Vegetation History and Archaeobotany
https://doi.org/10.1007/s00334-023-00982-6

DISCUSSION

Comparable quantification methodologies in archaeobotany – a work-in-progress and debate

Jennifer Bates[1] · **Carolina Jiménez-Arteaga**[2]

Received: 29 August 2023 / Accepted: 7 December 2023

Abstract

The way archaeobotanists name and quantify seed fragments is a determinant step not only in the interpretation of a given macrobotanical assemblage, but also in the degree of comparability across different sites. However, seed terminology and quantification have yet not been standardised among scholars but rely on the various geographical and laboratory traditions, as well as specific research needs and circumstances. This has created two major biases: first, the main focus has been put on plants of economic importance, specially Near Eastern and European cereals (barley, wheat, rye and oats); and second, while there has been notable discussion about quantification methods such as ubiquities, densities, proportions, or more complex multivariate statistics, there is often little explicit discussion of the actual first counting stage (i.e. how one goes from things under a microscope to things in raw data or how the Minimum Number of Individuals -MNI- is calculated). In the case of South Asian archaeobotany, the economic role of other cereal species (e.g. millets, rice) and non-cereal crops (e.g. pulses, oilseeds), as well as the usually high fragmentation state in which macrobotanical remains are found, lead us to reflect on the need to establish a more accurate and comparable quantification methodology in the region. We believe that applying this to all seed fragments will also be an important tool to better understand the role of wild taxa (e.g. fruits) in ancient diets and improve the potential contribution of weeds to disentangle past agricultural systems. In this paper, we propose a new work-in-progress terminology and counting method which, far from concluded, is intended to be a starting point for future fruitful debate and development.

Keywords Quantification · Terminology · Methodology · Identification · MNI · Comparability

Introduction

Archaeology is always working backwards from a fragmentary record to try to piece together what the original whole picture of the past looked like (Schiffer 1983). Quantification, the act of figuring out how much of something existed, is a critical element of this process. Within this there are counting methods (the summing up of raw numbers of things) and descriptive aspects of how to describe the different elements present and equate them back to the sum of the whole that originally existed. This is all worked through taphonomic and preservational processes of fragmentation, with the aim of breaching the gap between the original living and final archaeological assemblages.

Archaeobotanical quantification in particular is subject to numerous preservational and taphonomic changes that can result in alteration from whole botanical objects to fragmentary seeds/diaspores, and much work has been conducted on the impacts of conditions of preservation (e.g. carbonisation) in the process of differential seed survival, alteration and destruction (Boardman and Jones 1990; Smith and Jones 1990; Mangafa et al. 2001; Jupe 2003; Margaritis and Jones 2006; García-Granero 2015). However, when it comes to how we figure out the number of things present in a sample and how to quantify this in ways that are comparable descriptively or statistically (be this simple MNIs, ubiquities, densities, proportions, or more complex multivariate methods) there is often little explicit discussion of how one

Communicated by F. Antolín.

✉ Jennifer Bates
jbates01@snu.ac.kr

[1] Department of Archaeology and Art History, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, South Korea

[2] Department of Humanities, Universitat Pompeu Fabra, Ramon Trias Fargas 25-27, Barcelona 08005, Spain

Published online: 06 January 2024

goes from things under a microscope to things in raw data or a summarized table. The ultimate goal of quantification in archaeobotany is to have a robust idea of quantity that won't radically change or affect the statistics and subsequent interpretations if an additional sample is added into the work.

Discussion of sampling strategies has been scrutinized for many years (see van der Veen 1984; Lennstrom and Hastorf 1992, 1995; d'Alpoim Guedes and Spengler 2014; Diehl 2017; Banning 2021), as has work on morphometrics such as discussions on the changes in seed size under charring conditions or under domestication (Boardman and Jones 1990; Smith and Jones 1990; Mangafa et al. 2001; Jupe 2003; Colledge 2004; Willcox 2004; Margaritis and Jones 2006; Maass and Usongo 2007; Purugganan and Fuller 2011; García-Granero 2015), but rarely has seed quantification (describing the 'bits' present and how to count them) and quantification terminology been subjected to scrutiny. In the few cases it has been explored, it tends to focus on 'staple' cereals, predominantly from the Near East or Americas (e.g. Boardman and Jones 1990; Adams et al. 1999; Bennetzen et al. 2001; Willcox 2004; Jacomet 2006; Antolín and Buxó 2011).

In this paper we argue that differences in how we name parts of seeds and other botanical elements has a critical impact on our counting methods as part of the overall quantification methodology. Differences between regions and within regions can hamper the comparability of data. This occurs even if data are fully reported in raw form as the decisions made on what to count as half, part or whole could affect the formation of a final 'number' used later in quantification. Terminological variance, method of counting and analysis application are explored in this paper to consider what the potential miscommunication(s) across datasets might be and how this could impact replicability and comparability. We look specifically to South Asia as a case study, and within this the Indus Civilization, as this is a region we are most familiar with and has a diversity of plants both domesticated and wild (Pokharia and Srivastava 2013; Bates 2019a), but argue that disparities in quantification methodologies are a problem globally. More explicit explanation in papers of the basic quantification methodology used is needed (including within our own work) and through comparing our own terms we (JB and CJA as the authors) suggest ways we can move forwards to make our own datasets explicitly comparable.

This paper is not designed as a finalised quantification methodology or set of terminologies for use in Indus Civilization, South Asia or global studies, but instead is a demonstration of the discussions we have had, trying to reach a point where we (JB and CJA) could compare our own datasets. In talking about our own materials, we realized we were struggling to communicate, and throughout this paper, we outline how we tried to reach a point not necessarily of consensus but of comparability. It is far from perfect – instead this paper is part of an ongoing exploration of quantification standardization from a terminological point of view with a mind to considering the impact of terminology on methods of counting, data reproduction, comparability and archaeobotanical reconstruction. We hope to spark a conversation that can continue between us, and move beyond that to other archaeobotanists as well. It is work-in-progress deliberately designed as such to encourage debate.

Background

MNI is typically associated with zooarchaeology but is important also within archaeobotany, even if not explicitly stated. Figuring out the minimum number of individuals, standardizing the data to create a baseline for quantifying the fragmentation patterns and linking taphonomic changes to life assemblages is critical so that botanical objects are not over or under counted. However, MNI can be calculated in different ways depending on how a fragment is assigned/categorized within the whole it was originally part of, with respect to taxonomic diagnostic elements, and decisions regarding what the 'whole' (seed, ear, but, fruit etc.) counts as. MNI (sometimes also termed minimum number of plant parts) is also important because it can allow us to explore depositional episodes and excavation methods – for example through the refitting of parts found in two sample bags or across two contexts. As such we must ask – how do we annotate a feature as identifiable, and then link this to the counting of individual botanical objects, both fragmentary and then within the final 'individual', the whole. When dealing with poorly preserved and/or very fragmented samples to answer specific questions requiring precise quantification, any variations or vagaries in how we count and label fragments may affect the whole interpretation of the assemblage (though it can be noted that in some instances such as highly fragmented or extremely large samples, specific research questions, or different non-charred preservation settings gross/approximate quantification though imprecise may allow for accurate interpretations also). This is the reason why it is crucial to not only design accurate quantification methods but also standardize them to make data comparable and replicable across sites and regions.

However, to date there is no single standard method of quantifying archaeobotanical remains. Instead, quantification methods should be designed and applied according to specific research needs and circumstances, as data recording is often a trade-off between the nature and state of the assemblage and limitations in time and resources. Hillman et al. (1996, p 206) building from Jones (1990) discussed how MNI of Near Eastern and European cereals (barley, wheat, rye and oats) should be built from focusing on the embryo end fragments and/or whole grains only. This method has had a long impact (a tradition passed down to many researchers) – and one of the authors (JB) has used a version of this in her research after being trained to use embryo/apex and whole grain presence for MNI construction. However as Antolín and Buxó (2011) note - how do we handle other fragmentary patterns such as when the grain is split in half longitudinally so only the ventral or dorsal side is visible (as they term it Longitudinal Dorsal [LD] versus Longitudinal Ventral [LV], respectively) or when it is split in half longitudinally across the groove so that both the dorsal and ventral sides are visible (Longitudinal Ventral-Dorsal, LVD)? And what if the grain is cut transversally so neither apex nor embryo are present (Transversal Medial [TM] as they term it)? MNI would be complex at this point and we rarely see this discussed in this level of nuance in papers/reports. Antolín and Buxó (2011) explicitly explore it for this very reason in relation to European/Near Eastern cereals, but how it can apply to non-European and non-Near Eastern cereals (maize, rice, the numerous millets and pseudo-cereals), is open to debate. At the same time, CJA has come across other cereal fragment patterns which are not contemplated in Antolín and Buxó (2011), such as when the embryo end fragment (sensu Hillman et al. 1996 and Jones 1990) is split in half or further split so that only the dorsal side of half an embryo is visible. Azorín and Antolín (2014) have also shown the importance of quantification difference when applied to hazelnuts, demonstrating that different ways of counting radically alter the interpretations possible – this implies that other non-grass foods can be affected by the choices made at the basic counting stages. In another paper wild gathered fruits including acorn, wild grape, hazelnut, mastic, strawberry tree and swamp saw-grass sedge were also explored with regard to quantification-difference impacts on interpretation (Antolín and Jacomet 2015, see especially Table 3 therein), but it is notable again the focus was on species deemed economically important, even if wild gathered, and how to standardise across to other taxa is still up for debate.

