**Article Info**

**Keywords:**
- Biometry
- Log-ratios
- Osteometric references
- R package

**Abstract**

Log Size Indexes (LSI) allow the increase of the number of data and have been used in a number of zooarchaeological studies since 1950. However, some standards to calculate the log ratios remain unpublished, and it is further hindered by the diversity of data recording practices. The R package ‘zoolog’ enables calculation of thousands of log-ratios in seconds, with the advantage that the users can choose between different public references, which increases the repeatability and comparability of the results, allowing the smooth integration of references and databases with heterogeneous nomenclatures. Alternatively, the users may use their own references. This paper presents the main functionalities and procedures enabled by the package ‘zoolog’, together with some examples of use. A real dataset and several examples with R code are provided with the aim of facilitating osteometrical analyses in zooarchaeology.

1. Log ratios in zooarchaeology

In general, zooarchaeological datasets are composed of skeletal remains representing many different anatomical body parts from different species. In investigation of animal size, the analysis of measurements from a given anatomical element provides the best control for the variables affecting size and shape and, as such, it is the preferable option. Log size indexes (LSI) are used to compare the relative (rather than the absolute) dimensions of animals from archaeological contexts (Simpson, 1941; Meadow, 1999; Wolfhagen, 2020). LSI methods compare archaeological measurements to a standard, producing a value that indicates how much larger or smaller the archaeological specimen is compared to that standard (Simpson, 1941; Meadow, 1981). Similar to other scaling techniques (e.g. Ducos, 1968; Uerpmann, 1979) the use of LSIs allows for measurements from different anatomical elements to be combined, and therefore to be plotted and analysed together. This increases the size of osteometrical datasets and therefore – potentially – their statistical robustness (Albarella, 2002). However, length, width and depth measurements of the anatomical elements are better analysed separately and should not be directly aggregated (Davis, 1996; Wolfhagen, 2020).

LSI methods have been used in a number of works to explore the effects of domestication, insularity and market production among other themes (e.g. Davis, 1996; Vigne et al., 2000; Albarella, 2002; Duval et al., 2013; Groot and Deschler-Erb, 2015; Trentacoste et al., 2018; Kunst and Gál, 2021, to cite just but a few). Unfortunately, some of the standards cannot be easily found in the publications, which limits the comparability of the results with those obtained by other researchers, as well as the reproducibility of the results. In addition, the calculation of log ratios may be long and tedious, and the use of a programming language is most convenient when dealing with large datasets.

We developed zoolog with the aim of facilitating osteometrical analysis in zooarchaeology and enhancing the comparability of results between different researchers using different standards. We made a particular effort to build a robust thesaurus, capable of reading datasets in different languages and from different scientific traditions. This paper presents the main functions available in the package and some examples of use, with the hope that it will be useful for researchers analysing animal size around the world.
2. Why zoolog?

R times is a versatile tool for statistical analyses (R Core Team, 2020). Most times zooarchaeologists face small bone assemblages with few osteometrical data, but sometimes large datasets need to be processed, most notably when regional syntheses are intended. zoolog allows the user to choose between published standards or their own, although the use of the references available in the package is encouraged to facilitate the comparability of the results. The core spirit of zoolog is to provide a useful tool for researchers developing osteometrical analyses (no matter the language or scientific tradition) and to facilitate repeatability and comparison between researchers.

The package is available from CRAN, and it can be installed directly on the console:

```r
install.packages("zoolog")
```

In addition, a development version is available from GitHub:

Multiple sets of standard reference values are included in the package. These standards include published and widely used biometric datasets (e.g. Davis, 1996; Albarella and Payne, 2005) as well as other less known standards. These references, as well as the data example provided with the package, are based on the measurements and abbreviations defined in Von den Driesch (1976) and Davis (1996). Please note that not for the sheep standard from Davis, the original nomenclature was used for SD (Davis, 1996), which is the equivalent to DD in Von den Driesch (1976). Researchers strictly following Von den Driesch nomenclature consider this to avoid too small log ratio values when using Davis sheep reference (Evin et al., in this issue).

The package includes a zoolog thesaurus to facilitate its usage by research teams across the globe, and working in different languages and with different recording traditions. The thesaurus enables the zoolog package to recognise many different names for taxa and skeletal elements (e.g. “Bos taurus,” “Bos,” “BT,” “bovino,” “bota,” or, “Humerus,” “HU,” “Húmer”). Consequently, there is no need to use a particular standardized recording code for the names of different taxa or elements. Users can expand the thesaurus themselves, or contact the zoolog maintainers (Pozo and Valenzuela-Lamas) to make the expanded nomenclature available for other researchers.

The package also includes a zoolog taxonomy to facilitate the management of cases recorded as identified at different taxonomic ranks (Species, Genus, Tribe) and their match with the corresponding references.

3. Standards available in zoolog (version 1.0)

Currently, zoolog includes reference values for the main domesticates in Europe and their agriotypes (Bos, Ovis, Capra, Sus), red deer (Cervus elaphus), Persian fallow deer (Dama mesopotamica), mountain gazelle (Gazella gazella), donkey (Equus asinus), horse (Equus caballus), and European rabbit (Oryctolagus cuniculus). Dogs were not initially included due to the huge size variability and allometry between morphotypes, but a wolf standard will be shortly available in the next release, and further genera (e.g. Bison, Camelus) will be included in the near future. References are drawn from a variety of publications and resources, and correspond to either single individuals or mean data from a given population.

