The Indus Civilization (c.3200–1500BC) was one of the great early complex societies of the Old World, spanning large parts of modern Pakistan, India and reaching into Afghanistan during its urban phase (c.2600–1900BC) [18, 28, 8, 9, 36, 17, 16, 6, 27, 1, 37]. The expansive nature of the Indus Civilization meant that settlements geographically and culturally differed, and this is reflected in their modelled subsistence practices [26, 27, 32, 5, 33, 37, 20, 35, 21, 23]. There has been a long history of archaeobotany in the Indus Civilisation (see [10]), with reviews of published data as far back as [7, 8, 9]. Recent reviews include [12] exploring a pan-South Asian plant use and the role of the Indus within this [25], highlighting the eastern and southern Indus plant exploitation, and species specific reviews such as [19, 34, 24]. [13] used a rank-order analysis to look at specific crop species from Indus sites, while [10, 11] and [38] looked at specific sites in the Indus. Systematic reviews of the datasets can be found in the works of [31] and [15] who explored the wider archaeobotanical research of South Asia. This dataset seeks to build on these and update them to the current day by looking at the published Indus archaeobotanical data available up to October 2017.

The collection of this dataset of published archaeobotanical data from the Indus Civilisation (c.3200–1500BC) was carried out by the author as part of her doctoral work, and has continued up to October 2017. The dataset represents a systematic collation of all primary published macrobotanical data, regardless of their designation as ‘crop’, ‘fully domesticated’ or ‘wild/weedy’ species. The dataset comprises 63 sites and 339 ‘taxa’ (including less confidently identified elements such as ‘charred seed’). Data is presented as presence/absence due to different sampling, quantification and data presentation practices.

Keywords: archaeobotany; Indus Civilisation; South Asia; Bronze Age; macrobotanical; Harappan

Funding statement: This paper developed out of research conducted while the author was a PhD student working as part of the Land, Water and Settlement project, which has been investigating human-environment relations in northwest India. It presents material gathered for a literature review that formed part of the author’s PhD dissertation, and expanded upon during her first post-doctoral position as Trevelyan research fellow at Selwyn College, University of Cambridge. The PhD research was funded by the Arts and Humanities Research Council (AHRC), United Kingdom (Grant No. 1080510), and this paper has been written up while she has been a Post-doctoral Research Associate at the Joukowsky Institute for Archaeology and the Ancient World, Brown University, and published while she is a post doctoral fellow at the Department of Anthropology, University of Pennsylvania. The Land, Water and Settlement project ran from 2007 to 2014 and was primarily funded by a Standard Award from the UK India Education Research Initiative United Kingdom (UKIERI) under the title ‘From the collapse of Harappan urbanism to the rise of the great Early Historic cities: Investigating the cultural and geographical transformation of northwest India between 2000 and 300 BC’. Smaller grants were also awarded by the British Academy’s Stein Arnold Fund, United Kingdom, the Isaac Newton Trust, the McDonald Institute for Archaeological Research, United Kingdom, and the Natural Environment Research Council (NERC), United Kingdom.
This conclusion highlights the need for more systematic recording and reporting styles in Indus archaeobotany, and for the continued interaction with this dataset to incorporate the quantified and roughly quantified data, as well as the unquantified datasets.

**Spatial coverage**

Description

Indus Civilisation – modern day Pakistan, northwest India (states of Rajasthan, Haryana, Uttar Pradesh and Gujarat), and Afghanistan (Takhar Province).

Co-ordinates are provided below for the most extreme extents of the Indus, but a few of these (for example the northerly limits based on Shortughai) represent outliers. **Figure 1** shows the spread of sites with archaeobotanical data.

Northern extent: (based on Shortughai) 69.5 E 37.3 N

Southern extent: (based on Daimabad) 74.6007 E 19.4228 N

Eastern extent: (based on Alamgirpur) 77.49916667 E 29.05722222 N

Western extent: (based on Sutkagan dor) 62 E 25.5 N

**Temporal Coverage**

c.3200–1500BC

Whilst the majority of data falls within this time range, some sites may have associated radiocarbon dates that exceeds these boundaries.

(2) Methods

This dataset was obtained directly from source publications. PhD theses and unpublished reports were not included as these grey literature have not always been digitised or made available in the same way that journal articles are. Some are also under embargo or have author requests not to be used in such datasets, and it was decided that data from grey literature such as these should not be included until publication. However, this dataset is designed so that these literature can be incorporated into analyses with ease at a future date. Data has been converted into presence/absence information due to different sampling strategies, quantitative methods, reporting standards and publication methodologies which made comparability questionable at the ‘quantified’/‘rough quantified’ level. This dataset does not include radiocarbon dates for two reasons – this should be compiled as a separate but comparable dataset once a systematic wood charcoal dataset is created (as many encountered radiocarbon dates were on charcoal and not seeds), and there are numerous concerns relating to the earlier radiocarbon dating programmes undertaken in Indus archaeology, including incomplete reporting of sample locations ([10]: 299; [13]: S358), a lack of systematic programmes of dating [22], as well as bioturbation, concerns about half life and sample handling. A radiocarbon dataset will thus need to be compiled separately taking into account these quality control concerns, which differ from those of an archaeobotanical dataset.

**Steps**

Archaeobotanical data collection involved an initial search of the *Indian Archaeology, a Review* journal, in which excavated reports in India are published, and *Pakistan Archaeology* in which a large number of excavations in Pakistan are published. This was followed up by locating other journals and monographs online and in libraries. Where necessary, reports were translated using Google translate or through colleagues. Relevant

![Figure 1: Map of Indus sites with archaeobotanical data. Numbers correspond to site names, found in file “ICAchbotSites.csv” deposited in the online repository.](image-url)
site periodisation and information were archived, and an assessment of the quantification method and additional sampling strategy information (e.g.: flotation) were noted. The location of sites was determined through three site co-ordinate lists: [27, 29] and [30] and recorded in decimal degrees in a separate spreadsheet. All taxa and plant parts identified by the original authors were included in the dataset.

**Quality Control**

An inclusive approach was utilised in the dataset creation: data was entered from all reports irrespective of whether the dataset author was confident of the identification (for example the inclusion of New World species such as Tradescantia sp. and Argemone mexicana), and irrespective of whether the report outlined sampling strategy methods and taphonomic aspects relating to the assemblages. Names have been checked, as older synonyms were repeatedly encountered in the reports, and these have been recorded under a single taxonomic name to avoid duplication (e.g.: Sorghum vulgare and Sorghum bicolor have been entered as Sorghum bicolor). In a few instances taxa exist that cannot be found in a relevant Flora – these have been included in the dataset, but should be treated with caution (Pisum granum; Setaria tormentosa – possibly a synonym for S.intermedia but this is tenuous; Solarium sp. – possibly a misspelling of Solanum sp.; Vigna catjang – possibly Vungui culata or Cajanaus cajan). It should be noted that concerns have been raised over the misidentification of species (see for example [10] comments on the misidentification of millets, in particular Eleusine coracana, as well as wheat and pulses amongst others, see also [34] for more on millets and [14] on pulses). These taxa can be seen in Table 1. An assessment of the reporting method has been made in the dataset: quantified – numerical data present for each taxa, often by context or phase, and some form of analysis; rough quantification – for example c.100, 100+ or only some species presented with numerical information and/or analysis; species reports by period or species reports without periodisation – presence/absence information at differing levels of detail.

**Constraints**

The level of recording varied across reports. As noted above, this included full quantified reports through to reports listing taxa from the entire site with no reference to period. The dataset creation attempted to account for this by using a presence/absence recording system for all data, regardless of quantification approach, but this limits the dataset as well. The dataset includes material collected through flotation and hand-sorting, and allowances must be made for bias in collection method when comparing datasets. Mesh size was not recorded in this dataset as it was rarely encountered in the reports. Mesh size will have a significant impact on the archaeobotanical assemblages collected, and as such there is likely to be a bias towards larger species in this dataset [10, 22, 4]. Confidence in identifications is also a constraint that must be kept in mind when using this dataset. As noted above, the inclusion of New World species in this pre-1492AD Old World dataset provides a cautionary note, as does the presence of unknown nomenclature for several accessions.

**Dataset Description**

**Object name**

ICArchbotSites – one file providing the sites, phases and sampling information for all sites with associated archaeobotanical data. Sites are described by their names, and periodisation follows that reported in the papers. The periodisation can be found in ICArchbotPeriods.

ICArchbotPeriods – one file providing information on periodisation used.

ICArchbotTaxa – one file providing the taxa data collated from publications. Table organised with sites and periods in rows and taxa in columns. Accessions noted as x (present) and (absent).

ICArchbotReferences – one file with bibliographic information for archaeobotanical reports used in ICArchbotTaxa.

**Data type**

Primary

**Format Names and Versions**

.csv

**Creation Dates**

Records created 2011-October 2017 as part of AHRC funded PhD work and Selwyn College Trevelyan Fellowship. Current csv dataset created 2019.

**Dataset Creators**

The primary researcher responsible for the data collation was Jennifer Bates.

**Language**

English and Linnaean Taxonomic Latin.

**Licence**

CCo

**Repository Location**

https://doi.org/10.7910/DVN/WSHMAD

**Publication Date**

10/04/2019

**Reuse Potential**

The dataset represents a systematic collation of macro-botanical data in the Indus Civilisation building on the foundation of those created by [31, 15, 10, 38, 11, 13]. It demonstrates the breadth of plant interactions in the Indus Civilisation (63 sites, 1449 records of plant ‘presence’ at these sites, 339 ‘taxa’ – including the chaff types, and less confidently identified accessions – from 148 confidently identified genera), and as such it has analytical potential for future research. This Indus Civilisation archaeobotanical dataset provides up-to-date collation not only of those plants deemed ‘economically valuable’ by authors (see critiques in [10, 38, 3]) but also of wild...
Table 1: List of taxa with concerns relating to data quality.

<table>
<thead>
<tr>
<th>Taxa as recorded in publication</th>
<th>Taxa as noted in dataset</th>
<th>Concern from publication</th>
<th>Sites (Period code)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acacia rugata</td>
<td>Acacia concinna</td>
<td>Older synonym used</td>
<td>Banawali (MH)</td>
</tr>
<tr>
<td>Aregemon mexicana</td>
<td>Aregemon mexicana</td>
<td>New World species</td>
<td>Almigirpur (MH)</td>
</tr>
<tr>
<td>Borrera sp.</td>
<td>Spermacoce sp.</td>
<td>Older synonym used</td>
<td>Rojdi (A, C)</td>
</tr>
<tr>
<td>Capparis aphylla</td>
<td>Capparis decida</td>
<td>Older synonym used</td>
<td>Kunal (EH, LH), Sanghol (LH)</td>
</tr>
<tr>
<td>Carex spicate</td>
<td>Carex spicate</td>
<td>European species</td>
<td>Surkotada (MH)</td>
</tr>
<tr>
<td>Citrullus lanatus</td>
<td>Citrullus lanatus</td>
<td>African domesticate – debates over time of arrival into subcontinent</td>
<td>Balu (EH), Kanmer (LH)</td>
</tr>
<tr>
<td>Coccinia cordifolia</td>
<td>Coccinia grandis</td>
<td>Possibly an older synonym, but no consensus found, so left in database</td>
<td>Balu (MH), Hulas (LH), Kunal (LH)</td>
</tr>
<tr>
<td>Cordia rothii or dichotoma</td>
<td>Cordia dichotoma</td>
<td>Older synonym referenced</td>
<td>Kunal (EH)</td>
</tr>
<tr>
<td>Echinochloa colonum/colona</td>
<td>Echinochloa colona</td>
<td>Older spelling referenced in a few publications – when recording the dataset this column included both colonum and colona</td>
<td>Bahola (LH), Lotseshwar (EH), MSD I (MH), MSD VII (EH, MH), Rojdi (H)</td>
</tr>
<tr>
<td>Emblica officinalis</td>
<td>Phyllanthus emblica</td>
<td>Older synonym used</td>
<td>Banawali (EH), Kunal (LH)</td>
</tr>
<tr>
<td>Eriophorum sp.</td>
<td>Eriophorum sp.</td>
<td>Genus separated into genus and species and misspelt – also questionable if spread into Indus region</td>
<td>Surkotada (LH)</td>
</tr>
<tr>
<td>Ficus religiosa or glomerata</td>
<td>Ficus religiosa or racemosa</td>
<td>Older synonym used</td>
<td>Hulas (LH)</td>
</tr>
<tr>
<td>Goniogyna hirta</td>
<td>Crotalaria hirta</td>
<td>Older synonym used</td>
<td>Rojdi (C)</td>
</tr>
<tr>
<td>Heliotropium bacciferum</td>
<td>Heliotropium bacciferum</td>
<td>Unknown – cannot find in flora. Cf. Boraginaceae family from Africa</td>
<td>Surkotada (LH)</td>
</tr>
<tr>
<td>Pisum granum</td>
<td>Pisum granum</td>
<td>Unknown – cannot find in flora.</td>
<td>Almigirpur (MH)</td>
</tr>
<tr>
<td>Prunus amygdalus</td>
<td>Prunus dulcis</td>
<td>Older synonym used</td>
<td>Hulas (LH)</td>
</tr>
<tr>
<td>Psidium sp.</td>
<td>Psidium sp.</td>
<td>New World species</td>
<td>Rojdi (H)</td>
</tr>
<tr>
<td>Scirpus litoralis</td>
<td>Schoenoplectis litoralis</td>
<td>Older synonym used</td>
<td>Surkotada (LH)</td>
</tr>
<tr>
<td>Setaria tormentosa</td>
<td>Setaria tormentosa</td>
<td>Unknown – cannot find in flora, possibly synonym of S.intermedia.</td>
<td>Oriyo Timbo (MH), Rojdi (A, B, C)</td>
</tr>
<tr>
<td>Siteria sp.</td>
<td>Siteria sp.</td>
<td>Unknown – cannot find in flora, possible misspelling of Setaria.</td>
<td>Almigirpur (MH)</td>
</tr>
<tr>
<td>Solarium sp.</td>
<td>Solarium sp.</td>
<td>Unknown – cannot find in flora, possible misspelling of Solanum.</td>
<td>Babar Kot (H)</td>
</tr>
<tr>
<td>Sorghum vulgare</td>
<td>Sorghum bicolor</td>
<td>Older synonym used – when recording the dataset this column included both vulgare and bicolor</td>
<td>Babar Kot (late MH, LH), Banawali (EH, MH), Hulas (LH), Kanmer (LH, H, 2800–1500BC), Khirsara (MH, 2800–2000BC), Kunal (LH), Rohira (EH), Sanghol (LH)</td>
</tr>
<tr>
<td>Tradescantia sp.</td>
<td>Tradescantia sp.</td>
<td>New World species</td>
<td>Khirsara (MH)</td>
</tr>
<tr>
<td>Trifolium alexandrium</td>
<td>Trifolium alexandrium</td>
<td>Late introduction into South Asia</td>
<td>Banawali (MH), Sanghol (LH)</td>
</tr>
<tr>
<td>Vigna catjang</td>
<td>Vigna catjang</td>
<td>Unknown – cannot find in flora, possible synonym of Vunguculata or Cajanaus cajan.</td>
<td>Hulas (LH)</td>
</tr>
</tbody>
</table>
and weedy plants that are often overlooked, and thus the full published dataset can now be compared between sites and regions and between periods. Furthermore, the data are linked to site coordinates (checked against the most recent datasets in [27, 29, 30], which provided the most up-to-date information on site location), offering potential for spatial comparisons through mapping tools. The dataset provides a reference point for further development, including the addition of radiocarbon dates and non-seed archaeobotanical remains such as wood charcoal, which can be built easily into the existing collated dataset and thus compared to look at more holistic notions of ‘subsistence’ in the Indus Civilisation. The dataset thus presents not only the current sum of published seed knowledge for the Indus Civilisation but a baseline for further dataset creation.

Acknowledgements
This project has been supported by several researchers, through their generous advice, encouragement and sharing of papers. In particular I am grateful to the co-directors of the Land, Water, Settlement project and TwoRains project, Dr C.A. Petrie and Prof. R.N. Singh, who encouraged me to create this dataset and gave me suggestions on formatting, structure and final publication format. I am also grateful to my lab colleagues in the GPR lab in the McDonald Institute for Archaeological Research, University of Cambridge for their support during my PhD and post-doc, to my project colleagues at Cambridge and BHU, and to my colleagues at the Joukowsky Institute for Archaeology and the Ancient World, Brown University and the Department of Anthropology, University of Pennsylvania for their help and advice.

Competing Interests
The author has no competing interests to declare.

References
27. Possehl, GL. 2002 The Indus civilization: a contemporary perspective. Walnut Creek, CA; Oxford: AltaMira Press.