With the exception of the hazelnuts and the wild gathered fruits in Azorín and Antolín (2014) and Antolín and Jacomet (2015), discussion has been mainly focused on cereal caryopsis (Hillman et al. 1996; Antolín and Buxó 2011) and, to a lesser degree, other domesticated or economically 'important' crops (see Fuller 2000; Bates 2016 for pulses), leaving other taxa almost totally aside. Although there is an increasing recognition of weed ecology as a tool to better understand agricultural systems and practices (Wolff et al. 2022) and references therein), to the knowledge of the authors no publications have raised debate on how wild seeds are being and/or should be quantified. This links to their perceived 'economic value' and also to a sense of ease at identifying the 'whole' – a single seed of grain is seen as easier to identify than figuring out the relationship between seeds and nutlets, achenes and embryos for example.

The impact of a focus on 'economic value' can be seen where researchers have tried to tackle describing non-cereals. Very little explicit discussion of the quantification methodology is seen across papers and instead regarding fruits (for example), a vast range of terms are often used to represent the parts seen (Table 1): "stone", "endocarp fragment", "spine" (Tengberg 1999); "seed", "drupe fragment", "achene", "stone fragment", "pip" (Decaix et al. 2019) or "stone", "stone with skin", "stone with pericarp", "stone skin", "endocarp" (Dabrowski et al. 2018) to name but a few. This can often make it unclear what the actual counting criteria are for each of these categories. Even in crops that have a perceived 'high economic value' like pulses the same situation is reflected: terms are sometimes specified regarding whether there is a "seed" or a "spine" (Tengberg 1999), an "entire seed", a "cotyledon" (Pokharia et al. 2017b; Dabrowski et al. 2018), a "fg. [fragment] cotyledon" (Dabrowski

Table 1 Examples of terminological diversity across different various 'types' of plants

Taxa	Seed part terminology	Publication	Region
Cereals	Embryo ends	Jones 1990	Southeastern Europe
	Transversal Apical (AP), Transversal Embryonal (TE), Transversal Medial (TM), Longitudinal Ventral-Dorsal (LVD), Longitudinal Ventral (LV), Longitudinal Dorsal (LD)	Antolín and Buxó 2011	Southwestern Europe
Fruits	Stone, endocarp fragment, spine	Tengberg 1999	South Asia
	Stone, stone with skin, stone with pericarp, stone skin, endocarp	Dabrowski et al. 2018	Southwest Asia
	Stone fragment, seed, drupe fragment, achene, pip	Decaix et al. 2019	Southwest Asia
Pulses	Seed, spine	Tengberg 1999	South Asia
	Entire seed, cotyledon	Pokharia et al. 2017b	South Asia
	Entire seed, cotyledon, cotyledon fg. [fragment]	Dabrowski et al. 2018	Southwest Asia
Weeds	Endocarp fragment	Tengberg 1999	South Asia
	Nut	Pokharia et al. 2017a	South Asia
	Nutlet	Dabrowski et al. 2018	Southwest Asia

et al. 2018), but how these are comparable is not clear due to such a diverse choice of descriptors. It logically follows that with complex morphology seen in the vast number of wild taxa exploited at sites, wild plants are commonly oversimplified as just “seed” or “caryopsis”, with few publications distinguishing “endocarp fragment” (Tengberg 1999), “nutlet” (Dabrowski et al. 2018) or “nut” (Pokharia et al. 2017a).

Traditions of terminology – that which is taught to us during our training and passed down during ‘learning lineages’, or more pervasively expected within labs or regions – feeds into this. We repeat what we learn, pass this on to the next generation and often this becomes unquestioned wisdom. If we focus on pulses to provide a concrete example of the impact of the disparities in terminology on counting and quantification, we can see serious implications for South Asian archaeology. As pulses are formed of 2 cotyledons Fuller (2000) proposed a method by which entire seeds equal MNI = 1. Each cotyledon then became labelled as ‘0.5’, as they were half of the whole. Should the researcher find 2 of the 0.5 fragments this equalled an MNI of 1 whole seed. At the same time finding only 1 of the 0.5 fragments also equalled 1 seed as it shows that at least 1 seed was present at the site (there is the minimum that 1 individual was present). Fragments however became challenging under this terminology. Sticking to the numerical naming, Fuller (2000) went on to name fragments as ‘0.25’. The implications of this are that 4 of the 0.25 fragments will add up to 1 whole. However, these are fragments, with little indication of size, diagnostic features or fragmentation pattern or process. 0.25 is therefore a term of convenience, and not reflective of the actual sum to a whole seed, but it carries with it implications for MNI that without this inside knowledge could have serious consequences for comparability and reuse of the dataset.

This is critical for South Asian archaeobotany, for which Fuller’s (2000) method has become a commonly used tradition (see Bates 2016 as an example). Macrobotanical remains in South Asia are commonly found fragmented (Fuller 2000; cited in Bates 2016), which makes quantification particularly challenging in the area. While this counting method was originally designed for Fabaceae, it has been implemented in other taxa for convenience, (e.g. jujube, sedges), which is one of the reasons why we consider it necessary to establish a more systematic quantification method for South Asian Archaeobotany. When combined with the diversity of terminology seen in above few examples, we believe there is still space for discussion regarding how fragments are both named and counted, since the impact of this complexity may have serious consequences on the final MNI, leading to incomparable datasets, especially in fragmentary assemblages. We can expand this to note that there are numerous ways of reporting the raw and final summarized datasets: for example, Tengberg (1999) records the total number of seeds at Miri Qalat, Makran (Pakistan), whereas Desse et al. (2008) present the results for Shahi-Tumb, Balochistan (Pakistan) as number of remains per taxon. In India, García-Granero (2015) refers to “macrobotanical remains” from Vaharvo Timbo, Loteswar, Datrana IV and Shikarpur, while Pokharia et al. (2017a, b) present “absolute counts”. There needs to be more discussion then within South Asian archaeology about the way we count, describe, and then present our data, in order that we can understand each other and reuse or compare our data.

This suggests three major variables influencing terminologies of quantification: a focus on ‘economically valuable’ staples (mainly cereal) and lack of discussion of other plants, traditions leading to diverse and uncritically used methodological lineages, and terms/methods created for specific needs that do not translate to other datasets. As a result, it can make our datasets difficult nigh impossible to compare both globally but even within regions, as JB and CJA have found.

What follows then is the result of discussions between the authors on their own counting methods and quantification methodologies. Building from the background above it is likely there is no single quantification methodology that can be created to satisfactorily be applied across all taxa, sites or regions, so instead what we hope to show is how we have worked through our different descriptive terms, how this affects the way in which these terms carry assumptions that affect the way other people view them and the final ‘raw number’ and MNIs reported, and the implications this might have that we did not necessarily mean to give. We are not exhaustively reporting all taxa or suggesting our solutions be adopted, simply that the underlying reasoning behind a term be reported in supplementary materials, or reported in a raw data table or as a text explaining the raw data table to show what the categories means, perhaps in the supplementary materials or in the methods’ sections of papers. What follows then is a few examples of the debates and outcomes of our conversations, some of our new terms that have been adopted in light of this. We aim simply to open up a discussion with the broader archaeobotanical community both within the South Asian realm and beyond as it has possible impacts for cross-regional comparisons. The tables (Tables 2, 3, 4, 5 and 6) are not exhaustive, comprehensive nor finalised. Instead, we lay them out here to spark further debates and to extend our conversations to the community.

Taxon-by-taxon discussions of quantification

The first aspect that we determined to be needed was a way to standardize the format of terms we used. Rather than thinking about the ‘type’ of plant we are looking at (cereal/

Table 2 Terms for cereal elements

JB	CJA	New	Description
1.0	Complete	ENT-C (Entire-Caryopsis) MNI = 1 per seed	Total grain including apex, embryo, body. Can be slightly damaged but both feature elements (apex and embryo) and the dorsal and ventral side are present. The MNI is one as there is only one whole seed present in an entire seed
Apex	Transversal apical	FF-C-A (Feature Fragment-Caryopsis-Apex) MNI: 1 per seed	Apex end of a grain including dorsal and ventral side, regardless size of fragment. Embryo end not present. The MNI is one as there is only a single apex end per entire seed
Embryo	Transversal embryonal	FF-C-E (Feature Fragment-Caryopsis-Embryo) MNI: 1 per seed	Embryo end of a grain including dorsal and ventral side, regardless size of fragment. Apex end not present. The MNI is one as there is only a single embryo end per entire seed
0.25/ fragment	Transversal medial	NFF-C (Non Feature Fragment-Caryopsis) MNI: unknown per seed	Central part of a grain including dorsal and ventral side, regardless size of fragment. Apex and embryo ends not present. The MNI is unknown as the seed could fragment in many different elements. (CJA and JB discussed whether fragments with a curve that indicates significant proportions i.e.: everything but the apex and embryo present, could be counted as one, but decided this was too risky as there could still be small NFFB present)
0.5/Long split	Longitudinal ventral-dorsal	Still to be figured out. It is a FF-C, but unclear how to describe the positionality. It might be best named FF-C-L-VD (Feature Fragment-Caryopsis-Longitudinal-Ventral to Dorsal) MNI: 2 per seed	Split down the ventral groove leaving half the apex and half the embryo visible. The MNI is two per grain as the seed splits in half down the centre. So the MNI of 2 FFVD would be 1 whole grain
Dorsal long split	Longitudinal dorsal	FF-C-L-D (Feature Fragment-Caryopsis-Longitudinal-Dorsal) MNI: 1 per seed	Split down the grain leaving only the dorsal side visible from embryo to apex end. This means the embryo will be visible. MNI is 1 per grain as there can only be one dorsal side per seed. When counting through this means if you have 1 FFV and 1 FFD you have only 1 whole grain, these two things add up to a whole
Ventral long split	Longitudinal ventral	FF-C-L-V (Feature Fragment-Caryopsis-Longitudinal-Ventral) MNI: 1 per seed	Split down the grain leaving only the ventral side visible from embryo to apex end. This means the dorsal groove/hilum will be visible. MNI is one per grain as there can only be one ventral side per seed. When counting through this means if you have 1 FFV and 1 FFD you have only 1 whole grain, these two things add up to a whole
(not used, haven't yet found these in my samples)	Longitudin-al ventral transversal apical Longitudin-al ventral transversal embryonal Longitudin-al dorsal transversal apical Longitudin-al dorsal transversal embryonal	FF-C-LTr-D-E (Feature Fragment-Caryopsis-Longitudinal and Transversal-Dorsal-Embryo) MNI: 1 per seed NFF-C-LTr-D/V (Non Feature Fragment-Caryopsis-Longitudinal Transversal-Dorsal or Ventral) MNI: 1 NFF-C-LTr-D per seed and 2 NFF-C-LTr-V per seed Alternatively for millets and grasses where hilum relevant - FF-C-LTr-D-E (Feature Fragment-Caryopsis-Longitudinal Transversal-Dorsal-Embryo) MNI: 1 per seed NFF-C-LTr-D (Non Feature Fragment-Caryopsis-Longitudinal Transversal-Dorsal) MNI: 1 per seed FF-C-LTr-V-H (Feature Fragment-Caryopsis-Longitudinal Transversal-Ventral-Hilum) MNI: 1 per seed NFF-C-LTr-V (Non Feature Fragment-Caryopsis-Longitudinal Transversal-Ventral) MNI: 1 per seed	Grain split so, only the ventral OR dorsal side is visible and only one feature end is visible. This creates a complex challenge for MNI. Thinking about wheat and barley (and perhaps rice) there is no hilum to help orientate the ventral side. So, the MNI should always be that there is only one embryo per seed found on the dorsal side, and one apex on the dorsal side, but two potential ends that cannot be distinguished on the ventral side. A hypothetical example of this then is that you could have 1 FF-C-LTr-DE, 1 NFF-C-D, 3 NFF-C-V and this means you have at least 2 seeds. Seeds with hilum rather than ventral grooved however the MNI of the fragments is reduced to one per seed, as all are either distinguished by features (dorsal or ventral like embryo or hilum) or the lack of features (dorsal apex end or ventral apex end)

Table 3 Terms for pulse fragments

JB	CJA	New	Description
1.0/whole	Entire	ENT-Leg (Entire-Legume) MNI: 1 per seed	Whole pulse, not necessarily including seed coat but with both cotyledons present. Preferably with hilum present but at least with scar visible. Nuancing of terms can be done to divide ENT-LEG (Entire legume with no hilum) from ENT-LEG-Hi (Entire hilum) if this becomes a necessity during counting. MNI is one as this is single whole seed made of both cotyledons
0.5/cotyledon	Half	FF-Leg-COT (Feature Fragment-Legume-Cotyledon) MNI: 2 per seed	One of the two cotyledons that make up a dicot. Should be entire cotyledon preferably with plumule/scar visible As there are two of these per seed, a count of 3 makes for an MNI of 2 seeds present.
D split 0.5	Unnamed and only noted in comments	NFF/FF-COT-Tr (Non/Feature Fragment Cotyledon-Transversal) MNI: 2 per seed, unless specific feature is present, then 1 per seed	Split against the natural divide in the seed. FF if contains elements like the Hilum, Micropyle, Chalaza etc. Add notation (e.g.: FF-COT-Tr-Hi; FF-COT-Tr-Cz) MNI varies depending on what is present, but in general practice it will be 2 per seed.
0.25/fragment	Indeterminate	FF-Leg-named (Feature Fragment Legume-named, e.g.: plumule apex/basal end; micropyle; hilum etc.;) MNI: depends on feature NFF-Leg (Non Feature Fragment-Legume) MNI: unknown per seed	Central part of a seed, divided into fragments that can be defined by a feature and those that are featureless. The MNI is variable as ranges from unknown in featureless fragments where the seed could fragment in many different sized elements, to variable depending on how many of those features are present, such as an MNI of 1 per seed in the radicle of a chickpea to 2 per seed in the MNI of a radicle end of *Vigna*

Table 4 Terms for fruit/nut drupes and other nut shell and fruit seed equivalents

JB	CJA	New	Description
1.0	Entire endocarp	ENT (Entire) MNI: 1 per seed	Single 'objects'
0.5	Half	L (Longitudinal) MNI: 2 per seed	Half bisected down the 'object' splitting them so you see the length of the fruit through any seed. Could potentially have L-FF and L-NFF (longitudinal feature and non-feature fragments)
0.5 D	Half	Tr (Transversal) MNI: 2 per seed	Half 'object' bisected against the seed structure, splitting them so you do not see the length of any chamber but the width. Could potentially have Tr-FF and Tr-NFF (Transverse feature and non-feature fragments)
0.25	Indeterminate fragment	NFF (Non Feature Fragment) MNI: unknown per seed FF-named (Feature Fragment with names e.g.: funiculus) MNI: dependant on feature	NFF includes any fragment that is not entire or half, even if "0.75" (more than half but with a section missing) as we cannot fully quantify the missing section and any small fragments could be counted as NFF and therefore result in a skewing of the MNI. If we were to count this is 0.75 and then have a separate NFF for the small fragment in essence we might end up with double or undercounting, depending on how that small fragment is included. We argue it is better to include "0.75" as NFF or FF

pulse/fruit or 'economically valued'/'weed' for example), we started with the idea that we are handling botanical objects and the shared value is their botanical nature. As such, we turned to botanical terminologies to start describing things.

Across the different taxa there are some common shared aspects that can be noted when describing the various elements that we find in assemblages.

Firstly – is the plant part we are looking at whole or fragmented, and does that fragment have any useful features for quantifying it? This creates our first part of the descriptor:

1. Entire/Feature Fragment/Non-feature Fragment.

We have simplified these into ENT/FF/NFF for ease of writing.

Secondly – what tissue are we looking at, and does a combination of tissues in one 'bit' change the MNI when comparing it to something of the same taxon?

This creates the second part of the descriptor:

2. Caryopsis / Cotyledon / Exocarp / Endocarp / Mesocarp / etc. (and all possible combinations)

Again, we have simplified the terms, for example Ex for exocarp, En for endocarp and so on.

Thirdly – (applying only to fragments with feature) - what part are we looking at as this affects the MNI.

3. Apex / Embryo / Base / Dorsal / Ventral / Transversal / Longitudinal / etc. (and all possible combinations)

Terms are simplified e.g. Ap for Apex.

Fourth – What features are present as the MNI is affected if these are present in limited/specific numbers. There may

Table 5 Terms for *Rumex* sp. but can be used across several 'weedy' species of similar achene structure

JB	CJA	New	Description
1.0/whole	Exocarp and seed	ENT-Ac-ScE (Entire-Achene-Seed coat and Embryo) MNI: 1 per achene	The whole achene including the Seed coat and embryo. The Seed coat should be complete enough that the embryo has not "fallen out", though some damage is acceptable. The Seed coat should still have the apex and funiculum present and all ridges as well
Embryo plug	Entire seed/ only seed	ENT-S (Entire-Seed) MNI: 1 per achene	The seed only, with no Seed coat/perianth present. While this is not the entire achene, the MNI is still 1 as there can only be one seed per achene
Unnamed (not found, probably listed as seed coat)	Longitudinal exocarp and seed	FF-Ac-L-BA (Feature Fragment-Achene-Longitudinal-Base to Apex) MNI: 2 per achene	A lengthwise split that bisects the achene (seed coat/perianth and seed) leaving half the base and half the apex visible. As this splits the achene in half there should be two of per achene One could also subdivide into FF-Ac-L- BA-Sc(ENT)S to take into account the Entire seed as opposed to FF-Ac-L-BA-ScS where the seed is split lengthwise along with the Seed coat
Unnamed (not found, probably listed as seed coat)	Longitudinal exocarp	FF-Sc-L-BA (Feature Fragment-Seed coat-Longitudinal-Base to Apex) MNI: 2 per achene	A lengthwise split that bisects the achene leaving half the base and half the apex visible. This is exocarp only, no seed As this splits the achene in half there should be two of per achene
Unnamed (not found, probably listed as seed coat)	Longitudinal seed	FF-S-L-BA (Feature Fragment-Seed-Longitudinal-Base to Apex) MNI: 2 per achene	A lengthwise split that bisects the seed leaving half the base and half the apex visible. This is seed only, no Seed coat As this splits the seed in half there should be two per achene
Unnamed (not found, probably listed as seed coat)	Apex exocarp	FF-Sc-A (Feature Fragment-Seed coat-Apex) MNI: 1 per achene	Seed coat fragment from the apex end of the achene. As there is only one apex this must be an MNI of one per achene
Unnamed (probably 0.25/ fragment)	Embryo exocarp	FF-Sc-B (Feature Fragment-Seed coat-Base) MNI: 1 per achene; ONLY if certain *Rumex* sp.	Seed coat fragment from the base end of the achene. As there is only one base this must be an MNI of one per achene Note that the seed is not adhering to this. If we might name this FF-Sc-BS (Feature fragment- Seed coat base seed) and create a new category to distinguish it from both ends without seeds and from loose seeds, as this might impact the MNI overall
Unnamed (probably 0.25/ fragment)	Unnamed (probably fragment)	FF-Sc-R (Feature Fragment-Seed coat-Ridge) MNI: unknown per achene; ONLY if certain *Rumex*	Seed coat fragment from the ridge of the seed. MNI cannot be certain as fragmentation can be into many pieces. Note that the seed is not adhering to this. If we might name this FF-Sc-RS (Feature fragment- Seed coat-ridge seed) and create a new category to distinguish it from those without seeds and from loose seeds, as this might impact the MNI overall
Unnamed (probably 0.25/ fragment)	Unnamed (probably fragment)	NFF-Sc (Non Feature Fragment-Seed coat) MNI: unknown per achene; ONLY if certain *Rumex* sp.	Seed coat fragment from the body of the achene. MNI cannot be certain as fragmentation can be into many pieces. Note that the seed is not adhering to this. If we find it with seed adhering we could create a new category to distinguish this (NFF-ScS)

be overlap between features and parts (e.g. Hilum could be listed as a part or feature, but we have chosen it as feature as the part would be the embryo end; equally one could list embryo twice – embryo end and embryo visible).

4. Hilum / Plumule / Radicle.

Plumule simplified to Pl for example.

In practice, this might look like the examples in Fig. 1:

In order to utilise the system, good understanding of botanical terms is needed. We include a few of the terms we have used to label botanical remains in our discussions in the ESM Table S1. This is by no means exhaustive, and indeed we have tried to minimize the number of terms in order to keep things simple and user friendly.

We also have some terms for position:

– Longitudinal (L) – split lengthwise down the object.

– Transversal (Tr) – split across the width of the object.

Within this we define position using some of the botanical features (e.g. longitudinal ventral to dorsal – a split lengthwise down a caryopsis showing ventral and dorsal elements on the halves created by the split; compared with longitudinal ventral – a split lengthwise down the caryopsis leaving only the ventral side visible).

When handling Seed coat (Sc) or Endosperm (En) surface patterns descriptions are necessary, as this can be diagnostic even for Non-Feature Fragments (NFF). For this, we recommend and will be using this in our identification descriptions, and potentially in our quantification where relevant for research questions. For Other (e.g. dung and food) we are referring to specific papers in which people have handled these materials (e.g. Kubiak-Martens et al. 2015;

Table 6 Terms for *Chenopodium* sp. but can be used for similar thin-coated achene species and those with curled seeds

JB	CJA	New	Description
1.0/whole	Exocarp + seed	ENT-Ac (Entire-Achene) MNI: 1 per achene	The whole achene (including the Seed coat and seed inside). The Seed coat should be complete enough that the seed has not "fallen out", though some damage is acceptable. The Seed coat should still have the radicle attachment present and entire disc ridge as well
Embryo	Entire seed/ seed only	ENT-S (Entire-Seed) MNI: 1 per achene	The seed only, with no Seed coat present. While this is not the entire achene, the MNI is still 1 as there can only be one seed per achene
0.5	Unnamed and only noted in comments	FF-Sc-L-RM (Feature Fragment-Seed coat-Longitudinal-Radicle to Margin) MNI: 2 per achene	Split around the disc of the achene along margin all the way round to the radicle, dividing the achene and leaving half the radicle attachment point visible on both halves. This is seed coat only. As this splits the achene in half, there should be two of per achene. It could be further subdivided into seed coat and seed (ScS) or Seed coat (Sc) only if needed
D split 0.5	Unnamed and only noted in comments	NFF-Sc-Tr (Non Feature Fragment-Seed coat-Transversal) FF-Sc-TR-R (Feature Fragment-Seed coat-Transverse-Radicle) MNI: 1 of each per achene	Split not down the ridge but across the width of the achene. The 'top half' has no features, while the 'lower half' has the radicle attachment point. MNI is theoretically 1 per achene as there is only on 'lower' and one 'top' half if the achene is split exactly transverse. In practice, this may get more complex and best judgment may be needed. It could be further subdivided into seed coat and seed (ScS) or Seed coat (Sc) only if needed
0.25	Unnamed (notes in comments only)	FF-Sc-R (Feature Fragment-Seed coat-Radicle) MNI: 1 per achene FF-Sc-M (Feature Fragment-Seed coat-Margin) MNI: unknown per seed; ONLY if certain *Chenopodium* sp.	Seed coat fragment from the radicle attachment point of the achene. As there is only one radicle this must be an MNI of one per achene. Note that the seed is not adhering to this. If we might name this Sc or ScS and create a new category to distinguish it from both ends without seeds and from loose seeds, this might impact the MNI overall
0.25	Unnamed (notes in comments only)	NFF-Sc (Non Feature Fragment-Seed coat) MNI: unknown per achene; ONLY if certain *Chenopodium* sp.	Seed coat fragment from the body of the achene. MNI cannot be certain as fragmentation can be into many pieces. Note that the seed is not adhering to this. If we might name this Sc and ScS and create a new category to distinguish it from both ends without seeds and from loose seeds, this might impact the MNI overall

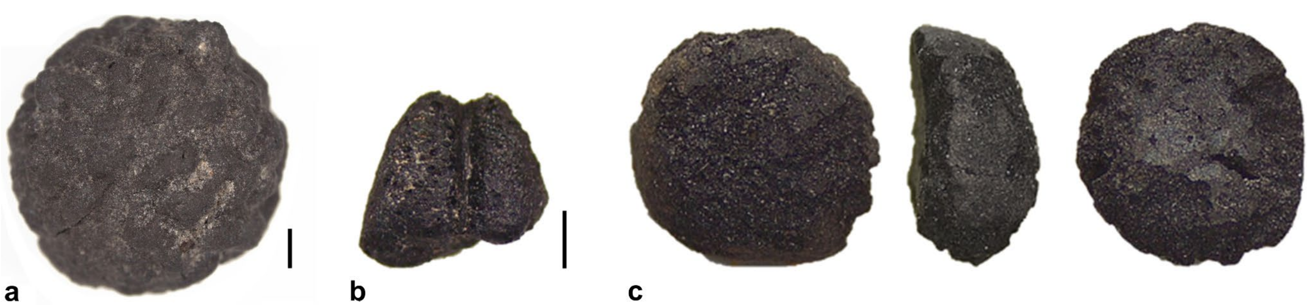

Fig. 1 From left to right: **a** *Ziziphus* sp. element showing the 'exterior' view, comprised of the endocarp (with seed comprised of seed coat, endosperm and embryo inside). This would be labelled ENT-En as it is entire (whole with no damage) and showing the endocarp; **b** fragment of a wheat/barley grain, showing a feature (apex). This would therefore be labelled FF-C-A (Feature fragment-caryopsis-apex); **c** half of a pulse showing the 'exterior', lateral and 'interior' views (Vicia sativa). This single element would be labelled as FF-COT (feature fragment-cotyledon). Scale = 1 mm

González Carretero et al. 2017; Arranz-Otaegui et al. 2018; Dunseth et al. 2019; Fuks and Dunseth 2021; Bates et al. 2022).

What follows is how we have applied these rules to specific examples of the materials and taphonomic patterns seen in our materials, how this affects our own descriptions, and subsequently how this might affect our quantification. We refer to ourselves by our initials (JB and CJA) rather than to specific sites as it is the discussions that are relevant rather than individual datasets.

Cereals

It was noted straight away that two very different sets of terms were being used for cereals within the work of the authors, perhaps in part due to sub-regional variations in the taxa and preservational patterns found. While both have large numbers of small millets, JB has rice present and a few (though rare) finds of wheat/barley, poorly preserved in general and usually whole or split in simple ways, while CJA has no rice, and wheat and barley often well preserved and in a diversity of splits (outlines below). This has resulted in a generally simpler set of descriptions in JBs work than CJA. This is also in part due to research traditions. JB works with a tradition building out of Fuller (2000), Jones (1990) and the UCL/Cambridge labs, learning from practiced traditions that have likely simplified down the years for application to a wide range of materials from across the world. CJA however works in a lab predominantly dealing with microbotanical remains (UPF) and has relied on terms borrowed from papers and created her own system for macrobotanical remains from Pakistan. These combined factors have led to JB favouring a set of terms that almost follows that for pulses outlined above (sensu Fuller 2000) and cereal terms from Hillman et al. (1996) and Jones (1990) and has tried to simplify these two approaches to make them useful across a diversity of crops found in the Indus, while CJA adopted and modified Antolín and Buxó (2011) though this became challenging on occasion with millets.

To create comparability, CJA and JB came up with a list of terms and descriptions (Table 2), and settled on new ways to describe things that both could understand and agree on. We have been testing if these work especially for rice and millets and are still refining them. It should be noted that these do not account for possible refitting and where this might occur, we are noting that as well in the lab-book/counting sheet (e.g. if a FF-C-A refits perfectly with a FF-C-E it could potentially be counted as ENT-C). Similarly, we would always count things in glumes/chaff separately from dehusked things, even when fragmented to reduce the chance of double-counting.

To put this into practice, below are some examples of the terms applied to samples in Fig. 2.

Fabaceae (pulses)

Fabaceae proved surprising easy, perhaps because it has more 'logical' (descriptively) breakage points than monocots like cereals. As dicots, legumes lean towards a split down the two cotyledons, and where this does not occur we found describing to one another the other fragmentation points seen in our assemblages proved surprising easy (Table 3). It was noted by both CJA and JB that more refinement of the fragments was needed in our descriptions going forward and issues like referring to '0.25' (JB) and 'half'/'frag' (CJA) will need to be checked in past works to think about the implications of our narratives.

The removal of numerical phrases like 1.0, 0.5/half and 0.25 make assumptions on the number of fragments needed to make up a single whole seed. Underlying some of these implicitly though not necessarily intentionally were notions that if you added up enough fragments of these categories you could reach a single entire seed. Similarly, there was a need to nuance the indeterminate/fragment category to take into account that we might be able to improve our MNIs and better incorporate some of the identifying features we see even on fragmentary remains. This has been done with the cereals which to all intents and purposes are similar to the feature fragments with plumule elements or hilum elements seen in pulses but often just listed as "fragment" without refinement.

Fruit/nut drupes

The diversity of fruits and nutlets within South Asian archaeology, indeed in the Indus alone (Bates 2019b, 2020), means that we did not have time to talk through the different permutations of taxa taphonomy, let alone have the word-space to work through all possible examples here. Instead we refined ourselves to debating what is perhaps the most ubiquitous taxon within South Asian archaeology, *Ziziphus* sp.

We have kept this table (Table 4) simple to focus on fruit/nut 'objects' found at sites, which could include endocarps alone or those with meso- and exocarp, because as we note below this can be complicated by preservation methods at sites, but at its essence there are a few simple basic descriptions we can set in place for fruit/nut drupes.

While this is a relatively simple table, this can be refined much further depending on the research questions and preservation of the material. Stepping outside the Indus but staying in South Asia, recent work by JB (Morrison et al. 2022) has resulted in identifying several variations in fruit

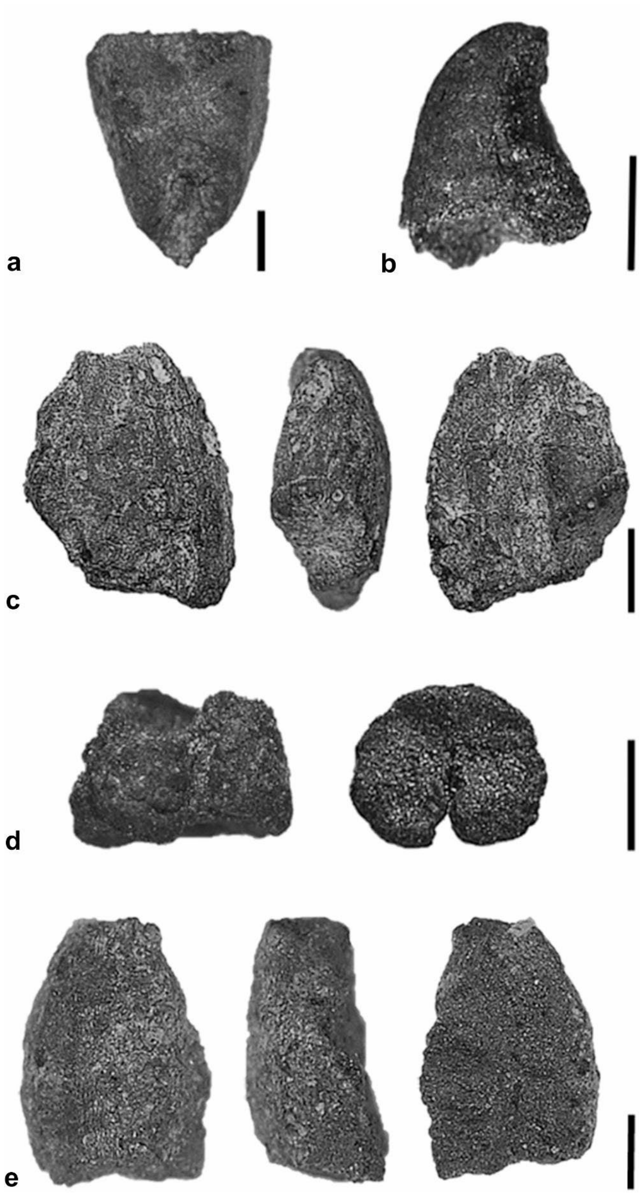

Fig. 2 Examples of cereal fragment patterns, of which b, c, d and e are not contemplated in Antolín and Buxó (2011). Notice that example e) is technically a fragment cut transversally, since both the apical and embryonal edges are cut; however, it is very unlikely that the apical edge, if preserved, will be identifiable as such (too small), so in practical terms we could say this example corresponds to a transversal apical fragment. **a** FF-C-E hulled barley; **b** FF-C-E-L-VD wheat/barley; **c** FF-C-A-L-D wheat/barley; **d** NFF-C-T wheat/barley; **e** NFF-C-LT-D hulled barley. Scale = 1 mm

preservation and fragmentation that necessitated complex description, and the discussions between CJA and JB have helped immensely with this. For example, JB has *Ziziphus* nutlets that are without any skin or flesh, with flesh, and with flesh and skin.

Using the new terminologies JB is able to identify a wide range of *Ziziphus* sp. elements at her site:

To describe the 'stone' only parts – ENT-En, FF-En-L, FF-En-Tr, NFF-En.

To describe the 'stone' with 'flesh' – ENT-EnM, FF-EnM-L, FF-EnM-Tr, NFF-EnM.

To describe the 'stone' with 'flesh' and 'skin' (full pericarp) – ENT-P, FF-P-L, FF-P-Tr, NFF-P, FF-P-Ch.

All of these are present in her samples. This splitting however can also be reduced back down later to simple MNI of ENT, FF-L, FF-Tr, NFF and FF-other for ease of basic MNI quantification and make it comparable with the data collected from earlier work in north India within the Indus which has been converted to the new terminology – we see the remains are ENT-En, FF-EN-L, FF-En-Tr, NFF-En.

By thinking through the different questions asked of the material and the nature of the taphonomy, and applying this at the beginning to the very identification and counting, JB is able to ask of it different questions and interrogate it in different ways throughout the quantification methodology. New questions are arising (Fig. 3) – why is there such diversity in JB's Southern Indian Neolithic-Iron Age *Ziziphus* sp. preservation yet only endocarps at Indus sites? How are *Ziziphus* sp. being prepared and consumed and/or reaching fire to be preserved in this fashion with the entire pericarp still present at the South Indian sites? This nuanced terminology could potentially help with understanding the functionality of the seeds or use behaviour. Margaritis and Jones (2006) studied grape processing and through ethnobotanical evidence showed distinct taphonomic changes due to processing. By noting specific categories of seed in the initial quantification, e.g. ENT-P as opposed to ENT-SEn or ENT-SEnM, we might be able to think about the changes Margaritis and Jones (2006) observed: are the exocarps missing leaving only the endocarp or endocarp and mesocarp present? This could indicate a specific processing of grapes and thus a specific behaviour depending on the exact parts still present.

Oil/vegetal/fibre

This category of seeds is arbitrary (Bates 2019b) – it does not define things taxonomically but instead socially. However, it is often cited in texts in relation to plants used for non-staple food, or non-food (e.g. fibre crops) purposes. This creates a disparate group that, like fruits, we cannot hope to discuss in its entirety, but instead in our discussions we decided to use this as an opportunity to think about whether there were terms we could pull from other categories that might make description easier.

For example, Brassicaceae often breaks into simple halves akin to those seen in pulses, so ENT, L, Tr, and FF/NFF can be useful here. Others like *Sesamum* sp. or the Cucurbitaceae family though dicots rarely split along their

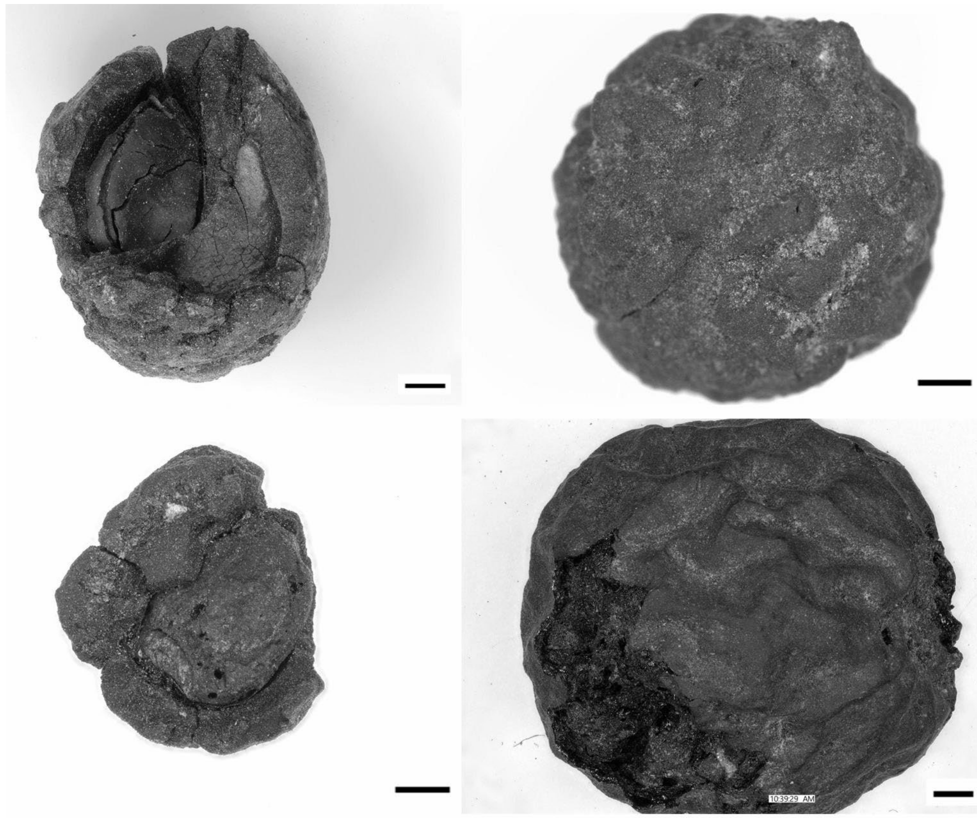

Fig. 3 Drupe fruits including *Ziziphus* sp. and *Phyllanthus emblica* from Kadebakele showing range of preservation and taphonomic changes. Top left: FF-EnS-L (Feature fragment-endocarp with seed inside-longitudinal – the MNI of this is 2 per fruit). Top right: ENT-En (Entire-endocarp – the MNI of this is 1 per fruit). Bottom left: challenging as we have ENT-En surrounded by FF-M-L, suggested terminology for this might be ENT-En-FF-M-L to incorporate this, and the MNI is 1. Bottom right: ENT-P (Entire-Pericarp, even though there is small damage to the exocarp of the pericarp it is more or less entirely there and the MNI of this is 1 as it is a complete fruit). Scale = 1 mm

Fig. 4 From left to right: sesame fragment (FF-T-R), flax fragment (FF-L-VD), flax fragment (FF-T-A). Scale = 1 mm

cotyledons, but their structure means that we can compare them with cereals in terms of descriptive elements. While seeds can be described as ENT (entire), Feature Fragment Apex (FF-A), Feature Fragment Radicle (FF-R), Longitudinal (where it splits from radicle to apex end), Transverse (splitting against any natural feature or shape) and Non-Feature Fragment (NFF), we might further start refining the fragments with similar terms from the cereals if we need to making Feature Fragment-Longitudinal-Apex (FF-L-A) if needed (for example) (Fig. 4).

Wild

We have chosen to create a category here of 'wild' taxa rather than calling it 'weeds' to avoid the trap of predetermining the analysis of any work (Bates 2019b; Wolff et al. 2022). Many of the things we might include here are already covered by discussions above – wild grasses for example, Lamiaceae that can be handled similarly to the Cucurbitaceae/*Sesamum* sp. discussions. There are many however, that will be more complex and we have chosen two examples from our discussions that posed particular challenges to our thinking: Polygonaceae and Chenopodiaceae. In particular we have focused on *Rumex* sp. (Table 5) and *Chenopodium* sp. (Table 6) as these provide complex shapes and features to work with, and are therefore examples of the many discussions we have had.

Rumex sp.

This was chosen as it is an achene – a simple dry fruit with a single seed that nearly fills the pericarp but does not adhere to it. This means *Rumex* sp. does not have a seed coat per se, but a pericarp (for simplicity of term we are sticking with seed coat), and the internal 'embryo plug' is in fact the actual seed. It also has specific morphological challenges as well – the achene is broadly ovate with an acute (not accuminated) apex, it is trigonous with rounded margins/ridges (Table 5).

At CJAs sites ironically the better preservation means that she has more elements to look at and describe than JB. JBs poor preservation meant that she only had a few rare whole achenes preserved or seed elements and what might charitably be called seed coat fragments, and as such did not feel the need to develop a terminology around this when originally analysing the material. However as more sites are analysed this is likely to change so we have decided that a shared language around them is better, and refining it so that it can be applied to more than just *Rumex* sp. (to Cyperaceae more broadly for example or beyond to other taxa) would be useful.

However, quantification can get complex with this kind of counting and description. For example, if an analyst has 1 ENT-S and 1 FF-Sc-A is this an MNI of 1 or 2? We would argue it is an MNI of only 1 as you cannot tell if the seed and seed coat apex came from the same or different achenes.

One could become granular with this – CJA for example sub divides the FF-Sc-L-BA category further into those with ridge and those without (Fig. 5), similar to the dorsal/ventral divisions in the cereal, to account for further impacts of fragmentation on the MNI.

Chenopodium sp.

Chenopodium was chosen because the lenticular shape means that we have different 'faces' and angles to handle in fragmentation. Like *Rumex* sp. it is an achene with seed coat/perianth and internal seed. However, the seed coat/perianth is thin and often does not survive and the harder seed inside appears like a curled shape. Some of the descriptions can be borrowed from pulses given the lenticular shape of the seed coat/perianth that encourages it to split 'logically' around the margin/ridge, while others have to take into account the internal structuring similar to that from Polygonaceae with the internal seed inside the seed coat, though reflecting the curled nature of the internal seed.

Fig. 5 *Rumex* sp. in three different states of preservation, from left to right: ENT-Ac-ScE; FF-Sc-B; FF-Sc-A. Scale = 1 mm

This refinement of our notes aims to help us think further about our materials and not just result in lumping things together. For example, when confronted with NFF-Sc (Non-Feature Fragment Seed coat), how far can we go with its identification? Are we confident putting it into the *Chenopodium* or *Rumex* genera, or should we be moving things to family level, or even to just "seed coat"? How many features do we feel we need to identify things, and what further work do we feel we need to do (SEM for example of seed coat pattern) to move things from NFF to FF? And equally, how confident are we to define seed coats, perianths/testas, pericarps (and within this exocarps, mesocarps and endocarps)?

Chaff

Chaff is not something we dived too far into in our discussions – this is a current lacuna in our debates. For the time being we are going to use terminology in Jacomet (2006) for cereals, Fuller and Harvey (2006) for pulses, and continue discussing as we go. We recognise this is a serious gap – however we are also faced with regional disparities. Jacomet for example is highly detailed on Near Eastern and European cereals, but less so on millets, rice and maize chaff. We are making up for this with Thompson (1996) and Reddy (1997, 2003) that look at rice and millets but these focus more on domestication changes or crop processing than refined diagnostic features. Fuller and Harvey (2006) discusses the processing of pulses but less so on the structure of pulse crop processing waste elements. JB is also dealing with sesame non-seed elements along with non-seed elements of *Ziziphus sp.*, and cf. *Citrus* sp. peels and more work to refine these in terms of identification is required even before counting terminology is needed. These discussions will continue and be the subject of further work as they develop.

Other

This is, we realize, a very unsatisfactory name for a group of materials. However, there are things we find in samples that need to be discussed that do not neatly fit into the 'seed' or 'chaff' groups.

Seed coat fragments are common (Fig. 6). Without specific diagnostic features to define them to taxa, for the time being we are suggesting NFF-Sc (Non-Feature Fragment Seed coat) is a suitable grouping, with perhaps descriptors if needed such as Thick Wall, Thin Wall, Rounded, Flat, and surface decoration information to perhaps refine them or group them further. FF (Feature-Fragments) can be refined to what feature is present and this should help them be determined at least to Family if not further. Seed coat quantification should be possible with feature fragments but not non

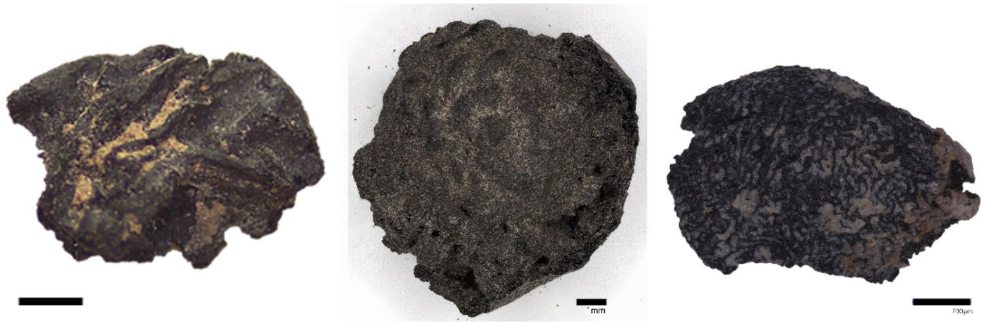

Fig. 6 Fruit skins and fruit flesh, showing the complexity of what might be encountered including those with no obvious features, those with distinct surface detailing and those with features that could be used in identification and/or MNI quantification. Left: fruit skin with no features (NFF-Sc); Middle: fruit skin and flesh attached to a nut making this the entire pericarp from exocarp, mesocarp to endocarp, *Ziziphus* (FF-P); Right: seed coat of cf. *Stellaria* (NFF-A-Sc). Scale = 1 mm for left and middle, 200 µm for right image

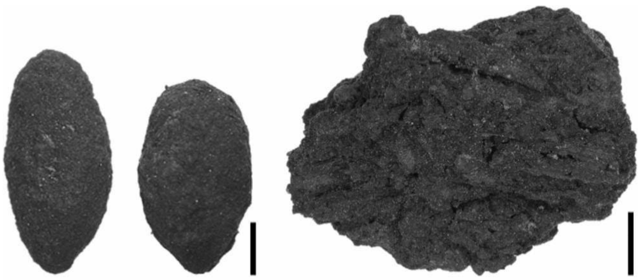

Fig. 7 Left: Entire pellets from a small-sized animal (cf. rodent), notice the edge-defined shape. Right: dung, likely from cattle, with no discernible shape and showing the layering of grassy-material. Scale = 1 mm

feature fragments. The same rules apply to loose fruit skins and fruit flesh, and we could stretch this to things like thorns for example, adding in rules on attachment and apex ends as further descriptors to refine the MNI.

Indeterminate items (JB used to term these IBF [indeterminate botanical fragments] and distinguish these from VCF [vesicular cerealia fragments], a distinction she no longer feels comfortable with) could potentially be named NFF-IDT (Non-Feature Fragment - Indeterminate), allowing them to be distinguished from the category of 'unknown' (UN). These would be things that we might be able to identify later but right now, the analyst does not recognise. The same naming (and thus quantification) conventions would be applied – FF-UN versus ENT-UN (Feature Fragment-Unknown or Entire-Unknown). Further description could then follow for each subgroup.

More complex items might include dung, pellets, plaster and food, all categories that JB and CJA have invariably used and encountered in papers. Having discussed this as a temporary measure we are using DUNG to refer to faecal matter of unrefined shape while PELLET is faecal matter with a specific edge-defined shape (Fig. 7). In addition DUNG has layered grass in it, while PELLET is more complex in composition and requires breaking open to see the contents (see also Valamoti and Charles 2005).

Food in the archaeobotanical context has been well defined by Valamoti (2002); Valamoti and Charles (2005); González Carretero et al. (2017); Arranz-Otaegui et al. (2018); Heiss et al. (2020); Bates et al. (2022) as amorphous, heterogeneous charred objects including plant particles, cell forms and tissue fragments, with pore spaces and other signs of transformation (Valamoti 2002; Valamoti and Charles 2005; Kubiak-Martens et al. 2015; González Carretero et al. 2017; Arranz-Otaegui et al. 2018; Heiss et al. 2020; Bates et al. 2022). Work is ongoing in several research groups to create identification methods and criteria.

Plaster/building materials are something that JB is going to work more on to identify, and for now a new category is created NSM (Non-Seed Mass). This new category is made because sometimes it is not easy to distinguish dung, plaster (which can contain dung), food and building/bedding materials without a lot more work. NSM therefore signifies materials that are not seeds/chaff/thorns etc. and instead are made of multiple botanical elements and not clearly categorizable as DUNG/PELLET/FOOD.

Discussion

This paper provides only the barest bones of our thoughts and discussions on quantification for macrobotanical remains, but we feel it is a start for future fruitful debate and development. Quantification is critical for a robust interpretative framework; we see in depth discussions around standardization of quantification methodologies in other subfields of archaeology like phytolith analysis (e.g. Piperno 1988, 2006; Madella 1997; Lentfer and Boyd 1998; Pearsall 1988; Barboni et al. 1999; Albert and Weiner 2001; Strömberg 2009; Zurro 2018), yet within macrobotanical studies it has seemingly lagged behind. This is likely due to the three major aspects highlighted earlier: the focus on 'economics' only, lab traditions, and a regionalisation/dataset myopia.

While we are not suggesting a globalized standard terminology or methodology for macrobotanical quantification (in comparison with phytolith analysis where the International Committee for Phytolith Taxonomy discusses this on a regular basis) due to the diversity in taxa, taphonomy and preservational pathways, we are suggesting that more open discussion on the decisions made regarding counting methodologies is needed. This will facilitate a better understanding of the published datasets, and thus more reproducibility in datasets, and further reuse of data in syntheses. We also push for more open access publishing in secure online storage/repository systems (see Lodwick 2019) in raw as well as MNI and other transformed format with a discussion of the quantification reasoning, so that such investigation and reuse can be done in rigorous format. It is critical to note that we recognise this terminology, as it stands, is highly complex and needs refining and simplifying. It also needs input from other researchers working in other regions and also on other preservation conditions (e.g. waterlogging where both depositional pathways are different but so too are fragmentation frequency). In addition, as part of Open Access efforts many researchers feed data into shared repositories (e.g. ArboDat in Europe) where such coding would not be accepted currently, and as such further debate is needed on how shared terminology and databases can develop together. This paper is, as we stress again, therefore only a starting point for discussion and meant to encourage reflection on the impact of quantification methodologies.

Quantification is not standardised and can't ever be fully standardised as our research questions, materials and goals will differ significantly. Time and experience also play a role, and if we were to try to count every fragment to such a fine level we might see extreme differences between the new undergraduate learner and the highly experienced lab technician expert in what is counted to each level. As such this is not what we are suggesting here in this thought-piece, but instead to ask what impact we are having when we do count and the terms we are having, to get a debate going in the wider community, and to see whether we can even compare our own datasets.

To highlight that this is a work in progress and that people need to outline how they make their own discussions as they go, we provide some final examples of challenges we are facing in our work that are still to be resolved and that will need to be explained as such. Within cereals – if we are to have 2 FF-C-V-T (grains split along the ventral axis and transversally to show either the apex or embryo end unclear because this is the ventral side) and 1 FF-C-A (grains split transversally to show only the apex with both the ventral and dorsal side present) is this an MNI or 1 or 2 or 3? This has turned up in our samples, as shown throughout this paper.

And this can get even more complex, as we could have 200 FF-C-V (grains split along the ventral axis), 100 FF-C-D-L (grains split along the dorsal axis and longitudinally), 50 FF-C-D (grains split along the dorsal axis but not longitudinally), 20 FF-C-A (grains split transversally to show only the apex), 10 FF-C-E (grains split transversally to show only the apex), 300 NFF-C-Tr (NFF of the caryopsis split transversally) which would make a potentially complex MNI given the many permutations. The implications for our interpretations are not small – changing the MNI could subsequently change basic stats of densities and proportions as well as more complex statistics down the line, which could alter interpretations on the role of this plant in our assemblage and the way humans and plants interacted in the past.

Taphonomy is a critical factor in our data – this is well established in discussions around all aspects of archaeobotanical interpretation from crop processing models to foodways and much in between and beyond. And we are asking for more taphonomy interrogation in the formation of the data used in making these interpretations. We provide this last thought to round out our argument: how far should we split or lump our data? Is it all okay so long as it does not affect MNIs, or could it have implications for other interpretations down the line? Could there be hidden questions within the data behind the counting, like the presence of achenes split in specific ways linked to crop processing for example, or NFFs as having seed coat patterns that could refine the identifications, or economic crops breaking in ways linked to cooking or consumption that have not being incorporated into analysis?

We are not expecting people to adopt this system as that is not the goal of this paper, instead we are simply laying out our own discussions for the sake of our open data practices, to open up the debate we have been having between ourselves to the wider community and to show how we are striving for comparability across our own datasets. Beyond that though we hope that this does prompt discussion, with panels at conferences or workshops to discuss counting and quantification terminology, with the goal being to move towards greater reproducibility, comparability and comprehensive across assemblages. This is a beginning, and we welcome suggestions on where to take it next.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00334-023-00982-6.

Acknowledgements The authors are grateful for the support of their universities and departments (Department of Archaeology and Art History, Seoul National University; Department of Humanities, Universitat Pompeu Fabra). They are also grateful for the thoughts and advice of people in their labs that this has been bounced off as an idea, and especially Matthew Conte, Marco Madella and Carla Lancelotti. The paper was written up and funded by JB's project at SNU "Ashmounds: Southern Indian Neolithic monuments of memory (c. 3000–1200 BC)" and M Madella's project at CASEs, UPF "ModAgrO" of which CJA is part. CASEs is a quality research group of the Catalan Government. This work was supported by the New Faculty Startup Fund from Seoul National University, the Palarq Foundation and the FI-2019 grant to hire research staff in training from the Catalan Government.

Funding Open Access funding enabled and organized by Seoul National University.

References

Adams KR, Muenchrath DA, Schwindt DM (1999) Moisture effects on the morphology of ears, cobs and kernels of a South-western U.S. Maize (*Zea mays*L.) Cultivar, and implications for the interpretation of Archaeological Maize. J Archaeol Sci 26:483–496. https://doi.org/10.1006/jasc.1998.0320

Albert RM, Weiner S (2001) Study of phytoliths in prehistoric ash layers from Kebara and Tabun caves using a quantitative approach. In: Meunier JD, Colin F (eds) Phytoliths: applications in Earth sciences and Human History. Balkema Publishers, Lisse, pp 251–266

Antolín F, Buxó R (2011) Proposal for the systematic description and taphonomic study of carbonized cereal grain assemblages: a case study of an early neolithic funerary context in the cave of Can Sadurní (Begues, Barcelona province, Spain). Veget Hist Archaeobot 20:53–66. https://doi.org/10.1007/s00334-010-0255-1

Antolín F, Jacomet S (2015) Wild fruit use among early farmers in the neolithic (5400–2300 cal BC) in the north-east of the Iberian Peninsula: an intensive practice? Veget Hist Archaeobot 24:19–33. https://doi.org/10.1007/s00334-014-0483-x

Arranz-Otaegui A, Gonzalez Carretero L, Ramsey MN, Fuller DQ, Richter T (2018) Archaeobotanical evidence reveals the origins of bread 14,400 years ago in northeastern Jordan. Proc Natl Acad Sci USA 115:7925–7930. https://doi.org/10.1073/pnas.1801071115

Azorín MB, Antolín F (2014) A les avellanes, foc i flames: tafonomia i quantificació de les closques d'avellana recuperades en contextos arqueològics. Revisió Del registre documentat a la península Ibèrica. Cypsela: Rev Prehistòria Protohistòria 19(2012):281–294

Banning EB (2021) Sampled to death? The rise and fall of Probability Sampling in Archaeology. Am Antiq 86:43–60. https://doi.org/10.1017/aaq.2020.39

Barboni D, Bonnefille R, Alexandre A, Meunier JD (1999) Phytoliths as paleoenvironmental indicators, West Side Middle Awash Valley, Ethiopia. Palaeogeogr Palaeoclimatol Palaeoecol 152:87–100. https://doi.org/10.1016/S0031-0182(99)00045-0

Bates J (2016) Social Organisation and Change in Bronze Age South Asia: a multi-proxy approach to urbanisation, deurbanisation and village life through phytolith and macrobotanical analysis. PhD thesis, University of Cambridge, Cambridge

Bates J (2019a) The published (to date October 2017) Archaeobotanical Data from the Indus Civilisation, South Asia, c.3200-1500BC. https://doi.org/10.7910/dvn/wshmad

Bates J (2019b) Oilseeds, spices, fruits and flavour in the Indus Civilisation. J Archaeol Sci Rep 24:879–887. https://doi.org/10.1016/j.jasrep.2019.02.033

Bates J (2020) Kitchen gardens, wild forage and tree fruits: a hypothesis on the role of the *Zaid* season in the Indus Civilisation (c.3200 – 1300 BCE). Archaeol Res Asia 21:100175. https://doi.org/10.1016/j.ara.2019.100175

Bates J, Wilcox Black K, Morrison KD (2022) Millet bread and pulse dough from early Iron Age South India: charred food lumps as culinary indicators. J Archaeol Sci 137:105531. https://doi.org/10.1016/j.jas.2021.105531

Bennetzen J, Buckler E, Chandler V et al (2001) Genetic evidence and the origin of Maize. Lat Am Antiq 12:84–86. https://doi.org/10.2307/971759

Boardman S, Jones G (1990) Experiments on the effects of charring on cereal plant components. J Archaeol Sci 17:1–11

Colledge S (2004) Reappraisal of the archaeobotanical evidence for the emergence and dispersal of the 'founder crops'. In: Peltenberg E, Wasse A (eds) Neolithic Revolution: New perspectives on Southwest Asia in Light of recent discoveries on Cyprus. Oxbow Books, Oxford, pp 49–60

D'Alpoim Guedes J, Spengler R (2014) Sampling Strategies in Paleoethnobotanical Analysis. In: Marston JM, d'Alpoim Guedes J, Warinner C (eds) Method and Theory in Paleoethnobotany. University Press of Colorado, Boulder, pp 77–94. https://doi.org/10.5876/9781607323167.c005

Dabrowski V, Tengberg M, Creissen T, Rougeulle A (2018) Plant supplying strategies in an Islamic Omani Harbour City: Archaeobotanical Analysis from a workshop (B39) in Qalhāt (XIVth-XVIth c. AD). J Islam Archaeol 5:17–38. https://doi.org/10.1558/jia.37690

Decaix A, Mohaseb FA, Maziar S, Mashkour M, Tengberg M (2019) Subsistence economy in Kohneh Pasgah Tepesi (eastern Azerbaijan, Iran) during the late Chalcolithic and the early bronze age based on the faunal and botanical remains. In: Meyer J-W, Vila E, Mashkour M, Casanova M, Vallet R (eds) The Iranian Plateau during the bronze age. MOM Éditions, Lyon, pp 75–88. https://doi.org/10.4000/books.momeditions.7996

Desse J, Desse-Berset N, Henry A, Tengberg M, Besenval R (2008) Faune et flore des niveaux profonds de shahi-tump (Balochistan, Pakistan): Premiers résultats. Paléorient 34:159–171. https://doi.org/10.3406/paleo.2008.5237

Diehl MW (2017) Paleoethnobotanical Sampling Adequacy and Ubiquity: an Example from the American Southwest. Adv Archaeol Pract 5:196–205. https://doi.org/10.1017/aap.2017.5

Dunseth ZC, Fuks D, Langgut D et al (2019) Archaeobotanical proxies and archaeological interpretation: a comparative study of phytoliths, pollen and seeds in dung pellets and refuse deposits at early Islamic Shivta, Negev, Israel. Quat Sci Rev 211:166–185. https://doi.org/10.1016/j.quascirev.2019.03.010

Fuks D, Dunseth ZC (2021) Dung in the dumps: what we can learn from multi-proxy studies of archaeological dung pellets. Veget Hist Archaeobot 30:137–153. https://doi.org/10.1007/s00334-020-00806-x

Fuller DQ (2000) The Emergence of Agricultural Societies in South India: botanical and archaeological perspectives, PhD thesis edn. University of Cambridge, Cambridge

Fuller DQ, Harvey EL (2006) The archaeobotany of Indian pulses: identification, processing and evidence for cultivation.

Environ Archaeol 11:219–246. https://doi.org/10.1179/174963106x123232
García-Granero JJ (2015) From gathering to farming in semi-arid northern Gujarat (India): a multi-proxy approach. PhD thesis, University of Barcelona, Barcelona
González Carretero L, Wollstonecroft M, Fuller DQ (2017) A methodological approach to the study of archaeological cereal meals: a case study at Çatalhöyük East (Turkey). Veget Hist Archaeobot 26:415–432. https://doi.org/10.1007/s00334-017-0602-6
Heiss AG, Azorín MB, Antolín F et (2020) Mashes to Mashes, Crust to Crust. Presenting a novel microstructural marker for malting in the archaeological record. PLoS ONE 15:e0231696. https://doi.org/10.1371/journal.pone.0231696
Hillman GC, Mason S, de Moulins D, Nesbitt M (1996) Identification of archaeological remains of wheat, the 1992 London workshop. Circaea 12:195–209
Jacomet S (2006) Identification of cereal remains from archaeological sites, 2nd edn. IPAS, Basel
Jones G (1990) The application of present-day cereal processing studies to charred archaeobotanical remains. Circaea 6:91–96
Jupe M (2003) The effects of charring on pulses and implications for using size change to identify domestication in Eurasia (BA). Dissertation, University College London, London
Kubiak-Martens L, Brinkkemper O, Oudemans TFM (2015) What's for dinner? Processed food in the coastal area of the northern Netherlands in the late neolithic. Veget Hist Archaeobot 24:47–62. https://doi.org/10.1007/s00334-014-0485-8
Lennstrom HA, Hastorf CA (1992) Testing old wives' tales in Palaeoethnobotany: a comparison of bulk and scatter sampling schemes from Pancán, Peru. J Archaeol Sci 19:205–229. https://doi.org/10.1016/0305-4403(92)90050-D
Lennstrom HA, Hastorf CA (1995) Interpretation in Context: Sampling and Analysis in Paleoethnobotany. Am Antiq 60:701–721. https://doi.org/10.2307/282054
Lentfer CJ, Boyd WE (1998) A comparison of three methods for the extraction of Phytoliths from Sediments. J Archaeol Sci 25:1159–1183
Lodwick LA (2019) Agendas for Archaeobotany in the 21st Century: data, dissemination and new directions. Internet Archaeol 53. https://doi.org/10.11141/ia.53.7
Maass BL, Usongo MF (2007) Changes in seed characteristics during the domestication of the lablab bean (*Lablab purpureus* (L.) Sweet: Papilionoideae). Aust J Agric Res 58:9–19. https://doi.org/10.1071/AR05059
Madella M (1997) Phytolith Analysis from the Indus Valley Site of Kot Diji, Sindh, Pakistan. In: Sinclair A, Slater E, Gowlett J (eds) Archaeological Sciences 1995: Proceedings of a conference on the application of scientific techniques to the study of archaeology. Oxbow Books, Oxford, pp 294–302
Mangafa M, Kotsakis K, Stratis G (2001) The experimental charring of products of the grape vine and an investigation of its uses in antiquity. In: Bassiakos Y, Aloupi E, Facorellis Y (eds) Archaeometric issues in Greek Prehistory and Antiquity. The Hellenic Society of Archaeometry, Athens, pp 495–505
Margaritis E, Jones M (2006) Beyond cereals: crop processing and *Vitis vinifera* L. Ethnography, experiment and charred grape remains from Hellenistic Greece. J Archaeol Sci 33:784–805. https://doi.org/10.1016/j.jas.2005.10.021
Morrison KD, Sinopoli CM, Wilcox Black K et al (2022) From the Southern Neolithic to the Iron Age: a view from Kadebakele. In: Korisettar R (ed) Beyond stones and more stones, vol 3. The Mythic Society, Bengaluru, pp 146–213
Pearsall DM (1988) Interpreting the meaning of Macroremain abundance: the impact of source and context. In: Hastorf CA, Popper VS (eds) Current Paleoethnobotany: Analytical methods and Cultural interpretations of Archaeological Plant remains. University of Chicago Press, Chicago, pp 97–118
Piperno DR (1988) Phytolith Analysis, an archaeological and geological perspective. Academic Press, London
Piperno DR (2006) Phytoliths: a Comprehensive Guide for archaeologists and paleoecologists. AltaMira Press, Oxford
Pokharia AK, Srivastava C (2013) Current status of Archaeobotanical studies in Harappan civilization: an Archaeological Perspective. Herit: J Multidiscip Stud Archaeol 1:118–137
Pokharia AK, Agnihotri R, Sharma S, Bajpai S, Nath J, Kumaran RN, Negi BC (2017a) Altered cropping pattern and cultural continuation with declined prosperity following abrupt and extreme arid event at ~4,200 yrs BP: evidence from an Indus archaeological site Khirsara, Gujarat, western India. PLoS ONE 12:e0185684. https://doi.org/10.1371/journal.pone.0185684
Pokharia AK, Sharma S, Tripathi D, Mishra N, Pal JN, Vinay R, Srivastava A (2017b) Neolithic – early historic (2500–200 BC) plant use: the archaeobotany of Ganga Plain, India. Quat Int 443 Part B 223–237. https://doi.org/10.1016/j.quaint.2016.09.018
Purugganan MD, Fuller DQ (2011) Archaeological Data reveal slow rates of Evolution during Plant Domestication. Evolution 65:171–183. https://doi.org/10.1111/j.1558-5646.2010.01093.x
Reddy SN (1997) If the Threshing Floor could talk: integration of agriculture and pastoralism during the late Harappan in Gujarat, India. J Anthropol Archaeol 16:162–187
Reddy SN (2003) Discerning palates of the past: an ethnoarchaeological study of crop cultivation and plant usage in India. Ethnoarchaeological series 5. International Monographs in Prehistory, Ann Arbor
Schiffer MB (1983) Toward the identification of formation processes. Am Antiq 48:675–706. https://doi.org/10.2307/279771
Smith H, Jones G (1990) Experiments on the effects of charring on cultivated grape seeds. J Archaeol Sci 17:317–327. https://doi.org/10.1016/0305-4403(90)90026-2
Strömberg CAE (2009) Methodological Concerns for Analysis of Phytolith assemblages: does count size matter? Quat Int 193:124–140
Tengberg M (1999) Crop husbandry at Miri Qalat Makran, SW Pakistan (4000–2000 B.C). Veget Hist Archaeobot 8:3–12
Thompson GB (1996) Ethnographic models for Interpreting Rice remains. In: Higham C, Thosarat R (eds) The excavations at Khok Phanom Di, a prehistoric site in Central Thailand. The Society of Antiquaries of London, London, pp 119–150
Valamoti SM (2002) Food remains from bronze age-archondiko and Mesimeriani Toumba in northern Greece? Veget Hist Archaeobot 11:17–22. https://doi.org/10.1007/s003340200002
Valamoti SM, Charles M (2005) Distinguishing food from fodder through the study of charred plant remains: an experimental approach to dung-derived chaff. Veget Hist Archaeobot 14:528–533. https://doi.org/10.1007/s00334-005-0090-y
Van der Veen M (1984) Sampling for seeds. In: van Zeist W, Casparie WA (eds) Plants and ancient man. Studies in Palaeoethnobotany. Balkema Publishers, Rotterdam, pp 193–199
Willcox G (2004) Measuring grain size and identifying Near Eastern cereal domestication: evidence from the Euphrates valley. J Archaeol Sci 31:145–150. https://doi.org/10.1016/j.jas.2003.07.003
Wolff AC, Westbrook AS, DiTommaso A (2022) In the ruins: the neglected link between archaeology and weed science. Weed Sci 70:135–143. https://doi.org/10.1017/wsc.2022.11
Zurro D (2018) One, two, three phytoliths: assessing the minimum phytolith sum for archaeological studies. Archaeol Anthropol Sci 10. https://doi.org/10.1007/s12520-017-0479-4. 1,673–1,691