



README file for South American dataset of FAIR Phytoliths Project

Data collection methodology

List of south american countries:

https://en.wikipedia.org/wiki/List_of_South_American_countries_by_area

Download Publish or Perish software (<https://harzing.com/resources/publish-or-perish>) - you can run keyword queries on Google Scholar from there, and export the results to .csv.

1. Use Publish or Perish to run a search for keywords 'Phytoliths' and 'South America'.

2. Save this search as a csv file in the data-raw/search-data folder on google drive. 
3. Start working your way down the search list:
 - a. All articles must be saved in South-America/data-raw/articles folder.
 - b. Our criteria for our study is the article must have primary data, be from the region and published in the date range of 2016-2020 (first published on internet).
 - c. Once you have downloaded an article and started work on that article for the FAIR assessment - change the colour for text in the 'to edit' search doc - green for meets our criteria and red for not meeting it.
4. Each time you work on the FAIR assessment - please put notes below with the date, your name and what you have done in your working session. This will make it easier for us to work asynchronously on this task.
5. The R package 'roadoi' (uses unpaywall with R) has been used on our list of articles to find information about the open access status of the articles. Documentation for the data categories in this package can be found here:
<https://unpaywall.org/data-format>.
6. This information was integrated into the data set after the rest of the data was collected.

Moderation:

- After doing the assessment on the paper, highlight your comments in **yellow** (in this file) any papers in need of moderation. A moderation meeting will then be conducted to discuss this paper and any data categories that have been hard to assess.

Data collection work session details:

15 September 2021 - Emma Karoune

- Used publish or perish to do search of google scholar using key terms above - phytoliths and south america.
- Saved as csv on my local computer and also put in the google drive folder.
- Made a 'raw' and 'to edit' version of this as a google sheet.

20 September 2021 - Javier Ruiz-Pérez

- Selected 50 publications (highlighted with green in the 'edit' version; publications were highlighted in orange when they could not be accessed but might be potentially selected).
- Uploaded pdf copies of the selected publications in data-raw/articles.

7 October 2021 - Emma Karoune

- Renamed all of the articles
- Checked date range.
 - We cannot use: Watling 2016 as it was available online in 2015
 - Coe 2016 as available online in 2015.
- We also need to look at Lombardo et al 2016 - is it south america? - maybe take over to a separate list for methods papers. - Javier: samples were collected in Bolivia, so we can keep it.
 - Murungi & Bamford 2020 - not south america but is methods paper.
- Also added in any journals names into our form as there seemed to be quite a few new ones that are not on our list.

We need 4 more articles:

- Checked orange ones - Osterrieth(row33) and Pearsall(row 36 &14) - are book chapters so we don't include this - changed to red.
- Row75 - Crifo et al - is just a conference abstract so we don't want that - changed to red
- Row 101 - Morales - can't find anywhere on internet.

Added: Coe et al 2017 from Flora - -row58 - downloaded article

- Frayssinet et al 2019 - row 188
- Huismann 2019 - row 195
- Barros et al 2016 - row 206

11th October 2021 - Emma and Javier

- Javier has added closed to Journal-type - as some journals from s america do not give an option to be open access.
- Leave all categories for open access articles to the end so that this can be done with the R package above.
- Javier did the FAIR assessment of Aguilar et al. 2018, Astudillo 2018, and **Astudillo 2020**.

16th October 2021 - Celine

- I just assessed the article Astudillo 2018 who used ICPN 01. To me the ICPN 01 recommendations are not fully used as the phytoliths were studied in 2D with some morphotypes labelled with anatomy-histology features such as achene.

22th October 2022 - Celine

- Javier will work on the 22 articles from the bottom of the list while Celine will proceed from the top.

27th October 2021 - Javier

- Javier did the FAIR assessment of Zucol et al. 2019, Zucol et al. 2018, Watling et al. 2020, and Watling et al. 2018.

- Yapuchura et al. 2018 (row 130) was removed from the list because the publication does not provide information on the origin of samples.

