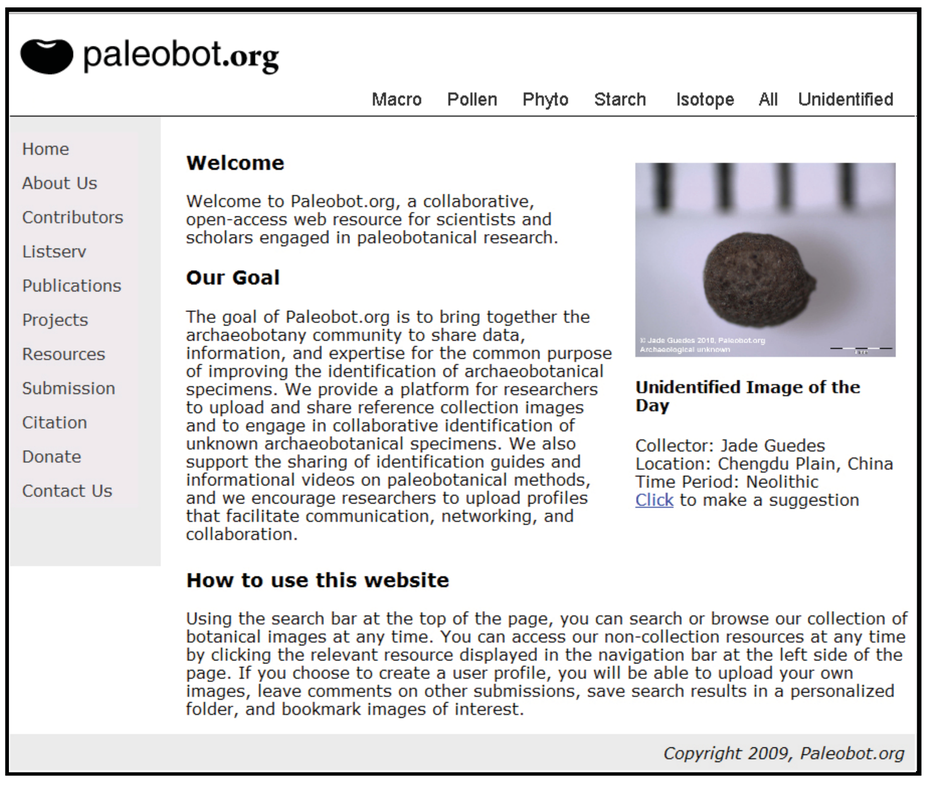
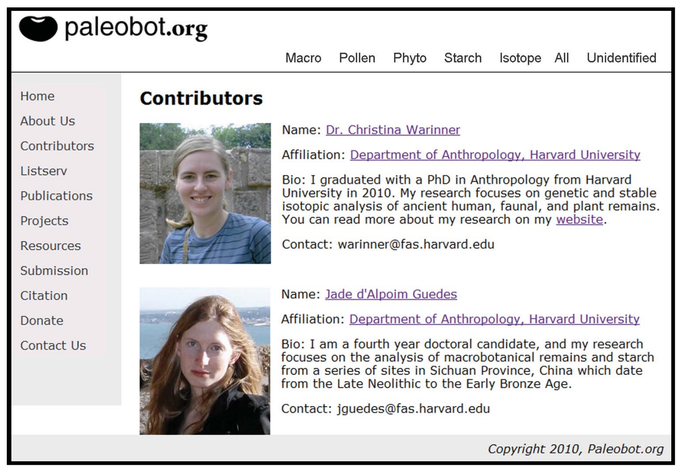
Establishing open-access online reference collections for archaeological research: macrobotanical, microbotanical, and isotopic data