The references’ database is organized per taxon. Table 1 details the standards currently available in zoolog and some of the main characteristics (e.g. single individual, mean from a population, geographical origin, male/female, etc.). All the references available can be seen in the R console once loaded the zoolog package:

```r
library(zoolog)
referenceSets
```

In addition, a detailed explanation is available in zoolog’s references Database help page accessible from the R console:

```r
help(referencesDatabase)
```

Given that zooarchaeological data generally include several taxa, predefined groups of references have been set to facilitate the calculation of log ratios. By default, the set combination is ‘combi’, which includes the references containing the largest number of standard measurements, and all the taxa for which a reference is available. In addition, the user can make their own combination of references (see below), and the provided references can be extended and updated integrating newer research data. Table 2 summarises the four current predetermined reference sets.

The measurements considered in each set are easily accessible from the console and also from the Environment tag in R Studio (under package:zoolog subtag). Here are two examples for the R console:

4. Main procedures enabled by zoolog (version 1.0)

zoolog includes several functions to organise and standardize databases, choose and create reference sets, expand the thesauri and calculate log ratios. The description of all the functions is available in the zoolog documentation. We detail here the main procedures that can be of interest for researchers aiming to explore zoolog with their own data. The examples here use the dataset provided with zoolog (see R script in supp. mat. 1), originally in Catalan and with special characters (Valenzuela-Lamas, 2008), as an example of how several languages are transparently handled with the package. A typical code to help users with their data is also provided (supp. mat. 1).

4.1. Database format

The first step is reading a data frame (typically a.csv or.xlsx file) containing the data. The data is supposed to be formatted with each individual bone fragment in a row. It must include at least two columns identifying taxon and bone element, respectively. In addition, each measurement (e.g. Gl, Bd, Dp) is expected to be recorded in a different specific column, identified by the measure name as column name. Any number of additional columns can also be included in the data frame. But columns with repeated names will originate warnings and eventually lead to errors. The dataset provided in zoolog represents an example of this expected format. We can read it by:

```r
dataFile <- system.file("extdata", "dataValenzuelaLamas2008.csv.gz", package="zoolog")
data <- read.csv2(dataFile, na.strings = "", encoding = "UTF-8")
```

install.packages("devtools")
develtools::install_github(\"josepzojo/zoolog@HEAD\", build_vignettes = TRUE)
4.2. Making zoolog understand a database: thesauri and taxonomy

4.2.1. Thesauri

As previously stated, zoolog is aimed for zooarchaeologists working in different languages and with different recording traditions. This introduced the challenge of integrating different nomenclatures for many different variables of databases and references, making them compatible without requiring their standardization. To enable this, zoolog includes a set of thesauri for four types of variables: taxa, elements, measures, and identifiers. In particular, this allows different options for taxa and elements to be directly recognized in different databases.

For instance, the current version of zoolog will understand as 'Bos taurus' all the following options in the user database, no matter if they are consistent or mixed through the database:

- Bos taurus
- Bos tauri
- Bovine
- Buffaloes
- Bos

Similarly, the options included in 'element' thesaurus can be easily seen from the R console:

Table 1
List of the references available in zoolog version 1.0

<table>
<thead>
<tr>
<th>Taxon</th>
<th>zoolog name</th>
<th>Main features</th>
<th>N measurements</th>
<th>Bibliographical reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bos taurus</td>
<td>Nieto</td>
<td>Bronze Age female individual. Catalonia (NE Spain)</td>
<td>68</td>
<td>Nieto-Espinet (2018)</td>
</tr>
<tr>
<td>Bos taurus</td>
<td>Johnstone</td>
<td>Mean cattle measurements from Late Iron Age - Roman period, Heybridge (Southern England)</td>
<td>24</td>
<td>Johnstone and Albarella (2002)</td>
</tr>
<tr>
<td>Bos primigenius</td>
<td>Degerbol</td>
<td>Female aurochs (non standard measures converted to Driesch, 1976)</td>
<td>50</td>
<td>Degerbol and Frejdskild (1970)</td>
</tr>
<tr>
<td>Bos primigenius</td>
<td>Steppan</td>
<td>Female aurochs with standard measures (Driesch, 1976)</td>
<td>84</td>
<td>Steppan (2001)</td>
</tr>
<tr>
<td>Ovis aries</td>
<td>Davis</td>
<td>Mean adult female Shetland sheep measurements from a single flock, Orkney islands, Scotland</td>
<td>23</td>
<td>Davis (1996)</td>
</tr>
<tr>
<td>Ovis aries</td>
<td>Clutton</td>
<td>Mean adult male Soay sheep measurements from a single flock, Hirta, Scotland</td>
<td>71</td>
<td>Clutton-Brock et al. (1990)</td>
</tr>
<tr>
<td>Ovis mosiorn</td>
<td>Basel</td>
<td>Present-day male individual, Switzerland</td>
<td>36</td>
<td>Stopp and Deschler-Erb (2018)</td>
</tr>
<tr>
<td>Ovis orientalis</td>
<td>Uerpmann</td>
<td>Present-day female sheep, Western Iran</td>
<td>50</td>
<td>Uerpmann and Uerpmann (1994)</td>
</tr>
<tr>
<td>Capra hirca</td>
<td>Basel</td>
<td>Present-day adult male, Switzerland</td>
<td>35</td>
<td>Stopp and Deschler-Erb (2018)</td>
</tr>
<tr>
<td>Capra hirca</td>
<td>Clutton</td>
<td>Mean measurements feral goats unknown age and sex, Holy Island, Scotland</td>
<td>59</td>
<td>Clutton-Brock et al. (1990)</td>
</tr>
<tr>
<td>Capra aegagrus</td>
<td>Uerpmann</td>
<td>Mean measurements from a male and a female present-day goats. Southern Turkey</td>
<td>50</td>
<td>Uerpmann and Uerpmann (1994)</td>
</tr>
<tr>
<td>Sus domesticus</td>
<td>Albarella</td>
<td>Mean measurements Late Neolithic pigs from Durrington Walls (Southern England)</td>
<td>42</td>
<td>Albarella and Payne (2005)</td>
</tr>
<tr>
<td>Sus scrofa</td>
<td>Basel</td>
<td>Present-day male individual, Switzerland</td>
<td>41</td>
<td>Stopp and Deschler-Erb (2018)</td>
</tr>
<tr>
<td>Sus scrofa</td>
<td>Hongo</td>
<td>Present-day female wild boar. Eastern Turkey (Elazig)</td>
<td>95</td>
<td>Hongo and Meadow (2000)</td>
</tr>
<tr>
<td>Sus scrofa</td>
<td>Payne</td>
<td>Mean measurements from a male and a female present-day wild boars (Sus scrofa libycus). Northern Turkey</td>
<td>33</td>
<td>Payne and Bull (1988)</td>
</tr>
<tr>
<td>Cervus elaphus</td>
<td>Basel</td>
<td>Present-day male adult individual, Switzerland</td>
<td>14</td>
<td>Stopp and Deschler-Erb (2018)</td>
</tr>
<tr>
<td>Dama mesopotamica</td>
<td>Haifa</td>
<td>Present-day adult female, Israel</td>
<td>60</td>
<td>Harding and Marom (2021)</td>
</tr>
<tr>
<td>Gazella gazella</td>
<td>Haifa</td>
<td>Present-day adult female, Israel</td>
<td>63</td>
<td>Harding and Marom (2021)</td>
</tr>
<tr>
<td>Equus ainsa</td>
<td>Haifa</td>
<td>Present-day adult male, Israel</td>
<td>48</td>
<td>Harding and Marom (2021)</td>
</tr>
<tr>
<td>Oryctolagus cuniculus</td>
<td>Nottingham</td>
<td>Present-day adult male. Essex (Southern England)</td>
<td>58</td>
<td>Ameen (2021)</td>
</tr>
</tbody>
</table>

