OPENQUATERNARY

Goring, S et al 2015 neotoma: A Programmatic Interface to the Neotoma Paleoecological Database. *Open Quaternary,* 1: 2, pp. 1-17, DOI: http://dx.doi.org/10.5334/oq.ab

METHOD

neotoma: A Programmatic Interface to the Neotoma Paleoecological Database

Simon Goring[1], Andria Dawson[2], Gavin L. Simpson[3], Karthik Ram[4], Russ W. Graham[5], Eric C. Grimm[6] and John W. Williams[1,7]

[1] Department of Geography, University of Wisconsin-Madison, Madison, WI, USA

[2] Department of Statistics, University of California, Berkeley, CA, USA

[3] Institute of Environmental Change and Society, and Department of Biology, University of Regina, Regina, SK, Canada

[4] Berkeley Institute for Data Science, University of California, Berkeley. Berkeley, CA

[5] Department of Geosciences, Pennsylvania State University, University Park, PA, USA

[6] Research and Collections Center, Illinois State Museum, Springfield, IL, USA

[7] Center for Climatic Research, University of Wisconsin-Madison, Madison, WI, USA

Corresponding author: Simon Goring

Paleoecological data are integral to ecological and evolutionary analyses. First, they provide an opportunity to study ecological and evolutionary interactions between communities and abiotic environments across time scales. Second, they allow us to study the long-term outcomes of processes that occur infrequently, such as megadroughts, hurricanes, and rapid climate change. Third, the past allows us to study ecological processes in the absence of widespread anthropogenic influence.

The R package `neotoma`, described here, obtains and manipulates data from the NeotomaPaleoecological Database (Neotoma Database: http://www.neotomadb.org). The Neotoma Database is a public-domain searchable repository for multiproxypaleoecological records spanning the past 5 million years. The Neotoma Database provides the data and cyberinfrastructure to study spatiotemporal dynamics of species and communities from the Pliocene to the present; `neotoma` provides a user interface to enable these studies. `neotoma` searches the Neotoma Database using terms that can include location, taxon name, or dataset type (e.g., pollen, vertebrate fauna, ostracode) using the Database's Application Programming Interface (API). The package returns a set of nested metadata associated with the site, including the full assemblage record, geochronological data to enable the rebuilding of age models, dataset metadata (e.g. age range of samples, date of accession into Neotoma, principal investigator), and site metadata (e.g. location, site name and description). `neotoma` also provides tools to allow cross-site analysis, including the ability to standardize taxonomies using built-in taxonomies derived from the published literature or user-provided taxonomies.

We show how key functions in the `neotoma` package can be used, by reproducing analytic examples from the published literature focusing on *Pinus* migration following deglaciation and shifts in mammal species distributions during the Pleistocene.

Keywords: R software; neotoma; paleoecology; database; pollen; mammal

Introduction

Paleoecological data are fundamental to understanding the patterns and drivers of biogeographical, climatic, and evolutionary change, ranging from the recent past to the dawn of life. Although individual site-level studies have provided significant insights into past ecological dynamics, the true power of paleoecological data emerges from networks of data assembled to study broad-scale ecological and evolutionary phenomena allowing us to understand ecological and evolutionary processes and patterns through deep time (Alroy et al.,

2008; Barnosky et al., 2011; Raup and Sepkoski, 1984; Sepkoski, 1997), and the rapid and individualistic responses of species to the climate changes accompanying recent glacial-interglacial cycles (Davis, 1981; Graham et al., 1996; Huntley and Webb, 1988; Tzedakis, 1994; Williams et al., 2004). Paleoecoinformatics (Brewer et al., 2012; Uhen et al., 2013) is dedicated to providing tools to researchers across disciplines to access and use large paleoecological datasets spanning thousands of years. These datasets may be used to provide better insight into regional vegetation change (Blarquez, Carcaillet, et al., 2014; Blois et al., 2013), patterns of biomass burning (Marlon et al., 2013), or changing rates of geophysical processes through time (Goring et al., 2012). The increasing interest in uniting ecological and paleoecological data to in order to better understand responses to a rapidly changing world (Behrensmeyer and Miller, 2012; Dietl and Flessa, 2011; Fritz et al., 2013) will require more robust tools to access and synthesize data from the modern and paleo time domains.

The NeotomaPaleoecological Database represents a consortium of paleoecological databases spanning the Pliocene to the Holocene (and Anthropocene), with distributed scientific governance and expertise, but sharing a common database infrastructure. Constituent databases include: the European, Latin American, and North American Pollen Databases; the North American Plant Macrofossil Database; FAUNMAP (Pliocene to Quaternary mammal fossils in the United States and Canada); the North Dakota State University Fossil Insect Database; the North American Non-Marine Ostrocode Database (NANODe); and the Diatom Paleolimnology Data Cooperative. More information is available from the Neotoma Working Group webpage (http://www.neotomadb.org/groups).

Neotoma is the outgrowth of a longstanding collaboration between the European Pollen Database and the North American Pollen Database (Grimm et al., 2013) and the desire to integrate these data with faunal and other paleo data. The database framework (Grimm, 2008) was generalized from the pollen databases (which had identical structures) and the FAUNMAP database to accomodate both macro- and microfossil data as well as other kinds of data such as geochemical, isotopic, and loss-on-ignition. Work is underway to include other taxonomic groups and depositional contexts (e.g. testate amoeba records, packrat midden data), thus further expanding the data that can be accomodated by Neotoma. Crucially, Neotoma is a vetted database. Through the use of data stewards—domain experts distributed among constituent bases who can check for inaccuracies, upload and manage data records—Neotoma can support quality control assurance for each of the constituent data types, and receive feedback from research communities involved with each specific data type (Grimm et al., 2013).

The Neotoma Database has also developed an Application Programming Interface (API) that allows users to query the database via web services, which return data using properly formed URL requests. For example, the URL: http://api.neotomadb.org/v1/data/datasets?siteid=960 will return all dataset information for the record associated with site 960, Guardipee Lake (Barnosky, 1989), originally published by Cathy Whitlock.

The analysis of paleoecological data is often performed using the statistical software R (R Core Team, 2014). There are several R packages designed specifically for paleoecological data analysis, including `analogue` (Simpson, 2007; Simpson and Oksanen, 2014) and `rioja` (Juggins, 2013) for paleoenvironmental reconstruction, `Bchron` (Parnell, 2014) for radiocarbon dating and age-depth modeling and `paleofire` to access and analyse charcoal data (Blarquez, Vannière, et al., 2014). Given the rapid proliferation and availability of these analytical tools in R, the rate-limiting step for analysis has become the difficulty of obtaining and importing data into R. This bottleneck has meant reliance on static datasets that are available as published, and rarely updated, and on more *ad hoc* methods such as the distribution of individual datasets from author to analyst.

With increasing pressure on the ecological community to perform and publish reproducible research that includes numerically reproducible results (Goring et al., 2013, 2012; Reichman et al., 2011; Wolkovich et al., 2012) it is important to provide tools that allow analysts to directly access dynamic datasets, and to provide tools to support reproducible workflows. The rOpenSci project (http://ropensci.org) is dedicated to developing tools that use R to facilitate a culture shift toward reproducible science in the ecology community. As part of this effort, rOpenSci provides a number of tools that can directly interact with APIs to access data from a number of databases including `rfishbase` for FishBase (Boettiger et al., 2012), and `taxize` for inter alia the Encyclopedia of Life, iPlant/Taxosaurus (Chamberlain and Szöcs, 2013) among others. The `neotoma` package was developed in conjunction with rOpenSci and represents one of two rOpenSci packages that interact with paleoecological data along with the `paleobioDB` package (Varela et al., 2014).

The `neotoma` package addresses concerns regarding data access and workflow reproducibility by providing users with tools that allow paleoecologists to query, download, organize, and summarize data from the Neotoma database using R. Here we describe the `neotoma` package, and illustrate its use with examples

drawn from the ecological literature, with the general objective of illustrating how neotoma provides tools to perform paleoecological research in an open and reproducible manner.

The neotoma package

The neotoma R package is an interface between the NeotomaPaleoecological Database (http://neotomadb.org) and statistical tools in R. neotoma uses an API to send data requests to the Neotoma Database, and then forms data objects that are compatible with existing R packages for environmental reconstruction, manipulation, and presentation of paleoecological data such as analogue (Simpson and Oksanen, 2014) and rioja (Juggins, 2013). The neotoma package also includes tools to standardize pollen taxon names across sample sites using a set of published pollen taxonomies for North America, or user defined taxonomies.