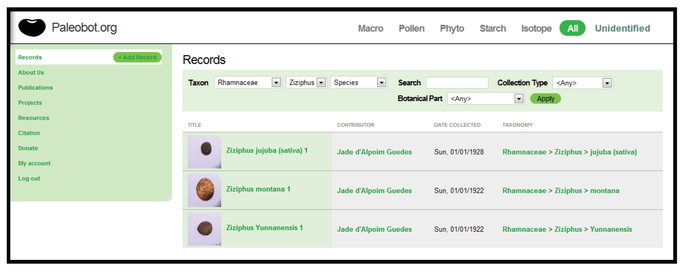
Jade d'Alpoim Guedes and Christina Warinner
Department of Anthropology, Harvard University

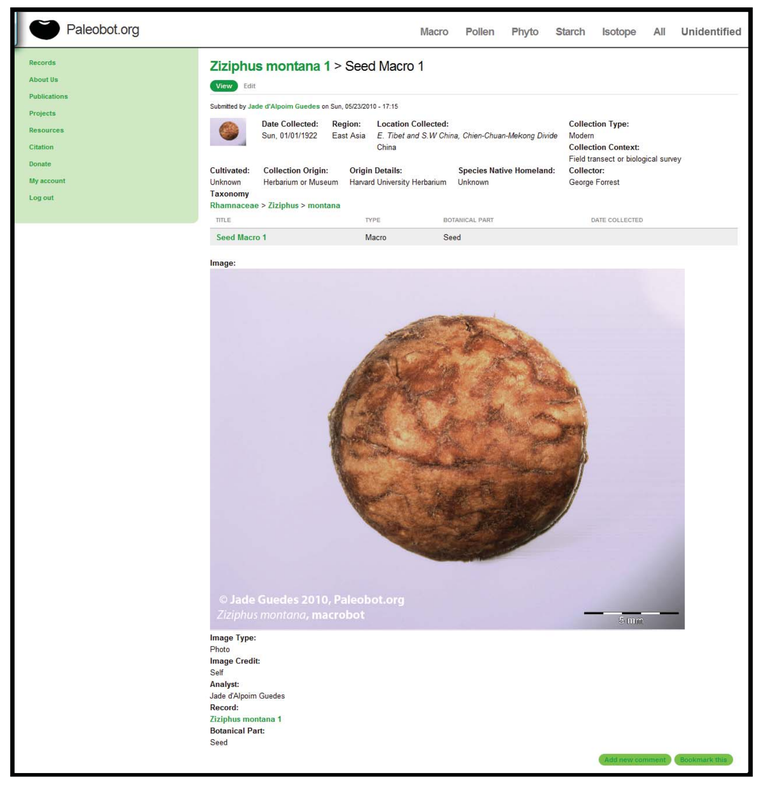
15th Conference of the International Work Group for Palaeoethnobotany
May 31-June 5, 2010 in Wilhelmshaven, Germany

1 paleobot.org









2 ABSTRACT

Limited access to high quality reference material has been a limiting factor in the advancement of archaeobotanical research. However, advancements in web-based database technology and faster downloading times make online curation of archaeobotanical reference images feasible for the first time. We describe the establishment of an open-access online reference collection database for macrobotanical, microbotanical, and isotopic data to help to standardize and improve the identification of archaeobotanical remains.

3 INTRODUCTION

Access to a suitable reference collection is a major limiting factor in the positive identification of unknown archaeobotanical specimens. As archaeobotanists are now working across the globe an increasingly large challenge has become the establishment of large regionally based reference collections. The identification of unknowns using Herbaria alone can be costly both in terms of travel and time. In many of these regions, access to Herbaria and identification guides is poor. Online reference collections can greatly help archaeobotanists narrow down the range of taxa which require viewing at Herbaria visits.

The development of new fields in microfossil analysis such as starch grain analysis and phytolith research have created new demands on the field. Currently no online reference collection exist for starch and as a result researchers have a poor understand of variability in starch grain morphology and positive identification of species remains highly problematic.

4 SURVEY RESULTS

Naomi Miller's 2010 survey
The level of confidence with which one is able to make an accurate identification of a given species is closely related to the size of an individual's reference collection. In 2010, Naomi Miller distributed an Archaeobotany Questionnaire (http://www.sas.upenn.edu/~nmiller0/AbotQ.pdf) to determine the state of current archaeobotanical research and identify areas in need of improvement.

