R. B., Moynihan, D. P., Roots, C. F., & Crowley, J. L. (2016). New Occurrences of *Oldhamia* in Eastern Yukon, Canada: Stratigraphic Context and Implications for Cambrian Deep-Marine Biostratigraphy. *Ichnos*, *23*(1–2), 33–52. https://doi.org/10.1080/10420940.2015.1127232

Malusà, M. G., Carter, A., Limoncelli, M., Villa, I. M., & Garzanti, E. (2013). Bias in detrital zircon geochronology and thermochronometry. *Chemical Geology*, *359*, 90–107. https://doi.org/10.1016/j.chemgeo.2013.09.016

Malusà, M. G., Resentini, A., & Garzanti, E. (2016). Hydraulic sorting and mineral fertility bias in detrital geochronology. *Gondwana Research*, *31*, 1–19. https://doi.org/10.1016/j.gr.2015.09.002

Mángano, M. G., & Buatois, L. A. (2004). Reconstructing Early Phanerozoic intertidal ecosystems: ichnology of the Cambrian Campanario Formation in northwest Argentina. *Fossils and Strata*, *51*, 17–38.

Mángano, M. G., & Buatois, L. A. (2014). Decoupling of body-plan diversification and ecological structuring during the Ediacaran-Cambrian transition: evolutionary and geobiological feedbacks. *Proceedings of the Royal Society B: Biological Sciences*, *281*, 1–9. https://doi.org/10.1098/rspb.2014.0038

Mángano, M. G., & Buatois, L. A. (2015). The trace-fossil record of tidal flats through the Phanerozoic: Evolutionary innovations and faunal turnover. In *Ichnia III* (pp. 157–177). Geological Association of Canada.

Mángano, M. G., & Buatois, L. A. (2017). The Cambrian revolutions: Trace-fossil record, timing, links and geobiological impact. *Earth-Science Reviews*, *173*, 96–108. https://doi.org/10.1016/j.earscirev.2017.08.009

Mángano, M. G., Buatois, L. A., Hofmann, R., Elicki, O., & Shinaq, R. (2013). Exploring the aftermath of the Cambrian explosion: The evolutionary significance of marginal- to shallow-marine ichnofaunas of Jordan. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *374*, 1–15. https://doi.org/10.1016/j.palaeo.2012.05.029

Mannila, H., & Räihä, K.-J. (1992). *The Design of Relational Databases*. Addison-Wesley.

Mao, Y. (2018). *Geodynamic Development of the South China Block from Precambrian to Cretaceous: Constraints from Geology, Geochemistry, and Geochronology*. University of Saskatchewan.

Markwick, P. J., & Lupia, R. (2002). Palaeontological databases for palaeobiogeography, palaeoecology and biodiversity: a question of scale. In J. A. Crame & A. W. Owen (Eds.), *Palaeobiogeography and Biodiversity Change: the Ordovician and Mesozoic-Cenozoic Radiations.* (pp. 169–178). The Geological Society of London, Special Publications. https://doi.org/10.1144/GSL.SP.2002.194.01.13

Matthews, K. J., Maloney, K. T., Zahirovic, S., Williams, S. E., Seton, M., & Müller, R. D. (2016). Global plate boundary evolution and kinematics since the late Paleozoic. *Global and Planetary Change*, *146*, 226–250. https://doi.org/10.1016/j.gloplacha.2016.10.002

McCall, G. J. H. (2006). The Vendian (Ediacaran) in the geological record: Enigmas in geology's prelude to the Cambrian explosion. *Earth-Science Reviews*, *77*, 1–229.

Meyers, S. R., & Peters, S. E. (2011). A 56 million year rhythm in North American sedimentation during the Phanerozoic. *Earth and Planetary Science Letters*, *303*, 174–180. https://doi.org/10.1016/j.epsl.2010.12.044

Miller, M. F., & Smail, S. E. (1997). A Semiquantitative Field Method for Evaluating Bioturbation on Bedding Planes. *Palaios*, *12*, 391–396.

Minter, N. J., Buatois, L. A., & Mángano, M. G. (2016a). The Conceptual and Methodological Tools of Ichnology. In M. G. Mángano & L. A. Buatois (Eds.), *The Trace-Fossil Record of Major Evolutionary Events Volume 1: Precambrian and Paleozoic* (pp. 1–26). Springer. https://doi.org/10.1007/978-94-017-9600-2

Minter, N. J., Buatois, L. A., Mángano, M. G., MacNaughton, R. B., Davies, N. S., & Gibling, M. R. (2016b). The Prelude to Continental Invasion. In M. G. Mángano & L. A. Buatois (Eds.), *The Trace-Fossil Record of Major Evolutionary Events Volume 1: Precambrian and Paleozoic* (pp. 157–204). Springer. https://doi.org/10.1007/978-94-017-9600-2

Minter, N. J., Buatois, L. A., Mángano, M. G., Davies, N. S., Gibling, M. R., MacNaughton, R. B., & Labandeira, C. C. (2017). Early bursts of diversification defined the faunal colonization of land. *Nature Ecology & Evolution*, *1*(0175), 1–10. https://doi.org/10.1038/s41559-017-0175

Morgenthaler, S. (2009). Exploratory data analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*. John Wiley & Sons.

Motulsky, H. (2014). *Intuitive biostatistics: a nonmathematical guide to statistical thinking*. *Oxford University Press* (Third Edition). New York: Oxford University Press.

Müller, R. D., Sdrolias, M., Gaina, C., & Roest, W. R. (2008). Age, spreading rates, and spreading asymmetry of the world's ocean crust. *Geochemistry, Geophysics, Geosystems*, *9*(Q04006), 1–19. https://doi.org/10.1029/2007GC001743

Nance, R. D., & Murphy, J. B. (2013). Origins of the supercontinent cycle. *Geoscience Frontiers*, *4*, 439–448. https://doi.org/10.1016/j.gsf.2012.12.007

Oksanen, J., Guillaume-Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., … Wagner, H. (2018). vegan: Community Ecology Package. R package version 2.5-1. Retrieved from https://cran.r-project.org/package=vegan

Orr, P. J. (2001). Colonization of the deep-marine environment during the early Phanerozoic: The ichnofaunal record. *Geological Journal*, *36*(3–4), 265–278. https://doi.org/10.1002/gj.891

Patchett, P. J., & Tatsumoto, M. (1981). A routine high-precision method for Lu-Hf isotope geochemistry and chronology. *Contributions to Mineralogy and Petrology, 75*(3), 263–267. https://doi.org/10.1007/BF01166766

Pemberton, S. G., Frey, R., Ranger, M., & MacEachern, J. (1992). Conceptual framework of ichnology. In *SEPM Core Workshop.* (pp. 1–32). Society of Economic Paleontologists and Mineralogists.

Pemberton, S. G., & Frey, R. W. (1982). Trace fossil nomenclature and the *Planolites–Palaeophycus* dilemma. *Journal of Paleontology*, *54*(4), 843–881.

Pemberton, S. G., MacEachern, J., & Frey, R. W. (1992). Trace fossils facies models: environmental and allostratigraphic significance. In R. G. Walker & N. P. James (Eds.), *Facies Models and Sea Level Changes* (pp. 47–72). Geological Society of America Abstracts with Programs, v.31.

Peters, S. E. (2005). Geologic constraints on the macroevolutionary history of marine animals. *Proceedings of the National Academy of Sciences*, *102*(35), 12326–12331. https://doi.org/10.1073/pnas.0502616102

Peters, S. E., & Heim, N. A. (2011). Stratigraphic distribution of marine fossils in North America. *Geology*, *39*(3), 259–262. https://doi.org/10.1130/G31442.1

Peters, S. E., & McClennen, M. (2016). The Paleobiology Database application programming interface. *Paleobiology*, *42*(01), 1–7. https://doi.org/10.1017/pab.2015.39

Peters, S. E., Zhang, C., Livny, M., & Ré, C. (2014). A machine reading system for assembling synthetic paleontological databases. *PLoS ONE*, *9*(12), 1–22. https://doi.org/10.1371/journal.pone.0113523

Pohl, A., Nardin, E., Vandenbroucke, T. R. A., & Donnadieu, Y. (2016). High dependence of Ordovician ocean surface circulation on atmospheric CO_2 levels. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *458*, 39–51. https://doi.org/10.1016/j.palaeo.2015.09.036

Pu, J. P., Bowring, S. A., Ramezani, J., Myrow, P., Raub, T. D., Landing, E., … Macdonald, F. A. (2016). Dodging snowballs: Geochronology of the Gaskiers glaciation and the first appearance of the Ediacaran biota. *Geology*, *44*(11), 955–958. https://doi.org/10.1130/G38284.1

Puetz, S. J., Ganade, C. E., Zimmermann, U., & Borchardt, G. (2018). Statistical analyses of Global U-Pb Database 2017. *Geoscience Frontiers*, *9*, 121–145. https://doi.org/10.1016/j.gsf.2017.06.001

Qin, X., Müller, R. D., Cannon, J., Landgrebe, T. C. W., Heine, C., Watson, R. J., & Turner, M. (2012). The GPlates Geological Information Model and Markup Language. *Geoscientific Instrumentation, Methods and Data Systems*, *2*, 365–428.

Quinn, G. P., & Keough, M. J. (2002). *Experimental Design and Data Analysis for Biologists*. Cambridge: Cambridge University Press. https://doi.org/10.1016/S0022-0981(02)00278-2

R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Rainbird, R. H., Heaman, L. M., & Young, G. (1992). Sampling Laurentia: detrital zircon geochronology offers evidence for an extensive Neoproterozoic river system originating from the Grenville orogen. *Geology*, *20*, 351–354.

Rainbird, R. H., McNicoll, V. J., Thériault, R. J., Heaman, L. M., Abbott, J. G., Long, D. G. F., & Thorkelson, D. J. (1997). Pan-Continental River System Draining Grenville Orogen Recorded by U-Pb and Sm-Nd Geochronology of Neoproterozoic Quartzarenites and Mudrocks, Northwestern Canada. *The Journal of Geology*, *105*(1), 1–17. https://doi.org/10.1086/606144

Rainbird, R. H., Rayner, N. M., Hadlari, T., Heaman, L. M., Ielpi, A., Turner, E. C., & MacNaughton, R. B. (2017). Zircon provenance data record the lateral extent of pancontinental, early Neoproterozoic rivers and erosional unroofing history of the Grenville orogen. *Bulletin of the Geological Society of America*, *129*, 1408–1423. https://doi.org/10.1130/B31695.1

Reineck, H. E. (1963). Sedimentgefuge im Bereich der sudlichen Nordsee. *Abhandlungen Der Senckenbergische Naturforschende Gesellschaft*, *505*, 1–138.

Reiners, P. W., Ehlers, T. A., & Zeitler, P. K. (2005). Past, Present, and Future of Thermochronology. *Reviews in Mineralogy & Geochemistry*, *58*, 1–18. https://doi.org/10.1016/j.jmig.2010.03.005

Roberts, N. M. W., & Spencer, C. J. (2015). The zircon archive of continent formation through time. In N. M. W. Roberts, M. Van Kranendonk, S. Parman, S. Shirey, & P. D. Clift (Eds.), *Continent Formation Through Time* (pp. 197–225). Geological Society of America Special Paper 389. https://doi.org/10.1144/SP389.14

Rode, A. L., & Lieberman, B. S. (2004). Using GIS to unlock the interactions between biogeography, environment, and evolution in Middle and Late Devonian brachiopods and bivalves. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *211*, 345–359. https://doi.org/10.1016/j.palaeo.2004.05.013

Ross, G. M., Parrish, R. R., & Winston, D. (1992). Provenance and U-Pb geochronology of the Mesoproterozoic Belt Supergroup (northwestern United States): implications for age of deposition and pre-Panthalassa plate reconstructions. *Earth and Planetary Science Letters*, *113*, 57–76.

Ross, G. M., & Villeneuve, M. (2003). Provenance of the Mesoproterozoic (1.45 Ga) Belt basin (western North America): Another piece in the pre-Rodinia paleogeographic puzzle. *GSA Bulletin*, *115*(10), 1191–1217.

Rowley, D. B. (2002). Rate of plate creation and destruction: 180 Ma to present. *Geological Society of America Bulletin*, *114*(8), 927–933.

Schiøtte, L., Compston, W., & Bridgwater, D. (1988). Late Archaean ages for the deposition of clastic sediments belonging to the Malene supracrustals, southern West Greenland: evidence from an ion probe UPb zircon study. *Earth and Planetary Science Letters*, *87*, 45–58.

Schreiber, B. C., & El Tabakh, M. (2000). Deposition and early alteration of evaporites. *Sedimentology*, *47*, 215–238. https://doi.org/10.1046/j.1365-3091.2000.00002.x

Schuiling, R. D., De Meijer, R. J., Riezebos, H. J., & Scholten, M. J. (1985). Grain size distribution of different minerals in a sediment as a function of their specific density. *Geologie En Mijnbouw*, *64*(2), 199–203.

Seilacher, A. (1953). Studien zur Palichnologie. I. Uber die Methoden der Palichnologie. *Neues Jahrbuch Fur Geologie Und Palaontologie, Abhandlungen*, *96*, 421–452.

Seilacher, A. (1955a). Spuren und Fazies im Unterkambrium. In O. H. Schindewolf & A. Seilacher (Eds.), *Beitrage zur Kenntnis des Kambriums in der Salt Range (Pakistan)* (10th ed., pp. 373–399). Akademie der Wissenschaften und der Literatur zu Mainz, Mathematisch-Naturwissenschaftliche Klasse, Abhandlungen.

Seilacher, A. (1955b). Spuren und Lebensspuren der Trilobiten. In O. H. Schindewolf & A. Seilacher (Eds.), *Beitrage zur Kenntnis des Kambriums in der Salt Range (Pakistan)* (10th ed., pp. 342–372). Akademie der Wissenschaften und der Literatur zu Mainz, Mathematisch-Naturwissenschaftliche Klasse, Abhandlungen.

Seilacher, A. (1963). Lebensspuren und sakinitatsfazies. *Fortschritte in Der Geologie Rheinland Und Westfalen*, *10*, 81–94.

Seilacher, A. (1964). Biogenic Sedimentary Structures. In J. Imbrie & N. Newell (Eds.), *Approaches to Paleoecology* (pp. 296–316). Chichester, UK: Wiley.

Seilacher, A. (1967a). Bathymetry of trace fossils. *Marine Geology*, *5*, 413–428. https://doi.org/10.1016/0025-3227(67)90051-5

Seilacher, A. (1967b). Fossil Behavior. *Scientific American*, *217*, 72–80.

Seilacher, A. (1999). Biomat-related lifestyles in the Precambrian. *Palaios*, *14*, 86–93. https://doi.org/10.2307/3515363

Seilacher, A., Buatois, L. A., & Mángano, M. G. (2005). Trace fossils in the Ediacaran-Cambrian transition: Behavioral diversification, ecological turnover and environmental shift. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *227*, 323–356.

Seilacher, A., & Pflüger, F. (1994). From biomats to benthic agriculture: A biohistoric revolution. In W. E. Krumbein, D. M. Paterson, & L. J. Stal (Eds.), *Biostabilization of Sediments* (pp. 97–105). Bibliotheks und Informationssystem der Carl con Ossietzky Universitat Oldenburg.

Sepkoski, J. J. (1988). Alpha, beta, or gamma: where does all the diversity go? *Paleobiology*, *14*(3), 221–234.

Shaw, P. J. A. (2003). *Multivariate statistics for the Environmental Sciences*. *New York*. New York: Arnold.

Sircombe, K. N. (2004). AgeDisplay: An EXCEL workbook to evaluate and display univariate geochronological data using binned frequency histograms and probability density distributions. *Computers and Geosciences*, *30*, 21–31. https://doi.org/10.1016/j.cageo.2003.09.006

Spencer, C. J., & Kirkland, C. L. (2016). Visualizing the sedimentary response through the orogenic cycle: A multidimensional scaling approach. *Lithosphere*, *8*(1), 29–37. https://doi.org/10.1130/L479.1

Spencer, C. J., Kirkland, C. L., & Taylor, R. J. M. (2016). Strategies towards statistically robust interpretations of in situ U-Pb zircon geochronology. *Geoscience Frontiers*, *7*, 581–589. https://doi.org/10.1016/j.gsf.2015.11.006

Stevens, T., Carter, A., Watson, T. P., Vermeesch, P., Andò, S., Bird, A. F., … Sevastjanova, I. (2013). Genetic linkage between the Yellow River, the Mu Us desert and the Chinese Loess Plateau. *Quaternary Science Reviews*, *78*, 355–368. https://doi.org/10.1016/j.quascirev.2012.11.032

Stigall, A. L., Bauer, J. E., & Brame, H. M. R. (2014). The digital Atlas of Ordovician life: Digitizing and mobilizing data for paleontologists and the public. *Estonian Journal of Earth Sciences*, *63*(4), 312–316. https://doi.org/10.3176/earth.2014.36

Stigall, A. L., & Lieberman, B. S. (2006). Quantitative palaeobiogeography: GIS, phylogenetic biogeographical analysis, and conservation insights. *Journal of Biogeography*, *33*, 2051–2060. https://doi.org/10.1111/j.1365-2699.2006.01585.x

Systra, Y., & Jensen, S. (2006). Trace fossils from the Dividalen Group of northern Finland with remarks on early Cambrian trace fossil provincialism. *GFF*, *128*, 321–325.

Taylor, A., & Goldring, R. (1993). Description and analysis of bioturbation and ichnofabric. *Journal of the Geological Society*, *150*, 141–148. https://doi.org/10.1144/gsjgs.150.1.0141

Tukey, J. W. (1977). *Exploratory Data Analysis*. *Analysis*. Addison-Wesley.

Uchman, A. (2004). Phanerozoic history of deep-sea trace fossils. *Geological Society, London, Special Publications*, *228*(1), 125–139. https://doi.org/10.1144/GSL.SP.2004.228.01.07

Uhen, M. D., Barnosky, A. D., Bills, B., Blois, J., Carrano, M. T., Carrasco, M. A., … Säilä, L. K. (2013). From card catalogs to computers: Databases in vertebrate paleontology. *Journal of Vertebrate Paleontology*, *33*(1), 13–28. https://doi.org/10.1080/02724634.2012.716114

Vallon, L. H., Rindsberg, A. K., & Bromley, R. G. (2016). An updated classification of animal behaviour preserved in substrates. *Geodinamica Acta*, *28*(1–2), 5–20. https://doi.org/10.1080/09853111.2015.1065306

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (Fourth Edition). New York: Springer.

Vermeesch, P. (2012). On the visualisation of detrital age distributions. *Chemical Geology*, *312–313*, 190–194. https://doi.org/10.1016/j.chemgeo.2012.04.021

Vermeesch, P. (2013). Multi-sample comparison of detrital age distributions. *Chemical Geology*, *341*, 140–146. https://doi.org/10.1016/j.chemgeo.2013.01.010

Vermeesch, P. (2018). Dissimilarity measures in detrital geochronology. *Earth-Science Reviews*, *178*, 310–321. https://doi.org/10.1016/j.earscirev.2017.11.027

Vermeesch, P., & Garzanti, E. (2015). Making geological sense of “Big Data” in sedimentary provenance analysis. *Chemical Geology*, *409*, 20–27. https://doi.org/10.1016/j.chemgeo.2015.05.004

Vermeesch, P., Resentini, A., & Garzanti, E. (2016). An R package for statistical provenance analysis. *Sedimentary Geology*, *336*, 14–25. https://doi.org/10.1016/j.sedgeo.2016.01.009

Vervoort, J. D., Patchett, P. J., Blichert-Toft, J., & Albarède, F. (1999). Relationships between Lu-Hf and Sm-Nd isotopic systems in the global sedimentary system. *Earth and Planetary Science Letters*, *168*, 79–99. https://doi.org/10.1016/S0012-821X(99)00047-3

Voice, P. J., Kowalewski, M., & Eriksson, K. A. (2011). Quantifying the Timing and Rate of Crustal Evolution: Global Compilation of Radiometrically Dated Detrital Zircon Grains. *The Journal of Geology*, *119*, 109–126. https://doi.org/10.1086/658295

Volta, G. S., & Egenhofer, M. J. (1993). Interaction with GIS Attribute Data Based on Categorical Coverages. In A. Frank & I. Campari (Eds.), *Lecture Notes in Computer Science, Vol. 716* (pp. 215–233). Spinger-Verlag.

Whittaker, R. (1972). Evolution and Measurement of Species Diversity. *Taxon*, *21*, 213–251.

Wu, R., Stouge, S., Percival, I. G., & Zhan, R. (2014). Early-Middle Ordovician conodont biofacies on the Yangtze Platform margin, South China: Applications to palaeoenvironment and sea-level changes. *Journal of Asian Earth Sciences*, *96*, 194–204. https://doi.org/10.1016/j.jseaes.2014.09.003

Xiao, S., & Laflamme, M. (2009). On the eve of animal radiation: phylogeny, ecology and evolution of the Ediacara biota. *Trends in Ecology and Evolution*, *24*, 31–40. https://doi.org/10.1016/j.tree.2008.07.015

Xiao, S., Narbonne, G. M., Zhou, C., Laflamme, M., Grazhdankin, D. V., Moczydlowska-Vidal, M., & Cui, H. (2016). Towards an Ediacaran Time Scale: Problems, Protocols, and Prospects. *Episodes*, *39*(4), 540–555. https://doi.org/10.18814/epiiugs/2016/v39i4/103886

Young, G. M. (1979). Correlation of middle and upper Proterozoic strata of the northern rim of the North Atlantic craton. *Transactions of the Royal Society of Edinburgh*, *70*, 323–336.

Young, G. M., Jefferson, C. W., Delaney, G. D., & Yeo, G. M. (1979). Middle and late Proterozoic evolution of the northern Canadian Cordillera and Shield. *Geology*, *7*, 125–128.

Zaffos, A., Finnegan, S., & Peters, S. E. (2017). Plate tectonic regulation of global marine animal diversity. *Proceedings of the National Academy of Sciences*, *114*(22), 5653–5658. https://doi.org/10.1073/pnas.1702297114

Zheng, H., & Hu, Z. (2010). Atlas of Pre-Mesozoic Tectonic Lithofacies Paloegeography in China.