Data in the neotoma package is represented in three main classes (**Figure 1**): a "site", a "dataset" (grouped into a "dataset_list"), and a "download" (grouped into a "download_list"). A "site" is the most basic form of spatial information representing the spatial locations of datasets along with site names, descriptions and a unique site.id. A "site" is a "data.frame" with columns siteid, sitename, lat, long, elev, description, long_acc, and lat_acc. These column headings are generally self explanatory; long_acc and lat_acc are used to indicate the width of the bounding box for a sample site (with a midpoint of long and lat). In the Neotoma Database, examples of sites include a lake from which one or more cores are collected, a cave from which one or more faunal assemblages are collected, an archaeological dig with one or more excavation pits, and so forth. Each unique site returned by the get_site() search is placed in a row, providing enough descriptive data to plot locations and understand the spatial context of a site.

Although get_site() is useful for first-pass surveys of data availability, analysts more commonly will want to search for and retrieve datasets stored in the Neotoma Database. The function get_dataset() uses search terms almost identical to those used in get_site(), and it returns a more complete description of the datasets available. However, at this time only get_site() has the ability to search for sites based on site names.

A "dataset" associated with and individual site can be obtained using get_dataset(). Each search using get_dataset() returns a "dataset_list", a list of one or more "dataset"s, whose length is equal

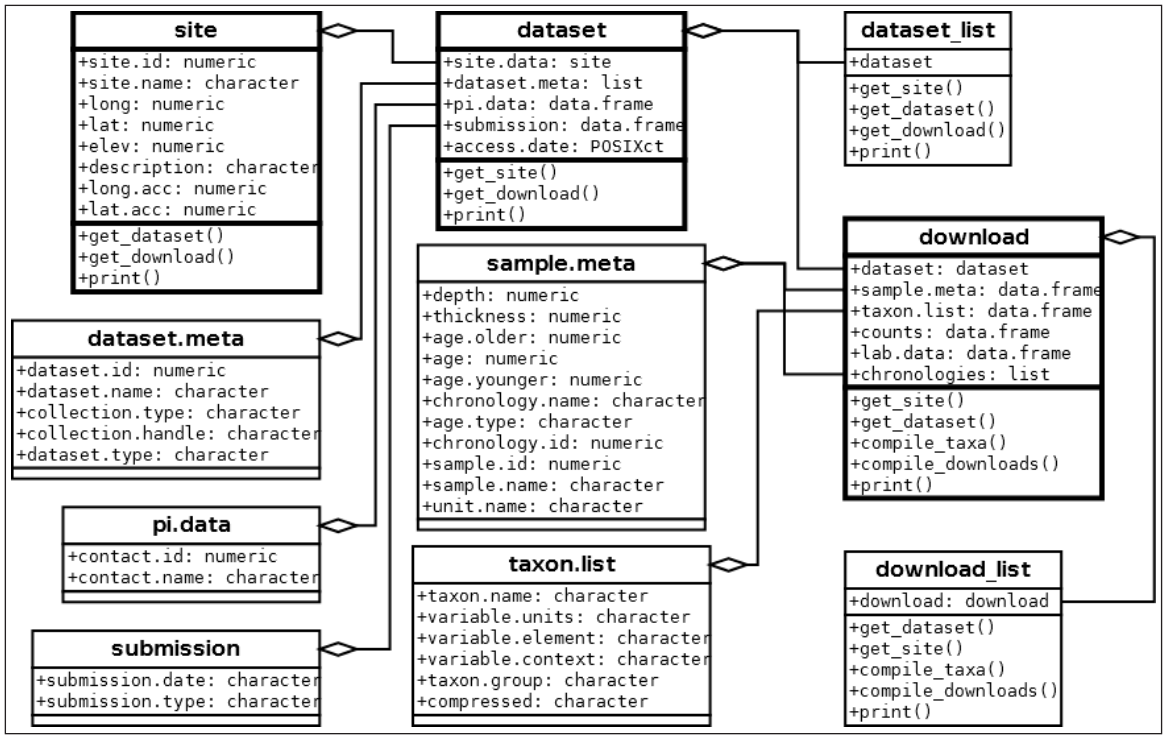

Figure 1: Major classes in neotoma, their relations to one another and the associated methods (functions). The classes described below have a heavier outline than their associated variables. The class download contains multiple lists, including dataset, which is a list and defined class. dataset itself contains lists, one of which is also a class, site. neotoma can interact directly with classes through the use of special methods for the various functions described here. These objects and classes are described in more detail below.

to the number of datasets returned by the function. `get_dataset()` returns the metadata associated with a dataset in the Neotoma Database. Neotoma's datasets are the containers for a set of samples of the same type from a single collection unit within a site. Examples of datasets in Neotoma include 1) all the pollen counts from a single core from a lake, 2) all the geochronological measurements (e..g radiocarbon dates) from a sediment core, 3) all the faunal data from an excavation in a cave, 4) all the plant macrofossil data from a packrat midden. In `neotoma`, a `"dataset"` includes the `"site"` for each `"dataset"` (as `"site.data"`, **Figure 1**), along with the `"dataset"` specific metadata (`"dataset.meta"` in **Figure 1**). The `dataset` also includes the principal investigator, submission date to Neotoma, and the date that the information was accessed via the Neotoma API using the R package. Each `"dataset_list"` and `"dataset"` can be used to access the full `"download"` using the `get_download()` method.

`get_download()` returns an object of class `"download_list"` containing one or more objects of class `"download"` (**Figure 1**). `get_download()` will also return a confirmation message for each individual API call as the function proceeds that can be turned off using the argument `verbose = FALSE`. Each `"download"` contains the associated `"dataset"` (**Figure 1**). The `"sample.meta"` component is where the core depth and age information is stored. The actual chronologies are stored in `"chronologies"`. If a core has a single age model then `"chronologies"` has a length of one. Some cores have multiple chronologies and these are also included in the `"chronologies"` list. The default chronology is stored in `"sample.meta"` and is always the first chronology in `"chronologies"`.

To build a new chronology with the same chronological controls as an existing chronology, perhaps using a different algorithim, `get_chroncontrol()` can be used to return the chronological controls and the `"chronology.id"` in either `"sample.meta"` or any one of the `"chronologies"` objects. While the chronological controls used to build a chronology may vary across chronologies for a single site, the default model contains the "best" chronological control data, as determined at the time the chronologies for the collection unit were last reviewed. It is important to note, however, that the "best" chronologies for most collection

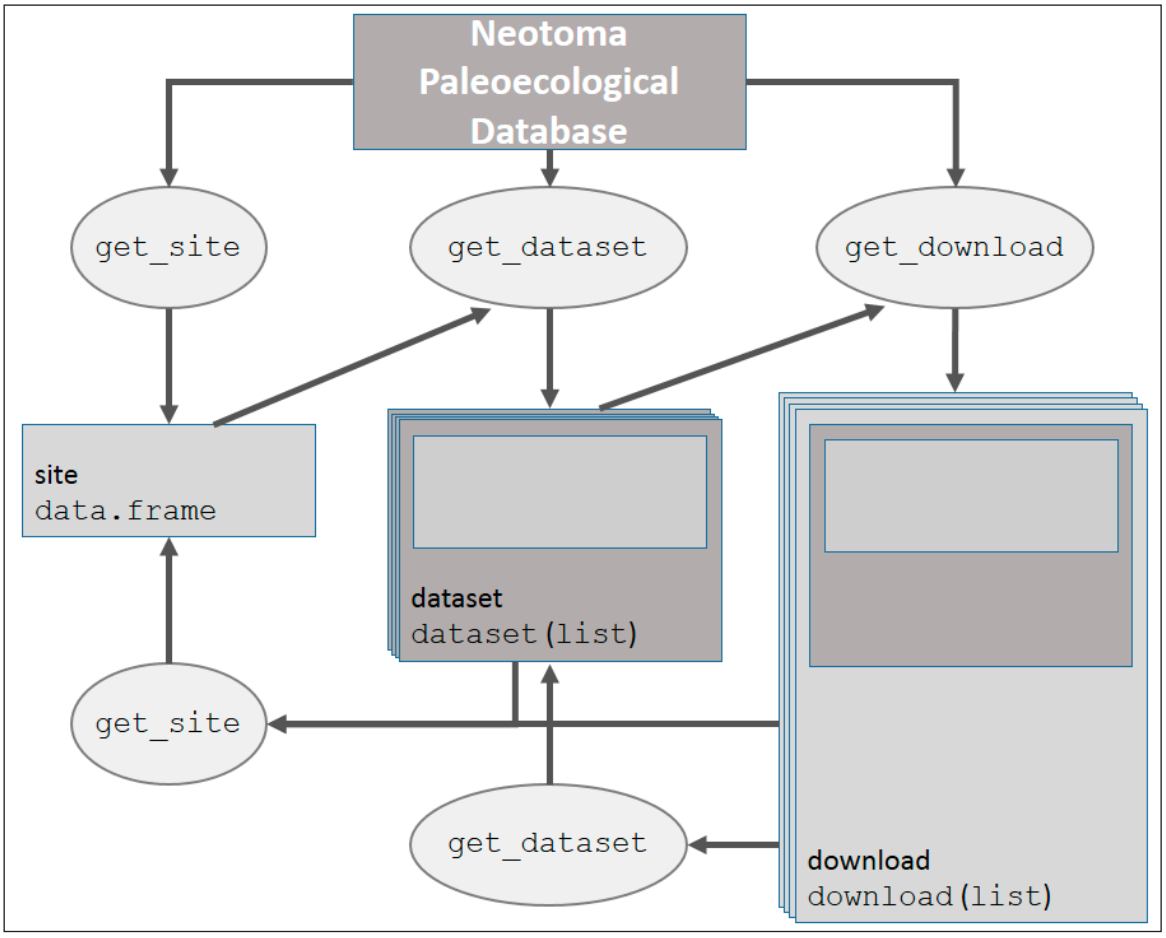

Figure 2: How the main data objects relate to one another in the `neotoma` package, and the helper functions used to move from one data type to another.