Call for online reference collections
One of the major challenges listed by archaeobotanists in this survey was the lack of identification tools and reference collection images for their specific region and overall. When asked what would facilitate or enhance their own archaeobotanical research in a practical or intellectual way, respondents listed the development of online identification databases as the single-most desired resource.

Need for unidentified forum
Several participants in Miller's 2010 survey expressed the need for a website dedicated to the identification of unknowns. The primary use of the current archaeobotany listserv has been to disseminate photographs of unidentified plant specimens. Paleobot.org will have a special page devoted to the identification of unknown specimens. Individuals will be able to upload their images to this portion of the website and the comments feature will allow the contributor to give additional details regarding potential identifications. Registered users of the website will be able to add comments to the image in order to assist with its identification. Once a positive identification has been achieved, the image can be moved to the complete database. A record of the comments which led to its identification will be maintained below the image.

5 IMAGE-BASED RESEARCH

Paleobot.org is structured to facilitate the kind of image-based research most relevant for today's archaeobotanists. At the click of a button you can search for a particular taxon or browse the entire image collection. There is no need to scroll through tedious lists of hierarchical links to access image files. You can also leave interactive comments on your own or other researchers' entries. Such comments might include 1) noting that a feature is diagnostic for a particular taxon, 2) commenting that a feature is rare or absent in most specimens, 3) updating taxonomic information to reflect new developments in botanical classification, etc.

6 CURRENT ONLINE SITES

There are currently a large number of web based resources available to archaeobotanists specializing in macro and microfossil remains.

Macrobotanical remains
Increasing numbers of flora and herbarium sheets have been digitized however only in a few instances do these contain high quality images of seeds which can be useful for as identification tools for paleoethnobotanists. Some exceptions to this include the Digital Atlas of the Netherlands and Discover Seeds which contain high quality images of seeds. The Digital Atlas of the Netherlands, funded by the Rijks University of Groningen and the German Archaeological Institute, is in the process of assembling a website for high quality images of plants from archaeological sites. In addition, several paleo-ethnobotanists (e.g., George Willcox, Gayle Fritz) have created individual websites in which they have uploaded seed drawings and guides to assist in seed identification.

Wood
A number of useful websites exist for wood identification. Examples of these include Inside Wood which contains a large collection of reference collection photographs and an online identification key based on morphological characteristics.

Phytoliths
Only two websites, the University of Missouri Phytolith Database and the UCL Phytolith Teaching and Research Images, currently host large numbers of reference collection images for phytoliths, and neither allow open access submission.

Starch
There are currently no websites hosting reference collection images for starch.

Practical Limitations of Current Online Resources
1. **Closed access**: Currently, if an individual wishes to place photographs from their reference collection online, it has been necessary for them to create an individual webpage. Aside from Inside Wood none of the current online resources allow individuals to upload images from their own reference collection.
2. **Regional focus**: With the exception of the Seed Identification Workshop run by Ohio State University and Discover Seeds hosted by Colorado State University most online collections are region-specific, which poses challenges to researches engaged in cross-regional research or who work in understudied regions.
3. **Difficult searches**: web-database design plays a large role in the usefulness of a particular online resource. Most websites have a taxonomic organization, meaning that you search for images by entering in taxonomic information, such as genus and species. Archaeobotanists, however, begin with an image and try work backwards to a taxonomic classification. Depending on the structure of the website, searches using this process can be overly slow and tedious.
4. **No unidentified forum**: Miller's 2010 survey found that many researchers wanted a website devoted to the identification of unknowns. Currently, no websites allow individual researchers to upload unknowns and solicit assistance with identification.
5. **Few archaeological images**: Few websites contain images of archaeological plant remains.

7 OPEN ACCESS

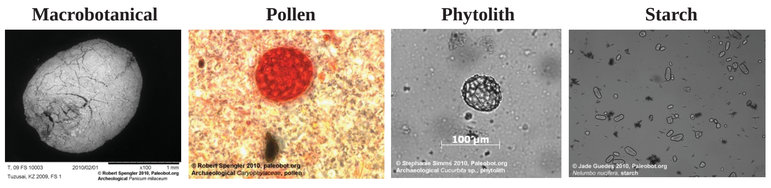
Open Access: Submission and Viewing
Many websites allow open access viewing, but with the exception of Inside Wood, none of the current websites allow open access submission of photographs from reference collections or archaeological sites. Open access submission has proven to be a powerful tool for the rapid development of new research and the dissemination of data to both professionals and the public. Examples of successful open access web-databases in biology include GenBank and Morphbank. Current non-open access websites contain data from either a regional herbaria or an individual's regionally specific collection. For individuals working in areas for which taxa from two different regions must be consulted such as the Eurasian Steppes, they are faced with trying to access reference collections from widely disparate sources. When such websites contain data from archaeological sites, these are also limited in regional scope and focus. Open access submission is important for allowing data from otherwise underrepresented regions to be accessible to the scholarly community. In addition open access submission will encourage individuals who otherwise would not have the means to develop a website to place their collection online.

Emerging Fields
We anticipate that Paleobot.org will play an important role in the development of identification procedures for newer fields such as phytolith research and starch grain analysis. In particular, little is understood of intra species/genus/family variation in starch due to a lack of access to large numbers of reference collection images. We hope that by making large numbers of images available online a better understanding of intra taxa variation will be achieved.

8 DATA TYPES

Archaeobotanists work with a wide variety of macro- and microbotanical remains and require access to large reference collections to be able to effectively and accurately identify specimens. Paleobot.org curates images of the most commonly analyzed types of macro- and microbotanical remains to facilitate this process. In addition to image files, C, N, O, & H isotopic data may also be uploaded.

Macrobotanical | **Pollen** | **Phytolith** | **Starch**



9 INTELLECTUAL PROPERTY

We acknowledge that reference collections take considerable time and effort to construct. For an open-access online database it is essential that contributors are properly credited for their reference collection images. Paleobot.org takes a number of measures to ensure that the authors of these images are given proper citation and acknowledgment for their work:

Individuals using images from the collection will be prompted to cite the author and paleobot.org using the following format: Author Name (2009). Published on the Internet http://www.paleobot.org [accessed DAY MONTH YEAR].

In order to ensure that images placed on our website cannot be used in publications without the authors credit our website will employ a software which will automatically embed copyright information (© [author's full name] [date], www.paleobot.org) on all downloaded images. We hope that individuals will be able to use content and images from Paleobot.org for their own identification purposes and in non-commercial scholarly and educational materials without special permission as long as they include proper photo credit and citation. submit images to the website.

Contributors of images to the website will also gain exposure through our contributors page. Contributors to the website will be required to create a personal profile in order to be able to submit images to the website.

Following creation each contributor's photograph, institutional affiliation, contact email address, websites and research interests will be featured on the contributors tab of the website. This will page will facilitate contact between the users and contributors of images. In addition it is expected that this portion of the website will give individuals research and personal websites additional exposure. As such, it will serve as a hub for connecting researchers around the world.

10 COLLABORATION

Collaboration
The costs and time associated with developing and maintaining an archaeobotanical research website have been deterrents to archaeobotanists wishing to upload reference collection material. Paleobot.org provides a forum where individuals can upload these images in a time efficient manner. We invite fellow researchers to submit their photographs and data to Paleobot.org and to participate in our uniquely collaborative archaeobotanical web-database.

Our Goal
The goal of Paleobot.org is to bring together a large academic community of archaeobotanists to share data, information, and expertise for the common purpose of improving the identification of archaeobotanical specimens. Our goal is not to duplicate the function of a herbarium, reference collection seeds, or slides. Rather, we hope that the images uploaded to the website can be used to assist researchers with preliminary identification ,which can then be confirmed with use of herbaria, literature, or assistance of a specialist. It is not a replacement for herbaria or plant taxonomists, but a way to use them more effectively.