29th October 2021-Celine

- Celine did the assessment of Caligari et al.2017 QI437 and Caligari et al.2017.

3rd November 2021 - Celine and Javier

- Celine did the assessment of Catil et al.2020.
- Javier did the FAIR assessment of Stevanato et al. 2019, De Rito et al. 2018, Ricardo et al. 2018 and Rasbold et al. 2020.

8th November 2021-Celine

- Celine did the assessment of Chueng et al. 2019, Coe et al. 2017, Contreras et al. 2019 and Contreras and Zucol 2019.

9th November 2021 - Celine and Javier

- Celine did the assessment of da Silva et al.2016 (article in Portuguese),
- Dias et al.2019 which is about the phytolith production of a specific plant and is not a reference collection, **should we add a new type of study choice such as plant morphology?**
 - Javi: I would say you can classify it as plant physiology
 - CK - Yes, I think you are right. Morphology can be understood under physiology and we do not create a new type of answer
- **Dickau et al.2016, Data (raw counts) are in the appendices. It is an open access article but I did not find any repository. Might be better to double check this one.**
- Javier did the FAIR assessment of Rasbold et al. 2019, Ramos et al. 2020, Ramírez et al. 2018, and Plumpton et al. 2019.

11th November 2021 - Javier

- Javier did the FAIR assessment of Plumpton et al. 2019, Plumpton et al. 2019, **Piperno et al. 2020**, Patterer et al. 2019, Patterer et al. 2020, da Silva Neto et al. 2018, da Silva Neto et al. 2020, Mourelle et al. 2020, Morcote-Ríos et al. 2016, and Luz et al. 2019.

12th November 2021 - Celine

- Celine did the FAIR assessment of Honaine et al. 2018

15th November 2021 - Celine

- Celine did the FAIR assessment of Frayssinet *et al.* 2019, Heider and Lopez, 2018 Hilbert *et al.* 2017 and Honaine *et al.* 2016.

16th November 2021 - Celine

- Celine did the FAIR assessment of Huisman et al. 2018, Huisman et al. 2019, Iriarte et al. 2020, Lombardo et al. 2016
- **Some articles are oriented on ecosystem research questions and are taking many countries into account. Shall we all list them according to Geonames or keep the denomination used by the authors (such as South West, Central and East Amazonia)?**

18th November 2021 - Celine

- Celine did the FAIR assessment of Lombardo et al. 2019 and Lombardo et al. 2020.
- Celine integrated the article DOI in the 2021-09-15_phytoliths-south-america-google-scholar-search_to-edit excel file.

26th November 2021 - Javier

- Highlighted in green the assessed papers in file 2021-09-15_phytoliths-south-america-google-scholar-search_to-edit and added/corrected their DOIs when necessary. Obtained open access/repositories information of the papers using the roadoi R package (function oadoi_fetch), which collects the information from Unpaywall. Added the file Unpaywall_South_America containing the results obtained (both the open access status of the assessed papers and the available open repositories) in FAIR-assessment/South-America/data-raw.

21st January 2022 - Emma

- Changing open access column to oa_status from unpaywall R search data.
- Need to check row 27 as oa status not clear? Diamond or gold access? - will check with spanish speaker - now checked - it is diamond.
- Morcote-Ríos, G., Bernal, R., & Raz, L. (2016). - check type of journal as says bronze oa - so should be down as hybrid but is down as closed at the moment. - checked and it is hybrid.
- Checked journal-type column too
- Need to take out columns K and L ?

23rd January 2022 - Emma

- Added repository information for column K
- Leaving column L for now
- Downloaded final version of raw data file as all complete for south america now.

Need to do a final check though.

4th Feb 2022 - Emma

Just rechecked the data set and there were a few things missing:

- Row 38 - column P and Q missing
- Row 49 - same as above

I have checked this and made the changes and now it is complete.

The final raw data is called:

2022-02-04_FAIR-Phytoliths-Data-Assessment-Final-South-America-raw

This file should not be altered now.

You can use the other file to start data analysis - the one with responses at the end of the name.