units in the database were based on "classical" age models (Blaauw, 2010) that do not include estimates of uncertainty. Moreover, these default age models that are in calibrated radiocarbon years utilize radiocarbon dates that are calibrated a priori. Bayesian age modeling programs, such as `Bacon` (Blaauw and Christen, 2011), as well as the classical age modeling program `clam` (Blaauw, 2010), and return estimates of uncertainty, using uncalibrated radiocarbon dates as input. Thus, the calibrated ages of many existing age models in Neotoma will not be appropriate for these programs, and the age controls may have to be obtained from the geochronology table using `get_geochron()`.

The age controls of existing default radiocarbon-year chronologies may sometimes be appropriate for programs such as `Bacon` and `clam`; however, many of these chronologies rejected radiocarbon dates *a priori*, which could be included when using `Bacon`. The `neotoma` package has a function to interface directly with `Bacon` or `clam`, called `write_agefile()`, which will output a correctly formatted age file for either of these applications using a `"download"` object.

The `"taxon.list"` component lists taxa found in the dataset samples (`"counts"`) and any laboratory data (`"lab.data"`), along with the units of measurement and taxonomic grouping (**Figure 1**). The `"counts"` are the actual count, presence or percentage data recorded for the dataset. The `"lab.data"` component contains information about any spike used to determine concentrations, sample quantities and, in some cases, charcoal counts.

Each of these objects, `"site"`, `"dataset"` and `"download"` can be obtained using direct calls to the API, or using functions defined in the `neotoma` package (**Figure 2**).

Examples

Here we present several examples that both introduce users to the `neotoma` package, and highlight how `neotoma` can be used in a paleoecological workflow. We begin with a simple example in which we compare change in *Alnus* pollen between two sites during the late Quaternary, followed by two more involved examples where we look at *Pinus* migration and changes in late-Quaterary mammal distributions in the United States.

A simple example

A researcher is interested in finding the pollen record for Marion Lake, in British Columbia (Mathewes, 1973), and comparing changes in *Alnus* pollen to the Louise Pond record (Pellatt and Mathewes, 1997) from HaidaG'Waii, further north. We search for specific sites by name using `get_site()`, making use of the wildcard `"%"` to catch sites whose site names begin with the strings `"Marion Lake"` or `"Louise Pond"`:

```
library("neotoma")
marion <- get_site(sitename = 'Marion Lake%')
  The API call was successful, you have returned  1 records.
louise <- get_site(sitename = 'Louise Pond%')
  The API call was successful, you have returned  1 records.
louise
              site.id      long      lat elev
  Louise Pond    1618 -131.7533 53.41667  650

description
  Louise Pond Glacial scour lake. Physiography: Queen Charlotte Ranges,
Louise Island. Surrounding vegetation: Tsuga mertensiana, Pinus contorta.
              long.acc lat.acc
  Louise Pond        0       0
```

In each case `get_site()` returns a single `"site"` (**Figures 1 & 2**). Here we queried the Neotoma database for site based on `sitename`, but alternately we could have queried for sites within a geographical bounding box, or by geopolitical region.

To get `"dataset"`s for these records we can bind the two records together by rows using `rbind()`, and pass the result directly to `get_dataset()` (**Figure 2**):

```
western.sites <- rbind(marion, louise)
western.data  <- get_dataset(western.sites)
```

`western.data` is a `"dataset_list"`, containing two `"dataset"`s (**Figure 1**). The `"dataset"` for a site will be nested within a `"dataset_list"`, even if only a single site is returned, so that methods can be consistent across classes and functions. This means that a single `"dataset"` must be retrieved as e.g., `western.data[[1]]` (this is also the case for `"download"` and `"download_list"` objects). The use of `"dataset"` and

"dataset_list" classes allow us to easily move between get_dataset(), get_site() and get_download(). We can see the special print() method for both "dataset"s and "dataset_list"s:

```
western.data
  A dataset_list containing 4 objects:
  Accessed from 2015-01-19 21:01h to 2015-01-19 21:01h.
  Datasets:
   dataset.id                          site.name      long      lat
         1705 Marion Lake (CA:British Columbia) -122.5472 49.30833
         8377 Marion Lake (CA:British Columbia) -122.5472 49.30833
         1670 Louise Pond                       -131.7533 53.41667
         8356 Louise Pond                       -131.7533 53.41667
             type
   pollen
   geochronologic
   pollen
   geochronologic
western.data[[1]]
  A dataset for Marion Lake (CA:British Columbia)
  Accessed 2015-01-19 21:01h.
   dataset.id                          site.name      long      lat   type
         1705 Marion Lake (CA:British Columbia) -122.5472 49.30833 pollen
```

before we download the full records and print them:

```
western.dl <- get_download(western.data)
  API call was successful. Returned record for Marion Lake(CA:British Columbia)
  API call was successful. Returned record for Marion Lake(CA:British Columbia)
  The dataset ID 8377 is associated with a geochronology object, not count data.
  API call was successful. Returned record for Louise Pond
  Warning:
  Modifiers are absent from the lab objects Lycopodium tablets, Sample quantity,
  Lycopodium spike.
  get_download will use uniqueidentifiers to resolve the problem.
  API call was successful. Returned record for Louise Pond
  The dataset ID 8356 is associated with a geochronology object, not count data.
western.dl
  A download_list containing 2 objects:
  Accessed from 2015-01-19 21:01h to 2015-01-19 21:01h.
  Datasets:
   dataset.id                          site.name      long      lat
         1705 Marion Lake (CA:British Columbia) -122.5472 49.30833
         1670 Louise Pond                       -131.7533 53.41667
   age.younger age.older   type
            58     13051 pollen
            62     10065 pollen
western.dl[[1]]
  A download object for Marion Lake (CA:British Columbia)
  Accessed 2015-01-19 21:01h.
   dataset.id                          site.name      long      lat age.young
         1705 Marion Lake (CA:British Columbia) -122.5472 49.30833        58
   age.old   type
     13051 pollen
```

It is worth noting that it is possible to go directly from a "site" object or set of "site"s to get_download() as well if desired.

Pollen taxonomy can vary substantially across cores depending on the level of taxonomic resolution used by a pollen analyst, or as a result of changes to taxonomies over time. One analyst might discriminate subgenera of *Pinus*, another might simply identify *Pinus* to the genus level. Gramineae—a common pollen type in earlier pollen records—has now been renamed Poaceae. This variable and shifting taxonomy is a first-order challenge for analysts seeking to analyze the dynamics of taxa across multiple groups. neotoma provides several options for standardized taxonomic list, corresponding to three published taxonomies for the United States and Canada (Gavin et al., 2003; Whitmore et al., 2005; Williams and Shuman, 2008). This function can be helpful, but should be used with care. The aggregation table is accessible using the command data(pollen.equiv) and the function to compile the data is called compile_taxa(). It can accomodate

either the internal translation table provided with the package, or a user-defined table with the same structure as `pollen.equiv`.