Table 2
Sets of standards and their names in zoolog version 1.0

<table>
<thead>
<tr>
<th>Taxon</th>
<th>zoolog name</th>
<th>Main features</th>
<th>N measurements</th>
<th>Bibliographical reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bos taurus</td>
<td>Nieto</td>
<td>Bronze Age female individual. Catalonia (NE Spain)</td>
<td>68</td>
<td>Nieto-Espinet (2018)</td>
</tr>
<tr>
<td>Bos taurus</td>
<td>Johnstone</td>
<td>Mean cattle measurements from Late Iron Age - Roman period, Heybridge (Southern England)</td>
<td>24</td>
<td>Johnstone and Albarella (2002)</td>
</tr>
<tr>
<td>Bos primigenius</td>
<td>Degerbol</td>
<td>Female aurochs (non standard measures converted to Driesch, 1976)</td>
<td>50</td>
<td>Degerbol and Frejdskild (1970)</td>
</tr>
<tr>
<td>Bos primigenius</td>
<td>Steppan</td>
<td>Female aurochs with standard measures (Driesch, 1976)</td>
<td>84</td>
<td>Steppan (2001)</td>
</tr>
<tr>
<td>Ovis aries</td>
<td>Davis</td>
<td>Mean adult female Shetland sheep measurements from a single flock, Orkney islands, Scotland</td>
<td>23</td>
<td>Davis (1996)</td>
</tr>
<tr>
<td>Ovis aries</td>
<td>Clutton</td>
<td>Mean adult male Soay sheep measurements from a single flock, Hirta, Scotland</td>
<td>71</td>
<td>Clutton-Brock et al. (1990)</td>
</tr>
<tr>
<td>Ovis mosiorn</td>
<td>Basel</td>
<td>Present-day male individual, Switzerland</td>
<td>36</td>
<td>Stopp and Deschler-Erb (2018)</td>
</tr>
<tr>
<td>Ovis orientalis</td>
<td>Uerpmann</td>
<td>Present-day female sheep, Western Iran</td>
<td>50</td>
<td>Uerpmann and Uerpmann (1994)</td>
</tr>
<tr>
<td>Capra hirca</td>
<td>Basel</td>
<td>Present-day adult male, Switzerland</td>
<td>35</td>
<td>Stopp and Deschler-Erb (2018)</td>
</tr>
<tr>
<td>Capra hirca</td>
<td>Clutton</td>
<td>Mean measurements feral goats unknown age and sex, Holy Island, Scotland</td>
<td>59</td>
<td>Clutton-Brock et al. (1990)</td>
</tr>
<tr>
<td>Capra aegagrus</td>
<td>Uerpmann</td>
<td>Mean measurements from a male and a female present-day goats. Southern Turkey</td>
<td>50</td>
<td>Uerpmann and Uerpmann (1994)</td>
</tr>
<tr>
<td>Sus domesticus</td>
<td>Albarella</td>
<td>Mean measurements Late Neolithic pigs from Durrington Walls (Southern England)</td>
<td>42</td>
<td>Albarella and Payne (2005)</td>
</tr>
<tr>
<td>Sus scrofa</td>
<td>Basel</td>
<td>Present-day male individual, Switzerland</td>
<td>41</td>
<td>Stopp and Deschler-Erb (2018)</td>
</tr>
<tr>
<td>Sus scrofa</td>
<td>Hongo</td>
<td>Present-day female wild boar. Eastern Turkey (Elazig)</td>
<td>95</td>
<td>Hongo and Meadow (2000)</td>
</tr>
<tr>
<td>Sus scrofa</td>
<td>Payne</td>
<td>Mean measurements from a male and a female present-day wild boars (Sus scrofa libycus). Northern Turkey</td>
<td>33</td>
<td>Payne and Bull (1988)</td>
</tr>
<tr>
<td>Cervus elaphus</td>
<td>Basel</td>
<td>Present-day male adult individual, Switzerland</td>
<td>14</td>
<td>Stopp and Deschler-Erb (2018)</td>
</tr>
<tr>
<td>Dama mesopotamica</td>
<td>Haifa</td>
<td>Present-day adult female, Israel</td>
<td>60</td>
<td>Harding and Marom (2021)</td>
</tr>
<tr>
<td>Gazella gazella</td>
<td>Haifa</td>
<td>Present-day adult female, Israel</td>
<td>63</td>
<td>Harding and Marom (2021)</td>
</tr>
<tr>
<td>Equus ainsa</td>
<td>Haifa</td>
<td>Present-day adult male, Israel</td>
<td>48</td>
<td>Harding and Marom (2021)</td>
</tr>
<tr>
<td>Oryctolagus cuniculus</td>
<td>Nottingham</td>
<td>Present-day adult male. Essex (Southern England)</td>
<td>58</td>
<td>Ameen (2021)</td>
</tr>
</tbody>
</table>
In this case, zoolog will understand as ‘humerus’ all the following options in the user database (again, no matter if the name is consistent or mixed through the database): humerus, HU, hum, húmer, omero, Membre-antérieur humerus, Húmero. Both thesauri are case and punctuation insensitive, so that ‘Húmero’ will be considered the same as ‘húmero’ or ‘humero’.

The automatic identification of heterogeneous nomenclature is especially important for the computation of log ratios (function LogRatios), which needs to identify corresponding categories between reference and database. But it also facilitates other operations such as selecting subsets of the database by a category without requiring standardization (function InCategory). In addition, if desired, the database can be automatically standardized (function StandardizeDataSet).

4.2.2. Expanding the thesauri

It is probable that, despite our efforts, users will need to expand the options available in the thesauri. This can be done with the function AddToThesaurus as follows (see also R script, supp. mat. 1, example 1):

```r
elementThesaurusExtended <- AddToThesaurus(
  thesaurus = zoologThesaurusSetElement,
  newName = c("Schulterblatt", "Oberschenkelnknochen"),
  category = c("scapula", "femur")
)
```

This will add ‘Schulterblatt’ as an option for ‘scapula’, and ‘Oberschenkelnknochen’ as an option for ‘femur’ so that zoolog can recognize these options in the user’s database. For this extended thesaurus to be used by all functions by default, the zoologThesaurus could be overwritten:

```r
zoologThesaurusSetElement <- elementThesaurusExtended
```

For it to be saved for following sessions, the user can write the extended thesaurus set:

```r
WriteThesaurusSet(thesaurus = zoologThesaurus, 
  file = "myThesaurus")
```

And in the future it could be read when needed:

```r
zoologThesaurus <- ReadThesaurusSet("myThesaurus")
```

These operations will allow the user to privately expand their thesauri. However, users are encouraged to contact us so others can benefit from expanded thesauri.

4.2.3. Taxonomy

Sometimes, the available reference for a particular genus is of different species or subspecies from the one recorded in the data to analyze. For instance, the data could include cases of Bos taurus, but the reference is for Bos primigenius. In other cases, some data could not be completely identified and include cases recorded by a higher taxonomic rank. For instance, a doubtful case between sheep and goat could be recorded as Caprini.

In order to facilitate the management of cases recorded at different taxonomic ranks (species, genus, tribe, …) and coordinate their match with the corresponding references, a taxonomy for the most typical zooarchaeological taxa was introduced in the zoolog major release 1.0.0. The taxonomic hierarchy is structured up to the family level and can be easily seen from the R console with the command:

```r
zoologTaxonomy
```

The taxonomy is integrated in the function LogRatios, enabling it to automatically match measurements and references for different species of the same genus. For instance, data of Bos taurus can be matched with reference of Bos primigenius, since both are Bos. It also enables the function LogRatios to detect when a specimen has been recorded at a higher taxonomic rank, such as tribe or family, and to suggest to the user the set of possible reference species.

Additionally, it is complemented with a series of functions (Subtaxonomy, SubtaxonomySet, GetSpeciesIn) enabling the user to query for the subtaxonomy or the set of species under a queried taxon at any taxonomic rank.

4.3. Calculating log ratios in seconds

zoolog is able to calculate thousands of log ratios in seconds with the function LogRatios. Base-10 LSI calculations are enforced in zoolog’s current version (1.0) to facilitate the comparison between studies (i.e. base-e or base-2 are not currently enabled). By default, the most comprehensive set (reference$Comb) is used, but the user can always set their desired one. For the example, we use the dataset provided with zoolog and read above (see also R script, supp. mat. 1, example 2):

```r
dataLogs <- LogRatios(data)
```

The output is a data frame with the same structure as the input data, but with added columns at the right of the database collecting the computed log ratios (LSI). The new columns have the same name as the processed measures but prepended with ‘log’. For instance, the log ratio computed from the measure GL is given in the new column logGL. The calculated log ratios can be then saved if desired in a csv file for later analysis:

```r
write.csv2(dataLogs, "LogRatios.csv")
```

Here, the function LogRatios will report two warnings when applied to the dataset provided with zoolog.

Both of them reflect found relationships between taxonomic categories, which can be detected thanks to the taxonomy included in the package. The user can include their own taxonomy if desired.

The first warning informs the user that the data includes some cases of recorded taxa that did not match any species in the reference, but was matched by genus. In many cases this can be the behaviour expected by the user. For instance, using the reference of Sus scrofa for computing the log ratios of cases of Sus domesticus, can be acceptable if no better reference is available or if both (sub)species are being compared. The warning also includes cases recorded by genus, a recording practice that is understandable when working only with one species in the genus, such as Oryctolagus or Ovis above. However, the warning also informs the user that this matching by genus can be suppressed by setting the argument useGenusIfUnambiguous = FALSE. This can be the case if the user wants to exclude from the analysis the cases with not matching species. Besides, this warning could make the user realize that the wrong reference was set by mistake.

The second warning informs the user that the data includes some cases recorded with a taxon not specifying a species or genus, but at a higher taxonomic rank (for instance undecided Ovis/Capra, equivalent to tribe Caprini). It also informs of the cases recorded by the genus but for which the reference includes more than one species (as happens with Equus above). In addition, the user is reminded of the option to use the argument joinCategories to indicate which reference species should be used for them.

4.3.1. Manually joining categories for calculating log ratios

The function LogRatios includes the option joinCategories allowing several categories of taxa or anatomical elements to be considered together with the same reference one. For instance, the user could
Warning messages:
1: Reference for Sus scrofa used for cases of Sus domesticus.
   Reference for Sus scrofa used for cases of Sus.
   Reference for Oryctolagus cuniculus used for cases of Oryctolagus.
   Reference for Ovis aries used for cases of Ovis.
Set useGenusIfUnambiguous to FALSE if this behaviour is not desired.
2: Data includes some cases recorded as
   * Equus (which is a Genus)
     for which the reference for Equus asinus or Equus caballus could be used.
   * Caprini (which is a Tribe)
     for which the reference for Ovis aries or Capra hircus could be used.
Set jointCategories as appropriate if you want to use any of them.

indicate that log ratios of cases recorded with taxa Ovis, Capra, and
unknown Ovis/Capra should be all computed with the reference for
sheep (Ovis aries). This can be obtained with.

Conveniently, here the set of taxa under Ovis/Capra, equivalent to
tribe Caprini (thanks to the thesaurus), is obtained (according to the
taxonomy) using the function SubtaxonomySet. The result is equivalent
to manually setting.

This joining of categories can be a deliberate analysis decision, in
order to directly compare the size of those taxa, or a suboptimal
option when there is no available reference for one of them. Note that
without using joinCategories any species or genus not part of the selected
reference set will be excluded. For instance, reference$NietoDavisAlbarella
does not include a reference for Capra. Thus, if using this reference, log
ratios for goats will not be calculated unless joinCategories is set as above.

Analogously, joinCategories can be used with anatomical elements. A
typical example can be the presence of first and second phalanges
without anterior-posterior identification (“phal 1” and “phal 2”), which
the user could be interested to include by computing their log ratios with
respect to the reference of the corresponding anterior phalanges (“phal 1
ant” and “phal 2 ant”, respectively):

```r
categoriesPhal1 <- c("phal 1 ant", "phal 1")
categoriesPhal2 <- c("phal 2 ant", "phal 2")
dataExampleWithLogs4 <- LogRatios(
data,  
  joinCategories = list('phal 1 ant' = categoriesPhal1,  
                         'phal 2 ant' = categoriesPhal2)  
)
```

4.3.2. Using other references

It is easy to use another reference set if the user prefers it, just naming
it:

```r
dataLogs2 <- LogRatios(data, ref = reference$Basel)
```

It is also possible to use a customized combination of references for
each taxon with the function AssembleReference:

```r
dataLogs3 <- LogRatios(data, ref = customizedReference)
```

Finally, it is also possible to calculate the log ratios with the user's
own reference (see R script, supp. mat. 1, example 3). Note that the
reference must be formatted as a data frame with four columns: Taxon,
Element, Measure (code Von den Driesch), and Standard (measurement
in millimetres). The names of the taxa, elements and measures must be
included in the corresponding thesaurus for zoolog to interpret them
properly. Typically, it would be stored as a csv or excel file (in this
element ‘MyReference.csv’), which can be loaded when required and
directly used to calculate log ratios:

```r
PersonalReference <- read.csv("MyReference.csv")
dataLogs4 <- LogRatios(data, ref = PersonalReference)
```

4.4. Selecting a single summary value for lengths, widths and depths

Once the log ratios have been calculated for different measurements, an
key decision is how to analyse and plot the information. An unfrag-
mented anatomical element can yield several measurements, and this is
compounded in the case of a whole skeleton or articulating limb. In
order to avoid the over-representation of the specimens with several
measurements and introducing correlated values, only one LSI value per
specimen should be included in the analyses, and one specimen from an
articulating group of bones. In addition, widths, lengths and depths
should be considered separately to assess multi-dimensional size change,
and with a proper multivariate statistical analysis or properly compens-
ating for multiple comparisons (Hochberg and Tamhane, 1987). Assembling
without structure several measurements per bone, either from the same
or from different axes, would lead to correlated values, which would
invalidate any statistical analysis, generally based on the
independence of the observations. In addition, there can be the tendency
of the largest group of measurements (generally widths) to mask the
effects from the other axes.

Two different methods have been considered for selecting a single
summary value per specimen from a set of measurements per axis
(length, width and depth): (a) averaging for each specimen all available
values from that axis or (b) selecting one available value according to a
given priority order. Without the help of a programming language,
selecting log values (either by prioritisation or averaging) can be very
long and tedious. zoolog offers the possibility of performing it in seconds
with the function CondenseLogs. By default, the ‘priority’ method is
used. A default order of priority is also given on the basis of the fre-
quency and reliability of measurements (see Davis, 1996 for a discus-
sion). In the current version (zoolog 1.0), the order of priority for lengths
is “GL”, “GLf”, “GLm”, “HTC”, for widths is “BT”, “Bd”, “Bp”, “SD”,

```r
caprineCategory <- SubtaxonomySet("Ovis/Capra")
JointDataWithLogs4 <- LogRatios(data,  
  joinCategories = list(ovar = caprineCategory))
```
caprineCategory <- c("Ovis aries", "Ovis orientalis", "Capra hircus", "Capra aegagrus", "Ovis", "Capra", "Caprini")

customizedReference <- AssembleReference(
  combination = list(cattle = "Nieto", sheep = "Davis", Goat = "Clutton",
                     pig = "Albarella", redDeer = "Bascal")
)

"Bfd", "Bfp", and for depths is "Dd", "DD", "BG", "Dp". Users can easily choose their priority order and the measurements considered as follows (see also R script, supp. mat. 1, example 4):

The user can also easily set the method to 'average' to calculate the average of the LSI values per axis:

dataWithSummary <- CondenseLogs(
  dataLogRatios,
  method = "average"
)

CondenseLogs will add new columns (by default, Length, Width, and Depth) to the right end of the data frame, so no information is lost in the process. Users can also explore the effect of calculating the average or the prioritized value, and of changing the set of measurements or their order of priority using the code above.

4.5. Standardising the database to English and Scientific terminology (optional)

zoolog is very flexible and allows different languages to be included in a database (as long as the different names for taxa and elements are included in the thesauri). Their automatic processing is transparent to the user and the results are presented without changing the original nomenclature in the user’s data. However, if desired, the dataset can be standardized to follow a consistent terminology. This could be useful, for instance, to create tables and plots for international publications. To facilitate this, zoolog includes the function StandardizeDataSet. It will standardize data column names, and taxon, element, and measure names to the standard ones defined in the thesaurus. Using the default zoolog thesaurus, they will effectively be translated to English and with scientific terms for taxa and anatomical elements. This is accomplished even if the same category is written under different forms through the database, as ‘HU’ and ‘hum’, or ‘cattle’ and ‘Bos’, which will be converted into the standard names ‘humerus’ or ‘Bos taurus’, respectively. The standardisation can be done at any stage, before or after calculating the log ratios (in this example, after calculating and prioritising the log ratios; see R script, supp. mat. 1, example 5):

dataStandardized <- StandardizeDataSet(dataWithSummary)

dataWithSummary <- CondenseLogs(
  dataLogRatios,
  grouping = list(Length = c("GL", "LI", "GLI", "GLm", "GLe", "HTC"),
                  Width = c("BT", "Bd", "Bp", "SD", "Bfd", "Bfp", "GLP", "LG", "SLC"),
                  Depth = c("Dd", "DD", "BG", "Dp"))
)

4.6. Lazy datasets (optional)

The number of measurements (and therefore columns) that exist in the literature can be overwhelming, and sometimes, in the rush of recording thousands of bone fragments and to avoid navigating within the database, we zooarchaeologists nest some rare measurements under other categories. Thus, the requirement for each measurement to be in a specific column can be relaxed, provided that no ambiguity could arise. This is the case for some mutually exclusive measurement. For instance, for most usual taxa, the measurement GL is only applicable to the astragalus, for which the measurement GL does not apply. Thus, in the example dataset both are recorded merged in the column GL (this is referred to as ‘lazy’ datasets in zoolog’s help pages and vignettes).

The optional argument mergedMeasures of the function LogRatios, was introduced to enable the processing of this type of ‘lazy’ datasets. For the example data, where GL1 and GLpe have been nested into the column for GL we can do (see also R script, supp. mat. 1, example 6):

GLVariants <- list(c("GL", "GL1", "GLpe"))
dataLogs5 <- LogRatios(data, mergedMeasures = GLVariants)

This option tells the function to match measurements in data and reference considering any of the GLVariants as equivalent. Observe that for any particular anatomical element they are supposed to be mutually exclusive and no check is done on that. If they were not, the resulting behaviour is unspecified.

5. Choosing an osteometrical reference. Does it change the results?

The published references available in zoolog correspond to single individuals and mean values of particular populations (Table 1). Combining measurements from different anatomical elements can be a problem when strong allometries exist between populations (e.g. the standard used has a different ‘shape’ - size relation between forelimbs and hindlimbs compared to the population we want to analyse, see Evin et al. in this issue). This introduces additional variability in the obtained LSI, which can reduce the ability of the data to reflect patterns and group differences. Bias can also be introduced for unbalanced samples in which most bone remains correspond to a particular anatomical element, which has more problematic effects on the analysis.

The published standards for sheep and cattle in zoolog originate from traditional breeds or archaeological specimens, but two sheep standards (Davis and Clutton) and a goat standard (Clutton) originate from
populations living in islands and, therefore, potentially developing particular allometries (Evin et al. in this issue). In addition, different references have a different number of measures, which may also have an impact on the values obtained. We therefore explored the effect of the reference choice on the dataset provided with the package (see R script, supp. mat. 1, example 7, see also Table 3 for the raw measurements of each standard). Figs. 1 and 2 show the effects of the different standards on the osteometrical data for cattle and sheep from the three Iron Age archaeological sites (ALP, OLD and TFC) that compose the dataset available in zoolog. The effect of the reference choice on goats and pigs is available in the supplementary materials (Figs. S2 and S3).

5.1. Cattle

There are four references available for the Bos genus. The full description of each reference is available in the references help page in the R console:

```
help("reference")
```

Fig. 1 shows the results obtained using the four Bos references on lengths (left) and widths (right) after selecting one measurement per bone specimen in the three sites provided in the zoolog sample dataset ("ALP", "OLD" and "TFC"; priority method using CondenseLogs function; full R script available in S1, example 7). Note that the OLD assemblage has very few osteometrical data, and therefore does not have any cattle length value. The four references provide different results according to their origin: the Basel reference corresponds to a modern specimen of a traditional breed in Switzerland (Stopp and Deschler-Erb, 2018), and Degerbol values correspond to an aurochs (Degerbol and Fredskild, 1970). Consequently, Iron Age values from the three sites are notably below these two references (generally below −0.05). In contrast, the reference from Johnstone and Albarella originates from Iron Age data from Durrington Walls (UK; Johnstone and Albarella, 2002), and Nieto reference corresponds to a Bronze Age cow (Nieto-Espinet, 2018). Consequently, the log values obtained on the Iron Age archaeological material are much more similar to these latter references (and therefore close to zero). Those differences in the global level are expected and do not affect the analysis of the observed patterns between groups (although can affect the comparability with other works). Interestingly, the tendency in lengths is similar no matter the reference used: TFC cattle have higher values than ALP cattle. The tendency on widths is not so straightforward, but the OLD value is consistently smaller than the log ratios obtained from ALP and TFC no matter the reference used, and the differences between ALP and TFC width log ratios are small. The main difference introduced in the analysis by using a reference or another is the number of measurements available, which leads to more or less bone specimens that can be used. In this particular case, Basel and Nieto seem to provide the most appropriate references based on the number of specimens and the distribution of the measurements.

Let us emphasize that it is not the number of measurements, but the number of specimens retained for the analysis that is important. More measurements from the same specimen would be correlated, not providing more independent data. This is actually one motivation to use summary Length, Width, and Depth variables, choosing only one measurement per specimen and dimension.

5.2. Sheep

There are four references available for the Ovis genus. Fig. 2 shows the results obtained using the four references on Iron Age sheep lengths (left) and widths (right) from the three sites provided with zoolog sample dataset ("ALP", "OLD" and "TFC") after selecting one measurement per bone specimen (priority method using CondenseLogs function).

The number of measurements available in each reference is notably different (see Table 2), and this has an impact on the number of archaeological bone specimens that can be considered, most notably on lengths. In this particular case, the references from Clutton-Brock and Davis (1996) provide the higher number of sheep bones with length data. Archaeological bone specimens with widths data are more numerous (Fig. 2, right), and the number of width measurements in each reference does not seem to have a major impact on the final number of bones that can be considered at each site. Other than the number of measurements available for each reference, the expected differences in the mean level according to the used reference are also evident, more clearly for the length values. The specimens from Basel (adult male, Stopp and Deschler-Erb, 2018) and Uerpmann reference (female sheep from Western Iran, Uerpmann and Uerpmann, 1994) are notably larger than the Soay sheep males and the Shetland sheep females described in Clutton-Brock et al., (1990) and Davis (1996) respectively. Differences on widths are similarly important, but are less evident because the intragroup variability is higher (Fig. 2, right).

Despite the differences in size and the number of measurements available in each reference, the tendencies observed at the three archaeological sites are consistent: OLD specimens provide the higher values, both in lengths and widths, no matter the reference used, and TFC means are slightly higher than ALP ones, most notably on lengths.

6. Discussion

The use of log ratios in zooarchaeological analysis has benefits and also limitations. zoolog provides a tool to compute log ratios in seconds and in a flexible way, so that the users can explore and choose the most appropriate reference for their data, without being hindered by differences in language and recording traditions. For that, the package includes four thesauri regarding taxa, elements, measures, and identifiers. This is also helpful in the cases where a database is not fully consistent or several databases are combined, as it is the case of the data provided.

![Fig. 1. Comparison between the four references available for Bos applied to the cattle measurements provided with zoolog (three Iron Age sites from Catalonia, Spain).](image-url)
with the package ('Bos' and 'bota' both used to refer to 'Bos taurus'). It also includes functions that enable users to expand the thesauri and tailor the package to the nomenclature used in their databases (AddToThesaurus function). Following recommended best practices, zoolog enables the easy selection by predetermined criteria of one length, one width and one depth value per bone specimen (CondenseLogs) to avoid the over-representation and correlated values that could happen otherwise. The priority method was designed following Davis (1996) to privilege the more robust measurements minimising inter-operator differences. Nevertheless, users have full flexibility to modify the criterion and the order of measure’s priority (see example 4 in the R script, supplementary material). If desired, the package also enables the standardisation of the dataset (StandardizeDataSet), by default to English with scientific terminology.

More than anything, the package was designed to facilitate the calculation of log ratios and enhance the comparability of results between researchers across the globe. For this reason, the public availability of good reference standards is crucial.

According to the results obtained when using different references (standards) to the Iron Age data provided with zoolog, the main difference corresponds to the number of measurements in each standard, which has an impact on the final number of bone specimens that can be used in the osteometrical analysis. Users are encouraged to explore what reference(s) adapt(s) the best to their data (see example 7 in R script, supplementary material), both in terms of bone specimens and distribution of values. Similarly, we are open to suggestions and criticism to improve the package in future releases.

Author contributions


Data availability

The software code, osteometrical references, and example data are available the GitHub public repository https://github.com/josempozo/zoolog. They can also be installed as a CRAN package (https://CRAN.R-project.org/package=zoolog).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was developed as part of the ERC-Starting Grant ZooM-West (Grant agreement 716298). The package was first presented in the workshop 'Zooarchaeology in the Interconnected Mediterranean', that took place online in March 2021. This Special Issue in Quaternary International is the result of a collaborative effort between ZooMWest team members and colleagues participating in the online workshop. We thank them all for their feedback on the package and spotting a number of bugs in the first zoolog versions prior to the major release 1.0.

We are also particularly grateful to Sabine Deschler-Erb and Barbara Stopp, from the University of Basel (Switzerland) for making the reference values of several specimens available through the ICAZ Roman Period Working Group, which were included in zoolog with their permission. We also thank Francesca Slim and Dimitris Filiglou from the University of Groningen, Claudia Minniti from University of Salento, Sierra Harding and Nimrod Marom from the University of Haifa, Carly Ameen from the University of Exeter and Mikolaj Lisowski from the University of York for providing additional reference sets. The thesauri benefited from the contributions from Moussab Albeoso, Canan Cakırlar, Jwana Chaboud, Jacopo De Grossi Mazzorin, Sabine Deschler-Erb, Dimitrios Filiglou, Armelle Gardeisen, Sierra Harding, Pilar Iborra, Michael MacKinnon, Nimrod Marom, Claudia Minniti, Francesca Slim, Barbara Stopp, and Emmanuelle Vila. Mikolaj Lisowski (University of York) and Allowen Evin (CNRS-ISEM Montpellier) saw potential pitfalls in the use of cattle Degerbol and sheep Davis references that have been included in zoolog help pages and vignettes. We thank all colleagues for their contributions, comments, and help. In addition, zoolog users are encouraged to contribute to the thesauri and share other osteometrical references so that the package can be useful around the globe. We commit to thank all users providing feedback to improve zoolog usage.

Finally, we thank the referees that reviewed this article for their useful comments and suggestions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.quaint.2022.11.006.

References