We are interested in comparing the relative pollen abundances of *Alnus* between two sites. We compile the pollen data using the `'P25'` taxonomy from Gavin et al. (2003). The first record downloaded is Marion Lake. We can see the `"download"` for Marion Lake the `taxon.table` has 5 columns:

```
head(western.dl[[1]]$taxon.list)
```

taxon.name	variable.units	variable.element	variable.context	taxon.group
Asteraceaeundiff.	NISP	pollen	NA	Vascular plants
Betula	NISP	pollen	NA	Vascular plants
Pinusmonticola-type	NISP	pollen	NA	Vascular plants
Myrica	NISP	pollen	NA	Vascular plants
Poaceae	NISP	pollen	NA	Vascular plants
Alnus	NISP	pollen	NA	Vascular plants

Once we apply `compile_taxa()` to the dataset using the `'P25'` compiler:

```
western.comp <- compile_taxa(western.dl, list.name = 'P25')
names(western.comp) <- c("marion", "louise")
```

The `taxon.table` for Marion Lake now has an extra column (note that several columns were removed to improve readability).

```
head(western.comp[[1]]$taxon.list[,c(1, 5, 6)])
```

	taxon.name	taxon.group	compressed
2	Asteraceaeundiff.	Vascular plants	Prairie Forbs
29	Betula	Vascular plants	Betula
3	Pinusmonticola-type	Vascular plants	Pinus
4	Myrica	Vascular plants	Other
5	Poaceae	Vascular plants	Poaceae
6	Alnus	Vascular plants	Alnus

`compile_taxa()` returns a `"download_list"` or `"download"`, for which `taxon.list` gains a column named `compressed` to link the original taxonomy to the revised taxonomy. This linkage is an important reference for researchers who choose to use this package for large-scale analysis, but who might need to later check the aggregated taxonomic groups against the original data. In this example we see that all the spore types have been lumped into a single taxon *Other*.

The sample data (`"counts"`) contained in each `"download"` in the `"download_list"western.dl` are converted into percentages using `tran()` from the `analogue` package (Simpson, 2007). We can then compare *Alnus* pollen percentages from these two locations to learn about vegetation changes on the west coast of North America during the Holocene:

```
library("analogue")

marion.alnus <- tran(x = western.comp$marion$counts, method = 'percent')[,'Alnus']
louise.alnus <- tran(x = western.comp$louise$counts, method = 'percent')[,'Alnus']

alnus.df <- data.frame(alnus = c(marion.alnus, louise.alnus),
                       ages  = c(western.comp$marion$sample.meta$age,
                                 western.comp$louise$sample.meta$age),
                       site = c(rep('Marion', length(marion.alnus)),
                                rep('Louise', length(louise.alnus))))

plot(alnus ~ ages, data = alnus.df, col = alnus.df$site, pch = 19,
     xlab = 'Years Before Present', ylab = 'Percent Alnus')
```

Marion Lake (red, **Figure 3**) maintains much higher proportions of *Alnus* pollen throughout its history, and has a rapid increase in *Alnus* pollen during the historical period. This rapid shift in the last 200 years is likely as a result of rapid colonization by pioneer *Alnus rubra* following forest clearance and fire in the lower mainland of British Columbia (Mathewes, 1973).

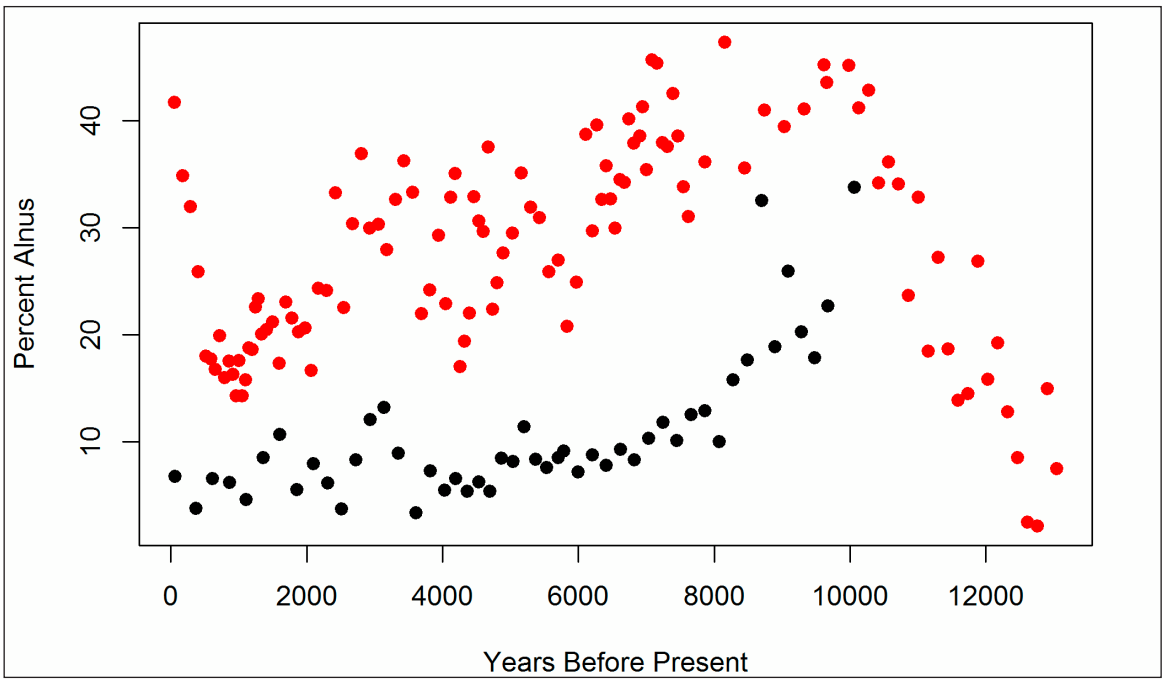

Figure 3: Plots of *Alnus* pollen percentages at two sites, one in the lower mainland of British Columbia (Marion Lake, red) and the other on HaidaG'waii (Louise Pond, black).

It is also possible to plot the pollen stratigraphy (**Figure 4**) at any one site using the `Stratiplot()` function in the `analogue` package for R (Simpson, 2007). Here we plot pollen percentages for Marion Lake, eliminating all taxa with maximum abundances less than 4% using `chooseTaxa()`:

```
core.pct <- data.frame(tran(western.comp[[1]]$counts, method = "percent"))
core.pct$age <- western.comp[[1]]$sample.meta$age

core.pct <- chooseTaxa(core.pct, max.abun = 4)
Stratiplot(age ~ ., core.pct, sort = 'wa', type = 'poly',
           ylab = "Years Before Present")
```

Pinus migration following the last Glacial Maximum

Macdonald and Cwynar (1991) used *Pinus* pollen percentages to map the northward migration of lodgepole pine (*Pinus contorta* var *latifolia*) following the retreat of the Laurentide Ice Sheet and the accompanying rise of temperatures in the late-Glacial and early Holocene. In their study a cutoff of 15% *Pinus* pollen was defined as the indicator of *Pinus* presence. Strong and Hills (2013) have remapped the migration front using a lower pollen proportion (5%) and more sites. Here, the analysis is partially replicated. Note that additional R packages must be installed and loaded for the following examples.

A spatial bounding box delimiting sites is defined for the search. Strong and Hills (2013) use a region approximately bounded by 54°N and 65°N, and from 110°W to 130°W. The function `get_site()` can return all sites within a slightly expanded bounding box:

```
#install.packages('ggmap', 'ggplot2', 'reshape2', 'Bchron', 'gridExtra')
library("ggmap")
library("ggplot2")
library("reshape2")
library("Bchron")
library("gridExtra")

all.sites <- get_site(loc = c(-140, 45, -110, 65))
  The API call was successful, you have returned  461 records.
```

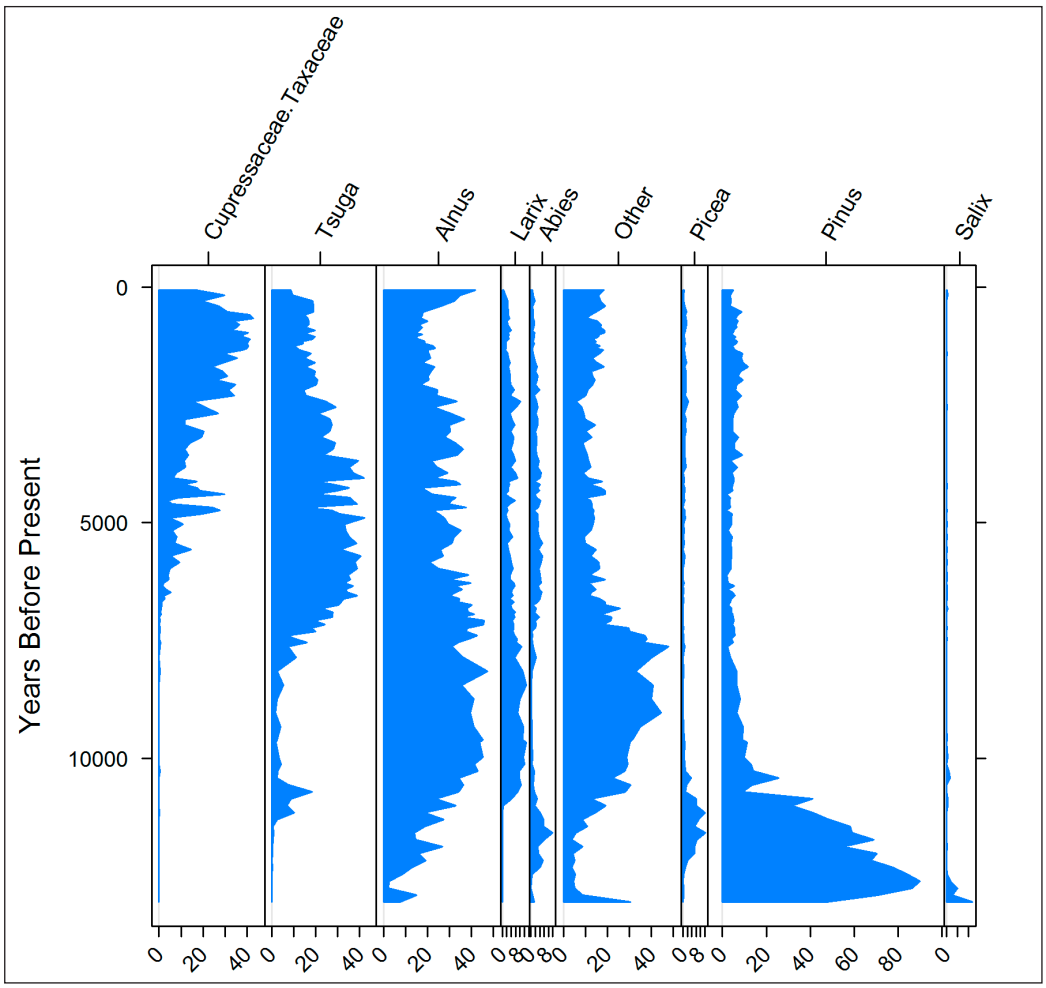

Figure 4: Stratigraphic plot for Marion Lake. Age is plotted on the y-axis in calibrated radiocarbon years before present. The `analogue` package provides the opportunity for users to further customize the stratigraphic plot if so desired.

The code above returned 461 sites. To narrow down the search we will use `get_dataset()` to search for all *Pinus* taxa within the same bounding box as above. `get_dataset()` can also limit the type of dataset, either by looking for specific taxa, or by describing the dataset type (e.g., `datasettype = 'pollen'` or `dataset-type = 'mammal'`). The `%` wildcard indicates that any characters may follow a string starting with `"Pinus"`:

```
all.datasets <- get_dataset(loc = c(-140, 45, -110, 65),
                            datasettype = 'pollen',
                            taxonname = 'Pinus%')
```

The API returned 74 datasets. Many dropped sites were pollen surface samples, or sites with datasets for other taxonomic groups. The distribution of the 74 fossil pollen sites can be plotted over the original 461 sites. We use the `ggplot2` package (Wickham, 2009) to make the figures:

```
map <- map_data('world')
ggplot(data = data.frame(map), aes(long, lat)) +
  geom_polygon(aes(group=group), color = 'steelblue', alpha = 0.2) +
  geom_point(data = all.sites, aes(x = long, y = lat)) +
  geom_point(data = get_site(all.datasets),
  aes(x = long, y = lat), color = 2) +
  xlab('Longitude West') +
  ylab('Latitude North') +
  coord_map(projection = 'albers', lat0 = 40, lat1 = 65,
            xlim = c(-140, -110), ylim = c(45, 70))
```

The map (**Figure 5**) shows a number of sites in the interior of British Columbia that have no fossil pollen. There are also other sites not shown here that may have relevant data but have yet to be entered into the database. This highlights a common challenge in paleoecoinformatics—the import of individual records into data repositories takes some time and is an on-going process that is aided by the collective contributions of the original analysts, data stewards, and large-scale research initiatives (e.g. PAGES 2K, PalEON). Fortunately, new software tools are speeding up the process of uploading and vetting data. The Tilia software (http://www.neotomadb.org/data/category/tilia) has been updated to allow direct upload to the Neotoma Database and includes a large number of automated data quality checks and standardized look-up tables for variable names. Because `neotoma` directly links to the Neotoma Database via APIs, analyses using `neotoma` can be updated continuously as new sites are added.

To obtain the data for the 74 sites we use `get_download()`:

```
all.downloads <- get_download(all.datasets, verbose = FALSE)
```

Only the percentage of *Pinus* is of interest, so we can again compile the taxa across the `"download_list"` using the 'P25' taxonomy (Gavin et al., 2003):

```
compiled.cores <- compile_taxa(all.downloads, 'P25')
```

We want to determine which sample has the first local *Pinus* presence in each core using a cutoff of 5% (Strong and Hills, 2013). We can find which rows in the *Pinus* column in each `"download"'s"count"` data frame have presence over 5% and then find the highest row number since the samples in a dataset are ordered stratigraphically, with the youngest sample in the top row and the oldest sample in the bottom row. This is a more complicated example. We first have to write a function that can take pollen count data

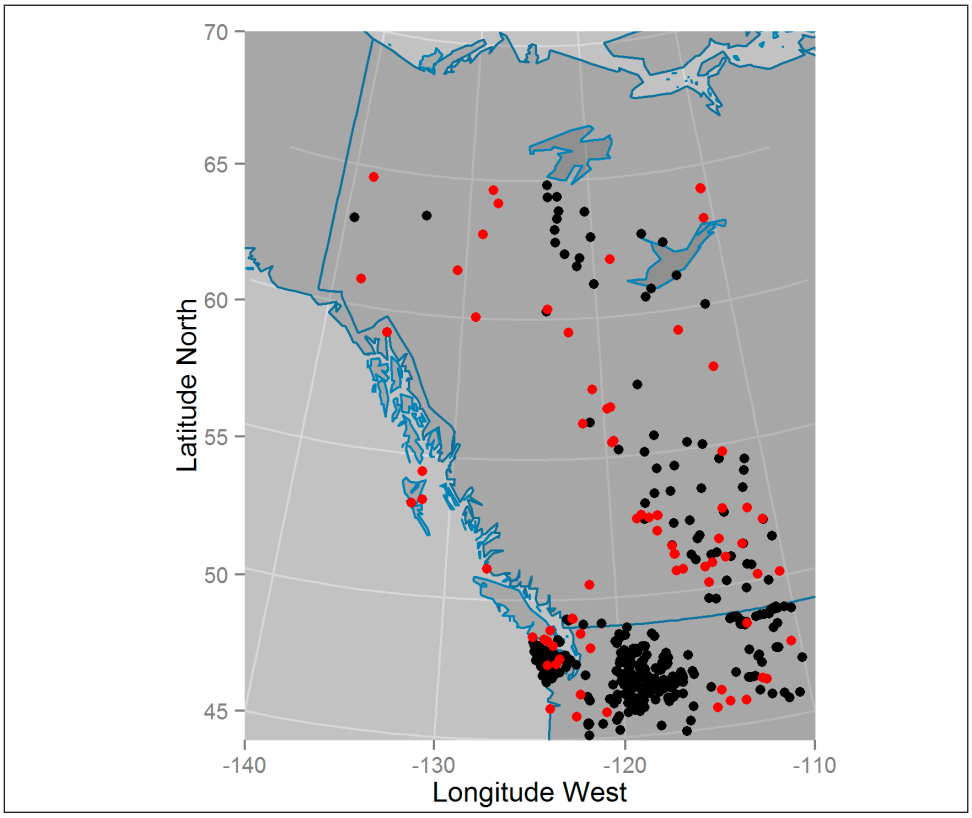

Figure 5: Mapped sites with pollen cores in British Columbia, Alberta and the Yukon Territory of Canada (red), including other Neotoma sites without stratigraphic pollen data (black).

```
top.pinus <- function(x) {
  x.pct <- tran(x$counts, method = "proportion")
  #  Cores must span at least the last 5000 years (and have no missing dates):
  old.enough <- max(x$sample.meta$age) > 5000 & !all(is.na(x$sample.meta$age))
  #  Find the highest row index associated with Pinus presence over 5%
  oldest.row <- ifelse(any(x.pct[, 'Pinus'] > .05 & old.enough),
                       max(which(x.pct[, 'Pinus'] > .05)),
                       0)
  #  return a data.frame with site name & location, and the age and date type
  #  (since some records have ages in radiocarbon years) for the oldest Pinus.
  out <- if (oldest.row > 0) {
      data.frame(site = x$dataset$site.data$site.name,
                 lat = x$dataset$site.data$lat,
                 long = x$dataset$site.data$long,
                 age = x$sample.meta$age[oldest.row],
                 date = x$sample.meta$age.type[oldest.row])
  } else {
      NULL
  }
  out
}
#  Apply the function 'top.pinus' to each core using lapply and rbind:
summary.pinus <- do.call("rbind.data.frame", lapply(compiled.cores, top.pinus))
```

We need to calibrate dates that are recorded in radiocarbon years using the `Bchron` package (Parnell, 2014). In most cases the original uncertainty for individual ages is not recorded in "`sample.meta`" so we assume a 100 year standard deviation. Note that direct recalibration of radiocarbon dates from interpolated age models is not the best approach, but is sufficient for our purposes here:

```
radio.years <- summary.pinus$date %in% 'Radiocarbon years BP'
sryears <- sum(radio.years, na.rm = TRUE)
# BChronCalibrate is in the BChron package:
calibrated <- BchronCalibrate(summary.pinus$age[radio.years],
                              ageSds = rep(100, sryears),
                              calCurves = rep('intcal13', sryears))

#  we want the weighted means from 'calibrated'
wmean.date <- function(x) sum(x$ageGrid*x$densities / sum(x$densities))

summary.pinus$age[radio.years] <- sapply(calibrated, wmean.date)
summary.pinus <- na.omit(summary.pinus)
summary.pinus <- subset(summary.pinus, subset=!(age < 2000 & long < -130))

#  A loess curve is straightforward, but not the best model:
regress <- ggplot(summary.pinus, aes(x = lat, y = age)) +
                 geom_point(aes(color = age), size = 2) +
                 scale_y_reverse(expand = c(0, 100)) +
                 xlab('Latitude North') +
                 ylab('Years Before Present') +
                 geom_smooth(n = 40, method = 'loess') +
                 geom_rect(aes(xmin = 59, xmax = 60, ymin = 7000, ymax = 10000),
                 color = 2, fill = 'blue', alpha = 0.01)

mapped <- ggplot(data = data.frame(map), aes(long, lat)) +
                geom_polygon(aes(group = group), color = 'steelblue', alpha = 0.2) +
                geom_point(data = summary.pinus,
                aes(x = long, y = lat, colour = age), size = 3) +
                coord_map(projection = 'albers', lat0 = 40, lat1 = 65,
                xlim = c(-140, -110), ylim = c(40, 70)) +
                theme(legend.position = 'none')

grid.arrange(mapped, regress, nrow=1)
```

The results show a clear pattern of northward expansion for *Pinus* in northwestern North America (**Figure 6**). These results broadly agree with the findings of Strong and Hills (2013) who suggest that *Pinus* reached a northern extent between 59°N and 60°N at approximately 10–7 kacal BP as a result of geographic barriers before continuing northward after 7 kacal BP.

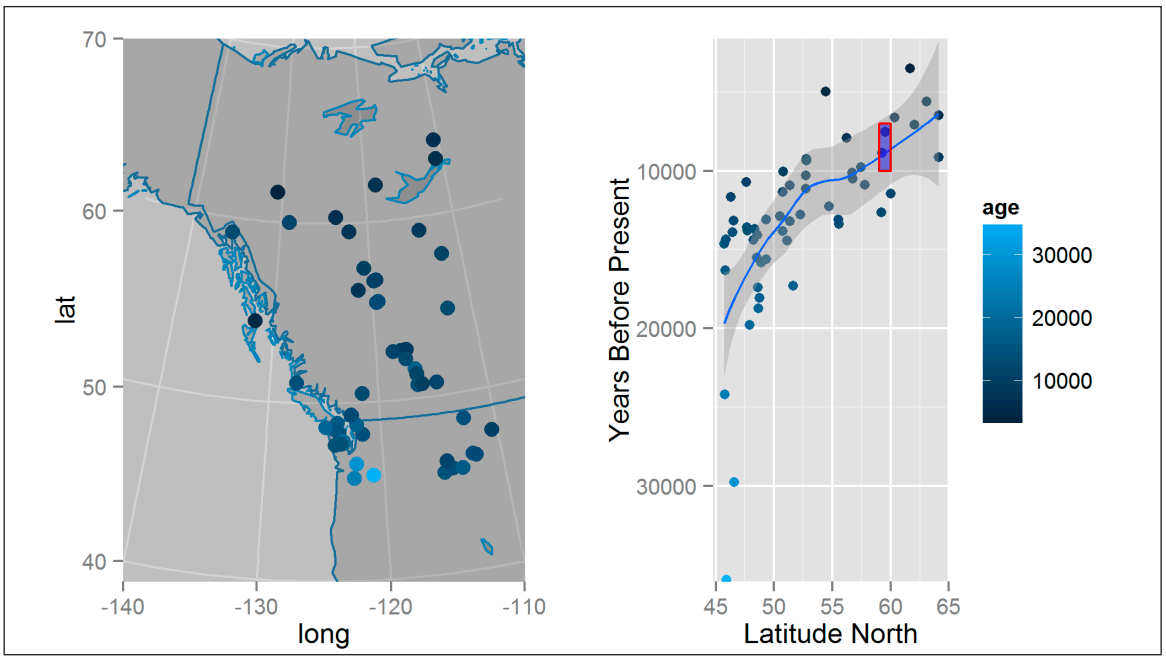

Figure 6: Mapped ages of first *Pinus* establishment in the interior of British Columbia and the Yukon Territory based on a 5% pollen cut-off. The age of first appearance is also plotted and smoothed with a loess curve.

Mammal Distributions in the Pleistocene

Graham et al. (1996) built and applied the FAUNMAP dataset (http://www.ucmp.berkeley.edu/faunmap) of fossil assemblages to elucidate patterns of change in mammal distributions through the Pleistocene to the present. The paper uses various multivariate analyses to show, in part, that mammal species have responded in a Gleasonian manner to climate change since the late-Pleistocene. Graham et al. (1996) show some species migrating northward in response to warming climates, others staying relatively stable, and some moving southward. FAUNMAP has been incorporated into Neotoma (and expanded with new records), and this example performs some simple analyses that show how different species responded to the changing climate of the last deglaciation.

First, all vertebrate fauna datasets are obtained from Neotoma:

```
#  Bounding box is effectively the continental USA, excluding Alaska
mam.set <- get_dataset(datasettype= 'vertebrate fauna',
                       loc = c(-125, 24, -66, 49.5))
#  Retrieving this many sites can be very time consuming
mam.dl <- get_download(mam.set)
```

Sites are assigned to time-period bins as in Graham et al. (1996); Modern (< 0.5 kacal BP), Late Holocene (0.5–4 kacal BP), Early-Mid Holocene (4--10 kacal BP), Late Glacial (10--15 kacal BP), Full Glacial (15–20 kacal BP) and Late Pleistocene (> 20 kacal BP). The first step is to build a large table with `time` and `xy` coordinates for each site. Time data in "`sample.meta`" for the mammal data is not the same as for for pollen where age is commonly found in the `sample.meta$age` component. Most vertebrate fauna samples are assigned only younger (`sample.meta$age.younger`) and older (`sample.meta$age.older`, **Figure 1**) bounds with no estimates of mean or median age. In this example we average the younger and older ages to determine sample age. We recognize that this averaging of ages is likely to be methodologically indefensible in the scientific literature, but is sufficient for illustrative purposes here. Here we make use of the `reshape2` package (Wickham, 2007) to help manipulate data structures.

```
compiled.mam <- compile_downloads(mam.dl)
time.bins <- c(500, 4000, 10000, 15000, 20000)
mean.age <- rowMeans(compiled.mam[,c('age.old', 'age.young', 'age')],
                     na.rm = TRUE)
```

```
interval <- findInterval(mean.age, time.bins)
periods <- c('Modern',
             'Late Holocene',
             'Early-Mid Holocene',
             'Late Glacial',
             'Full Glacial',
             'Late Pleistocene')
compiled.mam$ageInterval <- periods[interval + 1]
mam.melt <- melt(compiled.mam,
                 measure.vars = 10:(ncol(compiled.mam)-1),
                 na.rm = TRUE,
                 factorsAsStrings = TRUE)
mam.melt <- transform(mam.melt, ageInterval = factor(ageInterval, levels = periods))
mam.lat <- dcast(data = mam.melt, variable ~ ageInterval, value.var = 'lat',
                 fun.aggregate = mean, drop = TRUE)[,c(1, 3, 5, 6)]
#  We only want taxa that appear at all time periods:
mam.lat <- mam.lat[rowSums(is.na(mam.lat)) == 0, ]
#  Group the samples based on the range & direction (N vs S) of migration.
#  A shift of only 1 degree is considered stationary.
mam.lat$grouping <- factor(findInterval(mam.lat[,2] - mam.lat[, 4],
                                        c(-11, -1, 1, 20)),
                           labels = c('Southward', 'Stationary', 'Northward'))
mam.lat.melt <- melt(mam.lat)
colnames(mam.lat.melt)[2:3] <- c('cluster', 'Era')
```

The output of the above manipulations is displayed via a `ggplot2` plot using

```
ggplot(mam.lat.melt, aes(x = Era, y = value)) +
  geom_path(aes(group = variable, color = cluster)) +
  facet_wrap(~ cluster) +
  scale_x_discrete(expand = c(.1,0)) +
  ylab('Mean Latitude of Occurrance') +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

Even with this fairly simple set of analyses we see that species did not respond uniformly to climatic warming following deglaciation (**Figure 7**), consistent with the prior work of Graham *et al.* (1996). Although most range shifts were northward, a number of taxa show little change in their ranges and a number show southward range shifts (**Figure 7**). This example does not examine east-west movement and ignores the issues that may be associated with the complex topography of the mountainous west, or possible confounding effects introduced by temporal variations in the available set of sites. The broader point here is that the use of `neotoma` can support research that is synchronized with the data holdings of large repositories such as Neotoma and reproducible.

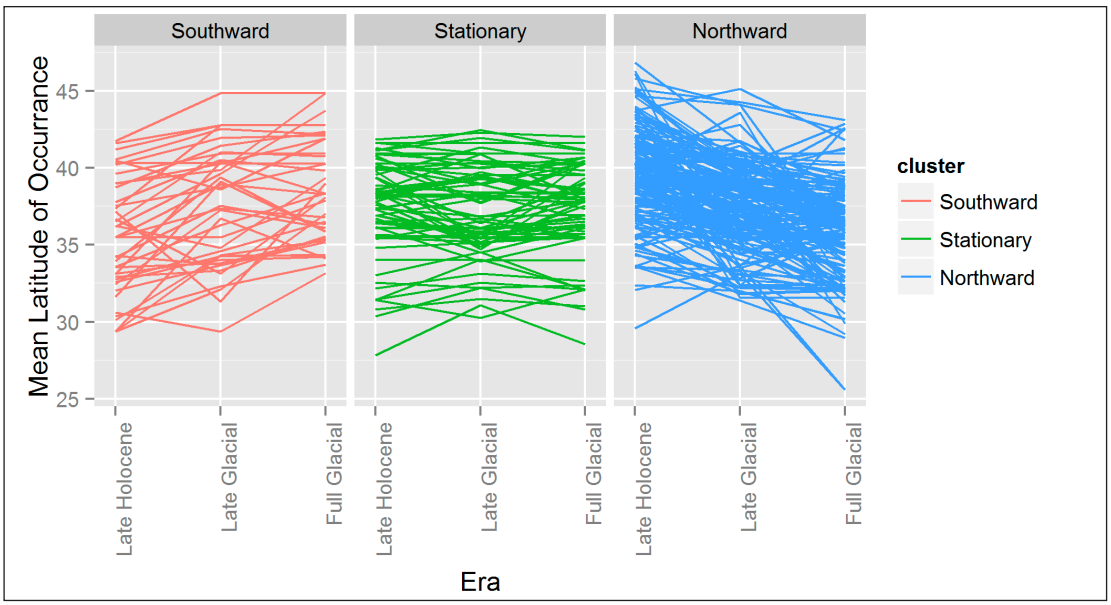

Figure 7: Mean latitudinal distribution of fossil mammal taxa during the LatePleistocene show that while there appears to be a net northward migratory pattern, a number of taxa appear not to shift their ranges, or move southward following deglaciation.

Conclusion

The whole of the fossil record is much greater than the sum of its parts. Many of our discipline's most important advances were made possible only by the synthesis of many individual fossil occurrences into regional- to global-scale databases of species occurrences, e.g., the Neotoma Paleoecology Database and the Paleobiological Database. Current frontiers in paleoecological informatics include 1) facilitating the input of data into these databases, 2) improved sophistication of the data models employed by these databases, enabling them to handle increasingly complex arrays of paleobiological and associated geochronological data, and 3) enabling the frictionless integration of these resources with other cyberinfrastructure (Brewer et al., 2012; Committee, 2014; Uhen et al., 2013).

Here we present the `neotoma` package for R and show how it can be used to directly transfer data from the Neotoma Paleoecology Database into the R statistical computing environment. The broader goals of this effort are 1) to ease the transfer of data from Neotoma into an environment widely used for paleoecological analyses (Juggins, 2013; Simpson, 2007; Simpson and Oksanen, 2014) and 2) to enable transparent and reproducible scientific workflows. The `neotoma` package itself is available either from the CRAN repository, or from GitHub (http://github.com/ropensci/neotoma) where ongoing open-source development continues. Suggestions for improvement and new code contributions by readers and users are welcome.

Acknowledgements

We would like to acknowledge the support of the rOpenSci project and the invaluable efforts made by data contributors across the globe who have provided the platform upon which Neotoma and the `neotoma` package are built. Brian Bills and Michael Anderson at Penn State University's Center for Environmental Informatics of provided considerable support in helping us to understand and use the API. This paper is a product of the PalEON project (http://paleonproject.net), funded through the NSF-Macrosystems program grants #1065656, 1241868 and 1241874, and the Neotoma Paleoecology Database, funded by the NSF Geoinformatics program grants #0947459 and #0948652. GLS was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant Program (RGPIN 2014-04032). We would also like to thank Lee Hsiang Liow, Pasquale Raia and one anonymous reviewer for their helpful comments. The authors would also like to thank the staff at Ubiquity Press for their considerable efforts in working with the authors to provide a clean layout for the code presented in this paper.

References

Alroy, J., Aberhan, M., Bottjer, D.J., Foote, M., Fürsich, F.T., Harries, P.J., Hendy, A.J., Holland, S.M., Ivany, L.C., Kiessling, W., Kosnik, M.A., Marshall, C.R., McGowan, A.J., Miller, A.I., Olszewski, T.D., Patzkowsky, M.E., Peters, S.E., Villier, L., Wagner, P.J., Bonuso, N., Borkow, P.S., Brenneis, B., Clapham, M.E., Fall, L.M., Ferguson, C.A., Hanson, V.L., Krug, A.Z., Layou, K.M., Leckey, E.H., Nürnberg, S., Powers, C.M., Sessa, J.A., Simpson, C., Tomašových, A., Visaggi, C.C., 2008. Phanerozoic trends in the global diversity of marine invertebrates. *Science* 321, 97–100. DOI: http://dx.doi.org/10.1126/science.1156963

Barnosky, A.D., Matzke, N., Tomiya, S., Wogan, G.O., Swartz, B., Quental, T.B., Marshall, C., McGuire, J.L., Lindsey, E.L., Maguire, K.C., Mersey, B., Ferrer, E.A., 2011. Has the earth's sixth mass extinction already arrived? *Nature* 471, 51–57. DOI: http://dx.doi.org/10.1038/nature09678

Barnosky, C.W., 1989. Postglacial vegetation and climate in the northwestern Great Plains of Montana. *Quaternary Research* 31, 57–73. DOI: http://dx.doi.org/10.1016/0033-5894(89)90085-9

Behrensmeyer, A.K., Miller, J.H., 2012. Building links between ecology and paleontology using taphonomic studies of recent vertebrate communities, in: Paleontology in Ecology and Conservation. Springer, pp. 69–91. DOI: http://dx.doi.org/10.1007/978-3-642-25038-5_5

Blaauw, M., 2010. Methods and code for "Classical" age-modelling of radiocarbon sequences. *Quaternary Geochronology* 5, 512–518. DOI:http://dx.doi.org/10.1016/j.quageo.2010.01.002

Blaauw, M., Christen, J.A., 2011. Flexible paleoclimate age-depth models using an autoregressive gamma process. *Bayesian Analysis* 6, 457–474. DOI: http://dx.doi.org/10.1214/ba/1339616472

Blarquez, O., Carcaillet, C., Frejaville, T., Bergeron, Y., 2014. Disentangling the trajectories of alpha, beta and gamma plant diversity of North American boreal ecoregions since 15,500 years. *Frontiers In Paleoecology* 2, 6.

Blarquez, O., Vannière, B., Marlon, J.R., Daniau, A.-L., Power, M.J., Brewer, S., Bartlein, P.J., 2014. paleofire: an R package to analyse sedimentary charcoal records from the Global Charcoal Database to reconstruct past biomass burning. *Computers & Geosciences* 72, 255–261. DOI: http://dx.doi.org/10.1016/j.cageo.2014.07.020

Blois, J.L., Williams, J.W., Fitzpatrick, M.C., Ferrier, S., Veloz, S.D., He, F., Liu, Z., Manion, G., Otto-Bliesner, B., 2013. Modeling the climatic drivers of spatial patterns in vegetation composition since the Last Glacial Maximum. *Ecography* 36, 460–473. DOI: http://dx.doi.org/10.1111/j.1600-0587.2012.07852.x

Boettiger, C., Lang, D., Wainwright, P., 2012. rfishbase: exploring, manipulating and visualizing FishBase data from R. *Journal of Fish Biology* 81, 2030–2039. DOI: http://dx.doi.org/10.1111/j.1095-8649.2012.03464.x

Brewer, S., Jackson, S.T., Williams, J.W., 2012. Paleoecoinformatics: applying geohistorical data to ecological questions. *Trends in Ecology & Evolution* 27, 104–112. DOI: http://dx.doi.org/10.1016/j.tree.2011.09.009

Chamberlain, S.A., Szöcs, E., 2013. taxize: taxonomic search and retrieval in R. *F1000Research* 2.

C4P Executive Committee, 2014. Paleobiology workshop report, in: EarthCube Collaboration and Cyberinfrastructure for Paleogeosciences (C4P). National Science Foundation EarthCube Research Coordination Network: Cyberinfrastructure for Paleogeoscience.

Davis, M.B., 1981. Quaternary history and the stability of forest communities, in: Forest Succession. Springer, pp. 132–153. DOI: http://dx.doi.org/10.1007/978-1-4612-5950-3_10

Dietl, G.P., Flessa, K.W., 2011. Conservation paleobiology: putting the dead to work. *Trends in Ecology & Evolution* 26, 30–37. DOI: http://dx.doi.org/10.1016/j.tree.2010.09.010

Fritz, S.A., Schnitzler, J., Eronen, J.T., Hof, C., Böhning-Gaese, K., Graham, C.H., 2013. Diversity in time and space: wanted dead and alive. *Trends in Ecology & Evolution* 28, 509–516. DOI: http://dx.doi.org/10.1016/j.tree.2013.05.004

Gavin, D.G., Oswald, W.W., Wahl, E.R., Williams, J.W., 2003. A statistical approach to evaluating distance metrics and analog assignments for pollen records. *Quaternary Research* 60, 356–367. DOI: http://dx.doi.org/10.1016/S0033-5894(03)00088-7

Goring, S., Lacourse, T., Pellatt, M.G., Mathewes, R.W., 2013. Pollen assemblage richness does not reflect regional plant species richness: a cautionary tale. *Journal of Ecology* 101, 1137–1145. DOI: http://dx.doi.org/10.1111/1365-2745.12135

Goring, S., Williams, J., Blois, J., Jackson, S., Paciorek, C., Booth, R., Marlon, J., Blaauw, M., Christen, J., 2012. Deposition times in the northeastern United States during the Holocene: establishing valid priors for Bayesian age models. *Quaternary Science Reviews* 48, 54–60. DOI: http://dx.doi.org/10.1016/j.quascirev.2012.05.019

Graham, R.W., LundeliusJr, E.L., Graham, M.A., Schroeder, E.K., Toomey III, R.S., Anderson, E., Barnosky, A.D., Burns, J.A., Churcher, C.S., Grayson, D.K., others, 1996. Spatial response of mammals to late Quaternary environmental fluctuations. *Science* 272, 1601–1606. DOI: http://dx.doi.org/10.1126/science.272.5268.1601

Grimm, E., Bradshaw, R., Brewer, S., Flantua, S., Giesecke, T., Lézine, A., Takahara, H., Williams, J.W., 2013. Databases and their application, in: Elias, S. and Mock, C. J. (Eds.), Encyclopaedia of Quaternary Science. Elsevier, pp. 831–838.

Grimm, E.C., 2008. Neotoma: An ecosystem database for the Pliocene, Pleistocene, and Holocene, *Illinois State Museum Scientific Papers E Series 1*.

Huntley, B.J., Webb, T., 1988. Vegetation History. Kluwer Academic Publishers Dordrecht. DOI: http://dx.doi.org/10.1007/978-94-009-3081-0

Juggins, S., 2013. rioja: Analysis of Quaternary science data, *R package version 0.8-5*.

MacDonald, G., Cwynar, L.C., 1991. Post-glacial population growth rates of *Pinus contorta* ssp. *latifolia* in western Canada. *The Journal of Ecology* 417–429. DOI: http://dx.doi.org/10.2307/2260723

Marlon, J.R., Bartlein, P.J., Daniau, A.-L., Harrison, S.P., Maezumi, S.Y., Power, M.J., Tinner, W., Vanniére, B., 2013. Global biomass burning: a synthesis and review of Holocene paleofire records and their controls. *Quaternary Science Reviews* 65, 5–25. DOI: http://dx.doi.org/10.1016/j.quascirev.2012.11.029

Mathewes, R.W., 1973. A palynological study of postglacial vegetation changes in the University Research Forest, southwestern British Columbia. *Canadian Journal of Botany* 51, 2085–2103. DOI: http://dx.doi.org/10.1139/b73-271

Parnell, A., 2014.Bchron: Radiocarbon dating, age-depth modelling, relative sea level rate estimation, and non-parametric phase modelling, *R package version 4.1.1*.

Pellatt, M.G., Mathewes, R.W., 1997. Holocene tree line and climate change on the Queen Charlotte Islands, Canada. *Quaternary Research* 48, 88–99. DOI: http://dx.doi.org/10.1006/qres.1997.1903

R Core Team, 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Raup, D.M., Sepkoski, J.J., 1984. Periodicity of extinctions in the geologic past. *Proceedings of the National Academy of Sciences* 81, 801–805. DOI: http://dx.doi.org/10.1073/pnas.81.3.801

Reichman, O., Jones, M.B., Schildhauer, M.P., 2011. Challenges and opportunities of open data in ecology. *Science* 331, 703–705. DOI: http://dx.doi.org/10.1126/science.1197962

Sepkoski, J.J., 1997. Biodiversity: past, present, and future. Journal of Paleontology 533–539.

Simpson, G.L., 2007. Analogue methods in palaeoecology: Using the analogue package. *Journal of Statistical Software* 22, 1–29.

Simpson, G.L., Oksanen, J., 2014. analogue: Analogue and weighted averaging methods for palaeoecology, *R package version* 0.14-0.

Strong, W.L., Hills, L.V., 2013. Holocene migration of lodgepole pine (*pinus contorta* var. *latifolia*) in southern Yukon, Canada. *The Holocene* 23, 1340–1349. DOI: http://dx.doi.org/10.1177/0959683613484614

Tzedakis, P., 1994. Vegetation change through glacial-interglacial cycles: a long pollen sequence perspective. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 345, 403–432. DOI: http://dx.doi.org/10.1098/rstb.1994.0118

Uhen, M.D., Barnosky, A.D., Bills, B., Blois, J., Carrano, M.T., Carrasco, M.A., Erickson, G.M., Eronen, J.T., Fortelius, M., Graham, R.W., Grimm, E.C., O'Leary, M.A., Mast, A., Piel, W.H., Pollyn, P.D. and **Säiläg, L.K.,** 2013. From card catalogs to computers: databases in vertebrate paleontology. *Journal of Vertebrate Paleontology* 33, 13–28. DOI: http://dx.doi.org/10.1080/02724634.2012.716114

Varela, S., Gonzalez-Hernandez, J., Sgarbi, L.F., 2014.paleobioDB: an R package for downloading, visualizing and processing data from the Paleobiology Database, *R package version 0.1*.

Whitmore, J., Gajewski, K., Sawada, M., Williams, J., Shuman, B., Bartlein, P., Minckley, T., Viau, A., Webb III, T., Shafer, S., Anderson, P., and **Brubaker, L.,** 2005.Modern pollen data from North America and Greenland for multi-scale paleoenvironmental applications. *Quaternary Science Reviews* 24, 1828–1848. DOI: http://dx.doi.org/10.1016/j.quascirev.2005.03.005

Wickham, H., 2007. Reshaping data with the reshape package. *Journal of Statistical Software* 21, 1–20.

Wickham, H., 2009. ggplot2: elegant graphics for data analysis. Springer.

Williams, J., Shuman, B., 2008. Obtaining accurate and precise environmental reconstructions from the modern analog technique and North American surface pollen dataset. *Quaternary Science Reviews* 27, 669–687. DOI: http://dx.doi.org/10.1016/j.quascirev.2008.01.004

Williams, J.W., Shuman, B.N., Webb III, T., Bartlein, P.J., Leduc, P.L., 2004. Late-Quaternary vegetation dynamics in North America: Scaling from taxa to biomes. *Ecological Monographs* 74, 309–334. DOI: http://dx.doi.org/10.1890/02-4045

Wolkovich, E.M., Regetz, J., O'Connor, M.I., 2012. Advances in global change research require open science by individual researchers. *Global Change Biology* 18, 2102–2110. DOI: http://dx.doi.org/10.1111/j.1365-2486.2012.02693.x

How to cite this article: Goring, S, Dawson, A, Simpson, G L, Ram, K, Graham, R W, Grimm, E C and Williams, J W 2015 neotoma: A Programmatic Interface to the Neotoma Paleoecological Database. *Open Quaternary,* 1: 2, pp. 17-17, DOI: http://dx.doi.org/10.5334/oq.ab

Published: 09 March 2015

Copyright: © 2015 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License (CC-BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/3.0/.

]u[*Open Quaternary* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS