Morphometrics of Starch Granules from Sub-Saharan Plants and the Taxonomic Identification of Ancient Starch

Julio Mercader¹, Matthew Abtosway¹, Robert Bird¹, Mariam Bundala¹, Siobhan Clarke¹, Jamie Inwood¹, Makarius Itambu¹, Fergus Larter¹, Robert Patalano¹, Maria Soto¹, Laura Tucker¹, Dale Walde¹

¹ Stone Tools, Diet and Sociality Laboratory, Department of Anthropology and Archaeology, University of Calgary, 2500 University Drive, Calgary, Alberta, Canada

Abstract The assumption that taxonomy can be ascertained by starch granule shape and size has persisted unchallenged since the late nineteenth and early twentieth century biochemistry. More recent work has established that granule morphological affinity is scattered throughout phylogenetic branches, morphotype proportions vary within the genus, granules from closely related genera can differ dramatically in shape, and size variations do not reflect phylogenetic relationships. This situation is confounded by polymorphism at the species and tissue level, resulting in redundancy and multiplicity.

This paper classifies morphological features of starch granules from 77 species, 31 families, and 22 orders across three African ecoregions. This is the largest starch reference collection published to date, rendering the dataset uniquely well suited to explore i) the diagnostic power of unique morphometric classifiers and their frequency, ii) morphotypes that cut across taxonomic boundaries, and iii) issues surrounding the minimum counts needed to accurately reflect granule polymorphism, variability, and identification.

In a collection of 23,100 granules, taxonomic identification occurred very rarely. In the instances it did, it was at the species level, with no occurrences of a single morphotype or complement identifying all species within a family or genus. Some families cannot be uniquely identified, and morphometric types are shared despite taxonomic distance for three quarters of the taxa. However, this reference collection boasts 98 unique identifiers located in the Arecaceae, Convolvulaceae, Cyperaceae, Dioscoreaceae, Fabaceae, Musaceae, Pedaliaceae, Poaceae, and Zamiaceae.

Keywords Starch granule Morphometry; Reference Collection; Starch Identification; Ancient Starch Research; Sub-Saharan Ethnobotany; Unique Identifiers

Correspondence

Dr. Julio Mercader

julio.mercader@gmail.com

1. Introduction

For decades, scientists have relied on the comparative study of starch granules from reference collections to infer phylogenetics and species identification for both fresh and ancient materials (Cortella and Pochettino, 1995; Parr, 2002; Langejans, 2006; Lentfer, 2009; Yang et al., 2012; Musaubach et al., 2013; Hart, 2014; Louderback et al., 2016). The assumption that taxonomy can be ascertained by granule shape and size dates back to late nineteenth and early twentieth century biochemistry (Nageli et al., 1858; Meyer, 1895; Reichert, 1913) and has persisted unchallenged (Czaja, 1978; Jane et al., 1994). Yet, an examination of phylogenetic origin for the morphological diversity of starch granules has been absent until recently: Matsushima et al. (2013) established that granule morphological affinity is scattered throughout phylogenetic branches, morphotype proportions vary within the genus, granules from closely related genera can differ dramatically in shape, and size variations do not reflect phylogenetic relationships (Shapter et al., 2008). Moreover, polymorphism at the species and tissue level yields a marked redundancy and multiplicity (Parr, 2002; Langejans, 2006).

Taxonomic identification of ancient starch granules often underpins reconstructions of human palaeoecology and ethnobotany, but remains problematic (Copeland and Hardy, 2018). The procedure utilized by the vast majority of researchers is a visual, morphometric comparison under light microscopy with fresh granules from a reference collection. A very limited number of studies have tackled automated systems of identification (Torrence et al., 2004; Wilson et al., 2010; Coster et al., 2015; Arráiz et al., 2016). And although the degree of subjectivity and

successful classification rates from visual versus computerized identification are controversial, direct and automated comparisons both face three unsolved methodological problems:

- i) The unreliability of small reference collections
- ii) The frequency of unique morphometric classifiers in a pool of targets, and issues surrounding minimum counts needed to accurately reflect granule polymorphism and variability per taxon.
- iii) The existence of many morphotypes that cut across taxonomic boundaries.

Moreover, very few atlases or reference collections for starch identification have been published to date (Nageli et al., 1858; Meyer, 1895; Reichert, 1913; Czaja, 1978; Jane et al., 1994; Lindeboom et al., 2004). It is important for the discipline to publish these reference collections, so that researches working in this area can access the classification, morphometric analysis, and images of a range of materials that might be present in archaeological assemblages. This work contributes *circa* 4,300 images of starch granules from taxa commonly exploited by Indigenous societies in Sub-Saharan Africa. It is the largest starch reference collection published to date, classifying 23,100 granules and morphological features from 77 species, 31 families, and 22 orders across the angiosperms and in one basal gymnosperm lineage, presented in phylogenetic order from primitive to derived (Fig. 1). Selection followed ethnobotanical criteria in three African ecoregions: the Guineo-Congolian, Zambezian, and Somalia-Masai zones. This paper is also a methodological contribution that examines the accuracy of using starch granule morphotypes and morphometrics to identify plant of origin and the chances for retrieval of unique identifiers as true diagnostic types per species. Blind testing was not conducted, as the present study was aimed at establishing criteria for future work.

2. Materials and methods

The botanical samples originated from 13 Sub-Saharan countries that fall within three phytochoria (Table 1) (White, 1983), with about three-quarters of the analyzed specimens native to the Zambezian woodlands of Mozambique and northern South Africa. Due to the human ecology focus, species and materials were specifically selected for their ethnobotanical significance (van Wyk and Gericke, 2000; Laden and Wrangham, 2005; Maundu and Tengnas, 2005; Williamson, 2005; Dharani and Yenesew, 2010). As a result, almost half of the species fall within two families (Fabaceae and Dioscoreaceae), as they play a key role in Indigenous utilization of natural resources. Taxonomic grouping and nomenclature come from the index of accepted names for the flowering plants of Sub-Saharan Africa (Klopper et al., 2006).

The method used to observe native starches is the extraction by powdering and suspension in aqueous media for light microscopy. Microscopic observations were conducted with two apparatuses: Motic BA410E (Photomicrography, Moticam 5+ camera, MoticImagesPlus 3.0) and Olympus BX51 (Photomicrography, SC50 camera, cellSens Standard). Granules were measured to an accuracy of 0.01 μm (Supplementary Fig. 1). The discovery and elimination process needed to find unique identifiers required counts per taxon >300 to ensure that variability and proportions are within 5% of the true value (counts of 500 achieve 95% confidence: Thompson, 1992). We followed a geometric morphometric typology that includes 50 possible morphotypes (Table 2), which can be broken down in 11 major classes, 18 three-dimensional descriptors, and 49 two-dimensional outlines. Morphotype richness per taxon is in Table 3. Elimination isolated primary morphometric types, defined as 3D/2D

combinations and maximum length range. There are two kinds of unique identifiers: one combines morphology and size with volumetric and textural features; the other is restricted to only morphometry. The original quantitative matrix, recording all observations and cases as well as all photomicrographs taken for granule measures, is with the Federated Research Data Repository (doi:10.20383/101.017).

3. Family Morphometrics and Identifiers

3.1. Zamiaceae

Represented by trunk samples from five *Encephalartos* species, with a morphotype complement broken down in Supplementary Table 1. The relative morphotype proportions differ within the family, but nil or low counts prevent valid statistical tests of difference. *Encephalartos friderici-guilielmi* is uniquely identified by Globular–Irregulars with a centric hilum and marked lamellae (n=1), as well as Lenticular–Orbiculars (n=46). Morphometrically, Globular–Trapeziforms with a maximum length range of 41.00-46.86 μm (n=4) uniquely identify the family, while the species *Encephalartos laevifolius* is uniquely identified by Hemisphere–Orbiculars with a maximum length range of 36.74-38.10 μm (n=2).

3.2. Nymphaeaceae

Represented by the USO of a single species: *Nymphaea caerulea*. The complement has Cylindroid–Ellipsoidals (n=24), Globular–Orbiculars (n=265), Globular–Trapeziforms (n=7), and Hemisphere–Orbiculars (n=4). The Cylindroid–Ellipsoidals with centric hilum and porous surface (n=1) and the Cylindroid–Ellipsoidals with centric hilum and psilate texture, with maximum length between 19.05-19.58 μm (n=2), have diagnostic power. However, if the hilum visibility and textural characteristics were not preserved, the species would not have unique identifiers.

3.3. *Araceae*

Represented by USOs of two *Zantedeschia* species. Morphotypes in these taxa differ significantly in content and proportion for the species *aethiopica* and *rehmannii*. Overall, the complement is Globular–Orbicular [n=100/8 (Z.a./Z.r.)], Globular–Trapeziform (n=172/0), Hemisphere–Orbicular (n=27/0), Hemisphere–Trapeziform (1/0), and Prismatic–Polygonal (0/292). *Zantedeschia aethiopica* is uniquely identified by a Hemisphere–Trapeziform morphotype with centric hilum, psilate surface, and a faceted orbicular volume (n=1).

Zantedeschia rehmannii is uniquely identified by a Prismatic–Polygonal morphotype with centric hilum, psilate surface texture, and a maximum length range of 3.68-12.22 μm (n=292). If we only had morphometric information and no textural or additional volumetric data, *Zantedeschia aethiopica* is still uniquely identifiable by Hemisphere–Trapeziforms with a length of 9.48 μm (n=1), while *Zantedeschia rehmannii* is uniquely identifiable by Prismatic–Polygons with a maximum length range of 3.68-5.31 μm (n=24).

3.4. *Taccaceae*

Represented by the USO of *Tacca leontopetaloides*. The complement consists of Globular–Orbicular (n=209), Globular–Trapeziform (n=39), and Hemisphere–Orbicular (n=52). It is uniquely identified by Hemisphere–Orbiculars with centric hilum, creased surface texture, and faceting, with a maximum length range of 12.09-25.81 μm . If degradation was to delete textural and volumetric characteristics, then there would not be unique identifiers for this species.

3.5. *Dioscoreaceae*

Represented by USOs from fourteen *Dioscorea* species with variable complements and proportions (Table 4), although most assemblages are dominated by Parabolic Prism–Prisms [cf. *Dioscorea cayenensis* (n=6), *D. diversifolia*, *D. dumetorum*, *D. rupicola*, *D. smilacifolia*, and *D. sylvatica* (n=0)]. This very large sample produced many unique identifier morphotypes for either individual or multiple species. When only morphometrics are considered, the Dioscoreaceae are uniquely identified by Conoid–Triangular [maximum length $>26.00 \mu\text{m}$ (n=11)], Lobate–Shell [maximum length 35.71-47.05 μm (n=2)], Parabolic Prism–Irregular (n=39), Parabolic Prism–Oblong Elongate (n=22), Parabolic Prism–Parabolic–Parabolic [maximum length $<14.90 \mu\text{m}$ (n=22) $>53.9 \mu\text{m}$ (n=195)], Prism–Shell (n=89), Parabolic Prism–Trapeziform (n=3), and Parabolic Prism–Triangular (n=303).

3.6. *Pandanaceae*

Represented by specimens from the trunk of *Pandanus livingstonianus*. Morphotypes include Globular–Orbicular (n=289), and Globular–Trapeziform (n=11). All granules have centric hilum and psilate texture. Faceting is common. There are no unique morphometric identifiers within this species.

3.7. *Orchidaceae*

Represented by granules from the USO of *Disa spp.* with types that include Globular–Orbicular (n=272), Globular–Trapeziform (n=16), and Hemisphere–Orbicular (n=12). All granules have centric hilum and psilate texture, with some faceting (n=214). There are no unique morphometric identifiers within this species.

3.8. *Iridaceae*

Represented by USOs from three *Gladiolus* species. The complement proportions amongst the species (Supplementary Table 2) are significantly different (Chi–square=201.623, df=6, p=.000), likely to result from the proportions seen in *G. undulatus*. However, this significance remains when *G. undulatus* is removed, confirming significant differences in the complement (Chi–square=12.815, df=2, p=.002). *G. undulatus* is uniquely identified by

Hemisphere–Orbiculars with a centric hilum, creased surface texture, and with maximum length 34.00-36.00 μm (n=2). This species is also uniquely identified by Hemisphere–Orbiculars with centric hilum, psilate surface, centric slit (-), and maximum length 26.82-26.92 μm (n=2). *G. spp.* is uniquely identified by Hemisphere–Orbiculars, centric hilum, psilate surface, and centric slit (v) (n=35). Removing volumetric and textural features with low taxonomic value would imply that there are no unique identifiers within this family.

3.9. *Hypoxidaceae*

Represented by USOs of three *Hypoxis* species (Supplementary Table 3). Most granules share features with other families: Globular–Orbicular with centric hilum, psilate texture, faceted/non–faceted; Globular–Orbicular with centric hilum, faint lamellae, faceted/non–faceted, and; Cylindroid-Ellipsoidal type with centric hilum, psilate texture, though sometimes with creasing or faint lamellae, and weak birefringence. Morphotype proportions in *H. nyasica* differ from those in *H. rigidula* and *H. zeihari* (Chi–square=56.379, $p=.000$, $df=2$), but these two latter species do not differ significantly between themselves (Chi–square =.293, $p=.394$, $df=1$). If morphometry is considered exclusively, then there are no unique identifiers within this family.

10. . *Asphodelaceae*

Represented by the USO of *Bulbine sedifolia*. All granules belong to a widely shared identifier [Cylindroid–Ellipsoidal (eccentric hilum, faint lamellae)] and maximum length overlaps with many others as well, resulting in no unique morphometric identifiers.

11. *Amaryllidaceae*

Represented by the USO of *Amaryllis belladonna*. Its complement includes Cylindroid–Oblong Elongates (n=80), Cylindroid–Ellipsoids (n=53), Globular–Orbiculars (n=3), Parabolic Prism–Parabolics (n=2), Pear Shaped–Irregulars (n=8), and Pear Shaped–Ovates (n=154). No other taxon shares this complement, and the taxon is uniquely identified by Parabolic Prism–Parabolics with eccentric hilum, marked lamellae, and maximum length 52.65–53.89 μm (n=2), as well as by Pear Shaped–Irregulars with eccentric hilum and marked lamellae (n=8). From a purely morphometric point of view, the species is uniquely identified by the presence of Cylindroid–Oblong Elongates with maximum length 60.80–89.70 μm (n=42).

12. *Arecaceae*

Represented by a trunk sample from *Hyphaene petersiana*. The complement includes Conoid–Bells (n=5), Conoid–Cuneiforms (n=27), Conoid–Oblong Elongates (n=2), Conoid–Orbiculars (n=3), Cylindroid–Oblong Elongates (n=2), Cylindroid–Ellipsoids (n=119), Globular–Orbiculars (n=81), Pear Shaped–Irregulars (n=1), and Pear Shaped–Ovates (n=60). No other taxa share these morphotypes in content or proportions. The species has unique identifiers

such as Conoid–Bells with centric hilum and faint lamellae (n=5), Conoid–Cuneiforms with centric hilum (n=27), Conoid–Oblong Elongate with centric hilum (n=2), Conoid–Orbicular with centric hilum and faint lamellae (n=3), Pear Shaped–Irregulars with centric hilum (n=1), and Pear Shaped–Ovates with centric hilum and creased surface (n=1). If we consider morphometrics only, *Hyphaene petersiana* is uniquely identified by Conoid–Bell with maximum length 16.41–19.23 μm (n=5), Conoid–Cuneiforms with maximum length <13.00 μm (n=27), Conoid–Oblong Elongates with maximum length 11.03–15.58 μm (n=2), and Conoid–Orbiculars with maximum length 19.83–23.86 μm (n=3).

13. *Typhaceae*

Represented by the USO of *Typha capensis*. The complement consists of Cyliindroid–Ellipsoidal (n=21), Globular–Orbicular (n=264), Hemisphere–Orbicular (n=2), and Pear Shaped–Ovate (n=13). There are no unique morphometric identifiers within this species.

14. *Poaceae*

Represented by specimens from *Eleusine africana* (grain), *Oxytenanthera abyssinica* (USO), *Sorghum bicolor* (grain), and *Sporobolus panicoides* (grain). Their complements are detailed in Supplementary Table 4. *Sorghum bicolor* is uniquely identified by Prismatic–Polygons with centric hilum and creasing (n=3) or granulation (n=2), and Prismatic–Polygons with centric hilum and psilate texture with maximum length 12.25–30.34 μm

(n=295). *Eleusine africana* is uniquely identified by Prismatic–Polygons with centric hilum and faint lamellae (n=255). If morphometric criteria are considered, *Eleusine africana* and *Sorghum bicolor* are mutually identified by Prismatic–Polygons with a maximum length 12.25-18.17 μm (n=121), while individually, *Sorghum bicolor* is uniquely identified by Prismatic–Polygons with maximum length 18.26-30.34 μm (n=108).

15. Cyperaceae

Represented by the USOs of three *Cyperus* species: *C. esculentus*, *C. rotundus*, and *C. tomaiophyllus*. They demonstrate dramatic differences in complements (Supplementary Table 5). *C. rotundus* and *C. tomaiophyllus* are uniquely identified by Lobate–Irregular granules with centric hilum and psilate texture [n=207/11 (C.e./C.t.)]. *C. esculentus* is uniquely identified by Lobate–Irregular granules with centric hilum and creased surface (n=16), Pear Shaped–Ovates with centric hilum and creased surface with maximum length 6.45-22.79 μm (n=214), and Pear Shaped–Ovates with centric hilum location and pitted surfaces (n=1). *C. rotundus* is uniquely identified by Pear Shaped–Ovates with centric hilum and psilate texture (n=14). *C. tomaiophyllus* is uniquely identified by Globular–Irregulars with centric hilum and psilate texture (n=6) and by Lobate–Shells with centric hilum and psilate texture with maximum length 6.45-22.79 μm (n=6). If we look at morphometrics only hoping to discriminate among unknown granules, only *C. tomaiophyllus* could be uniquely identified [Lobate–Shells with max. size range 14.45-21.65 μm (n=6)].

16. Strelitziaceae

Represented by fruit mesocarp from *Strelitzia nicolai* with a granule complement consisting of Conoid–Fusiforms (n=227), Conoid–Irregulars (n=26), Conoid–Orbiculars (n=14), and Conoid–Ovates (n=33). Uniquely identified by Conoid–Fusiforms (n=227), Conoid–Irregulars with centric hilum and psilate texture (n=26), Conoid–Orbiculars with centric hilum and psilate texture (n=14), and Conoid–Ovates with centric hilum and psilate texture (n=33). Morphometrically, this species is uniquely identified by Conoid–Fusiforms with a maximum length 7.61-30.32 (n=227), Conoid–Irregulars with maximum length 11.64-18.85 μm (n=18), Conoid–Orbiculars with maximum length 8.12-13.62 μm (n=14), and Conoid–Ovates with a maximum length 10.12-20.90 μm (n=33).

17. *Musaceae*

Represented by the USO of *Ensete ventricosum*. The starch granule complement comprises Conoid–Bells (n=2), Conoid–Irregulars (n=77), Conoid–Oblong Elongates (n=42), Conoid–Ovates (n=6), Cylindroid–Oblong Elongates (n=3), Cylindroid–Ellipsoidals (n=10), Cylindroid–Irregulars (n=34), Irregular–Irregulars (n=52), Lobate–Irregulars (n=57), Pear Shaped–Irregulars (n=4), Pear Shaped–Shells (n=29), and Pear Shaped–Ovates (n=4). It is uniquely identified by Pear Shaped–Irregulars with hyper-eccentric hilum (n=4), Pear Shaped–Shells with hyper-eccentric hilum and marked lamellae (n=4), and Pear Shaped–Ovates with hyper-eccentric hilum (n=29). If we look at morphometrics only, unique identifiers include Conoid–Bells with maximum length 33.35-38.01 μm (n=2), Conoid–Irregulars with maximum

length 30.08-82.36 μm (n=77), Conoid–Oblong Elongates with maximum length 31.47-90.42 μm (n=42), Conoid–Ovates with maximum length 35.82-48.43 μm (n=5), Cylindroid–Irregulars with maximum length 51.78-67.74 μm (n=7), Irregular–Irregulars with maximum length 38.93-93.33 μm (n=45), Lobate–Irregulars with maximum length 50.43-82.32 μm (n=15), and Pear Shaped–Shells with maximum length 31.49-53.46 μm (n=21).

18. *Menispermaceae*

Represented by the USO of *Jateorhiza palmata*. The complement includes Conoid–Ovates (n=1), Cylindroid–Oblong Elongates (n=36), Cylindroid–Ellipsoidals (n=186), Cylindroid–Irregulars (n=1), Globular–Orbiculars (n=15), Pear Shaped–Irregulars (n=4), and Pear Shaped–Ovates (n=57). It is uniquely identified by Conoid–Ovates with eccentric hilum and psilate texture (n=1), Pear Shaped–Irregulars with eccentric hilum and psilate textures (n=4), and Pear Shaped–Ovates with eccentric hilum and psilate texture (n=57), although if we restrict parameters to morphometric identifiers, there would be none within the species.

19. *Proteaceae*

Represented by the USO of *Protea welwitschii*. The two morphotypes present, Globular–Orbicular (n=274) and Hemisphere–Orbicular (n=26), are ubiquitous across families. There are no unique morphometric identifiers within this species.

20. *Euphorbiaceae*

Represented by seeds from two *Ricinodendron* species: *heudelotii* and *rautanenii*. The complement includes Globular–Orbiculars [n=271/169 (R.h./R.r.)], Globular–Trapeziforms (n=10/13), and Hemisphere–Orbiculars (n=19/118). While the morphotype complement is consistent within the Euphorbiaceae, as represented by these two species, the proportions are not. *R. heudelotii* is uniquely identified by Hemisphere–Orbiculars with granulate texture (n=19) while *R. rautanenii* is uniquely identified by Hemisphere–Orbiculars with psilate textures (n=118). Unique identifiers within this family would be none if we restricted criteria to morphometry alone.

21. *Fabaceae*

Represented by the legume or USO of sixteen species (Table 5). Granules can be identified using fifteen unique morphotypes combined with hilum location and surface texture. From a purely morphometric standpoint, *Eriosema preptum* (USO) is uniquely identified by Conoid–Triangulars with maximum length 16.13-21.65 μm (n=7) as well as by Prismatic–Irregulars with maximum length 18.11-25.20 μm (n=2), and Prismatic–Quadratics with maximum length 20.77-38.64 μm (n=80). *Cassia petersiana* (legume) is uniquely identified by Cylindroid–Ellipsoidals maximum length 3.48-4.90 μm (n=31). *Cajanus cajan* (legume) is uniquely identified by Cylindroid–Ovates with maximum length 41.90-41.99 μm (n=2). *Lablab*

spp. (legume) and *Vigna lobatifolia* (legume) are jointly identified by Reniform–Reniforms with maximum length 20.77–38.84 μm (n=14).

22. *Cannabaceae*

Represented by the fruit of *Celtis africana*. The granule assemblage is dominated by Globular–Orbicular granules (n=299) with granulate texture. The single remaining granule is a Cylindroid–Ellipsoidal with granulate texture. These morphotypes are ubiquitous across families and length ranges, resulting in no unique morphometric identifiers.

23. *Cucurbitaceae*

Represented by the seeds of *Acanthosicyos naudinianus* and *Momordica cardiospermoides*. The morphotype complements and proportions of the two species differ considerably: Cylindroid–Oblong Elongate [n=0/15 (A.n./M.c.)], Cylindroid–Ellipsoidal (n=3/171), Globular–Orbicular (n=289/96), Prismatic–Quadratic (n=8/0), and Pear Shaped–Ovate (n=0/18). *Acanthosicyos naudinianus* is uniquely identified by Prismatic–Quadratics with psilate textures (n=8). *Momordica cardiospermoides* is uniquely identified by Cylindroid–Ellipsoidals with eccentric hilum and faint lamellae (n=4). From a morphometric point of view, *Acanthosicyos naudinianus* is uniquely identified by Prismatic–Quadratics with maximum length 4.72–8.49 μm (n=8).

24. *Combretaceae*

Represented by the legume of *Combretum spp.* The granule assemblage is dominated by Cylindroid–Ellipsoidals (n=207) with declining representation of Globular–Orbiculars (n=62) and Pear Shaped–Ovates (n=31). Morphologically it is uniquely identified by Pear Shaped–Ovates with granulate texture (n=31), but there are no unique morphometric identifiers.

25. *Bombacaceae*

Represented by the fruit mesocarp of *Adansonia digitata*. All granules belong to a single morphotype: Globular–Orbicular with centric hilum, psilate texture, sometimes faceted. This is the single most widely distributed morphotype in the collection and several species from different families share the full complement of Globular–Orbicular granules. Additionally, the maximum length overlaps completely with several families noted here, resulting in no unique morphometric identifiers.

26. *Polygonaceae*

Represented by the USO of *Polygonum senegalense*. A single 3D-2D morphotype is present in the Polygonaceae sample: Globular–Orbicular with centric hilum and psilate texture. These identifiers are nearly ubiquitous and maximum length overlap amongst the various families, preventing this species from being uniquely identified morphometrically.

27. *Sapotaceae*

Represented by the fruit of *Englerophytum magalismontanum*. The morphotypes include Globular–Orbiculars (n=34), Globular–Trapeziforms (n=128), and Hemisphere–Orbiculars (n=138). All granules have centric hilum and psilate texture, with 266 faceted cases that include Orbiculars and Globular–Trapeziforms. These morphotypes and length ranges are ubiquitous across families and cannot be used in unique morphometric identification.

28. *Convolvulaceae*

Represented by the USO of *Ipomoea longituba*. All granules belong to a single morphotype: Globular–Orbiculars with psilate texture and equatorial groove. This morphotype is a unique identifier for this species; however, there are no unique morphometric identifiers within this species.

29. *Apocynaceae*

Represented by the USOs of *Ceropegia rendalii* and *Fockea angustifolia*. Their granule morphotype complement differ significantly in content and proportion to include Pear Shaped–Ovates [n=0/89 (C.r./F.a.)], Globular–Orbiculars (n=262/192), Globular–Trapeziforms (n=32/7), Hemisphere–Orbiculars (n=6/10), and Pear Shaped–Shells (n=0/2). *Fockea angustifolia* is

uniquely identified by Hemisphere–Orbiculars with centric hilum and marked lamellae (n=8) and by Pear Shaped–Shells with centric hilum (n=2). There are no unique morphometric identifiers within this family.

30. *Pedaliaceae*

Represented by the seeds of *Sesamum angolense* yielding a unique identifier: Lobate–Irregulars with psilate texture. This species is not morphometrically unique.

31. *Lamiaceae*

Represented by the USOs of *Plectranthus esculentus* and *Solenostemon rotundifolius*. The morphotype complement is noted in Supplementary Table 6. The complement proportions in the two species differ significantly (Chi–square=68.328, df=2, p=.000). *Plectranthus esculentus* is uniquely identified by Hemisphere–Orbiculars with centric hilum, psilate texture, facets and bifid pleating (n=74). There are no unique morphometric identifiers within this family, however, if we were to include morphology and size only.

4. Discussion

A recent overview of 176 publications in archaeological starch from the last two decades shows that one of the most pressing needs in the field is to establish reference collections that

evaluate the reliability of taxonomic identification and highlight diagnostic morphotypes (Barton and Torrence, 2015). The vast majority of ancient starch attributions to plant of origin are still conducted using starch granule morphotypes and morphometrics directly recognized and classified under the microscope, such as we did here. Only four publications to date have explored computerized manipulation, and they present divergent results and degrees of success (Torrence et al., 2004; Wilson et al., 2010; Coster and Field, 2015; Arráiz et al., 2016). A persistent problem with both visual and automated classification of reference collections is small size, often utilizing <30 taxa. Paradoxically, as the pool of target species grows, so do difficulties in isolating truly diagnostic types and therefore mistaken identifications, making even automated classification unsatisfactory (Torrence et al., 2004; Wilson et al., 2010; Arráiz et al., 2016).

In our analysis of the data, ten of the 31 families cannot be identified (Table 6) by combining morphometrics with volumetric and/or textural features. Using morphometrics exclusively results in the removal of an additional ten families. Morphotypes are shared despite taxonomic distance for 67.7% of the taxa. This cohort of confuser granules encompasses 37 morphotypes (Supplementary Table 7). Conversely, this reference collection boasts 98 unique identifiers, most of which concentrate in five families (Dioscoreaceae, Fabaceae, Musaceae, Cyperaceae, and Arecaceae) (Fig. 2a). The relationship between total number of granules per family and the appearance of unique identifiers is not significant ($r^2=.028$, $p=.371$). Even if we were to accept outlier bias and include families represented by a large number of species, the linear relationship between them and the identification of unique identifiers would be only weakly positive ($r^2=.629$, $p=.000$). Two of the families with the fewest species (Convolvulaceae, Pedaliaceae) have the highest percentage of identifier granules (Fig. 2b). The Fabaceae, on the

other hand, has the highest granule count (n=4,800) but only 5.3% are unique to the family.

Comparing species with similar granule counts, the Poaceae (n=1,200) had 46.3% unique granules, while the Zamiaceae (n=1,500) had three.

Reniforms are extremely rare, only 0.3% of all Fabaceae and 0.06% of all granules.

Reniforms in the Fabaceae consist of two examples from *Lablab spp.* and 12 from *Vigna*

lobatifolia. Planoconvex Lenticulars were recovered only amongst the Zamiaceae (n=46),

specifically from the trunk of *Encephalartos friderici-guilielmi*. The other four Zamiaceae did

not yield Lenticular granules. Lenticulars are 3% of all Zamiaceae and 0.2% of the full

collection. Irregulars constitute another rare morphotype in the collection (n=53, 0.2%). Fifty-

two of these (98.1%) were recovered from the Musaceae (17.3% of all Musaceae granules) and

one (1.9%) from a Fabaceae (a *Cajanus cajan* legume). Parabolic Prisms occur frequently

(n=2,628, 11.3%), and are overwhelmingly associated with the Dioscoreaceae (99.1%), although

0.08% were recovered from *Amaryllis belladonna* (n=2) and 0.8% from the Fabaceae (n=22, 21

from an *Eriosema preptum* USO and 1 from a *Vigna lobatifolia* legume). Parabolic Prisms

represent 62% of all Dioscoreaceae granules, 0.6% of all Amaryllidaceae granules, and 0.4% of

all Fabaceae granules. Parabolic Prisms with distal bifid shape are unique to eight of the fifteen

Dioscorea. Parabolic Prisms vary in maximum length 8.49-83.81 μm . In the present sample, all

Parabolic Prisms 26.40-52.00 μm in length are from *Dioscorea* species. Amaryllidaceae

specimens are >52 μm in length and all 22 Fabaceae granules are <26.40 μm .

We examined size distribution and found that maximum lengths overlap ubiquitously.

When creating a subsample with the upper ten percent sizes only (12 families, n=2,310), no

patterning or exclusive identificatory characteristics could be observed. Subsampling for the

largest granules was detrimental to identification, as it eliminated 19 families and 40 species from consideration and blocked previously recognized unique identifiers (10 of 22). The procedure also affected sample size and morphotype proportions from each family differentially.

Distributions of maximum lengths amongst the 77 species were examined for normality using the SPSS Statistics 25 Explore and Descriptives modules. Where distributions were found not to be normal, the data were transformed using $\log_{10}$ and/or square root calculations and re-examined for normality of distribution (Supplementary Table 8; Supplementary Fig. 2). The results indicate that 36% of the taxa follow a non-normal distribution (Supplementary Table 9), which demonstrates that using mean or median maximum length as an identificatory feature is an invalid way to describe central tendency in a granule population; at least not without previous confirmation that a given taxon actually yields a normal size distribution (Supplementary Table 10).

5. Conclusions

The size of a reference collection is key to ascertaining the uniqueness of a morphotype, which must have characteristic shapes and sizes to allow for solid identification. In a collection consisting of 23,100 granules, firm taxonomic identification occurred very rarely (7%), and when it did, it was at the species level, with no instances in which a morphotype identifies all species within a family or genus. We found no evidence that starch granule morphology, in isolation, is able to reveal plant part, species, family or order unambiguously: There are no

diagnostic 3-D or 2-D shapes, and granule length alone is not a useful variable in identification, but rather must be assessed in combination with morphology.

Positively, some species and families produce more morphometrically unique granules than others do. Therefore, the identification of starch granules is possible, but not guaranteed, depending on the type of plant that was exploited in the past and whether the specific taxa and ecoregion contain unique identifiers. If early humans utilized plants that bear starch granules without demonstrable diagnostic morphometrics, then identification will be improbable to impossible. The problem isolated here is not with the type of inspection conducted, visual or automated, rather it is with the optimism often implied in the use of ‘diagnostic’ criteria to identify archaeological starch granules to the species level, as the criteria lose uniqueness when using large datasets, and reflect an overwhelming morphometric overlap between many species, phylogenetically related or not.

6. Acknowledgments

This work was fully sponsored by the Canadian Social Sciences and Humanities Research Council under its Partnership Grant Program no. 895-2016-1017.

7. Conflict of interest statement

The work submitted to this journal was not conducted with the involvement of any personal, professional or financial relationships that could be construed as a conflict of interest.

References

Arráiz H, Barbarin N, Pasturel M, Beaufort L, Domínguez-Rodrigo M, Barboni D (2016)

Starch granules identification and automatic classification based on an extended set of morphometric and optical measurements. *J Archaeol Sci Rep* 7:169-179

Cortella R, Pochettino M (1995) Comparative Morphology of Starch of Three Andean Tubers.

Starch-Starke 47:455-461

Coster A C F, Field J H (2015) What starch grain is that? – A geometric morphometric approach

to determining plant species origin. J Archaeol Sci 58:9-25

Czaja A (1978) Structure of Starch Grains and the Classification of Vascular Plant Families.

Taxon 27:463-470

Dharani N, Yenesew A (2010) Medicinal Plants of East Africa: An Illustrated Guide. Najma

Dharani, Kenya

Hart T (2014) Analysis of Starch Grains Produced in Select Taxa Encountered in Southwest Asia.

Ethnobotany Letters 5:135-145

Jane J L, Kasemsuwan T, Leas S, Zobel H, Robyt JF (1994) Anthology of Starch Granule

Morphology by Scanning Electron Microscopy. Starch-Starke 46:121-129

Klopper R R, Chatelain C, Habashi C, Gautier L, Spichiger R E (2006) Checklist of the

Flowering Plants of Sub-Saharan Africa: An Index of Accepted Names and Synonyms.

Sabonet, Pretoria

Laden G, Wrangham R (2005) The rise of the hominids as an adaptive shift in fallback foods:

Plant underground storage organs (USOs) and australopith origins. J Hum Evol

49:482-498

Langejans GHJ (2006) Starch grain analysis on Late Iron Age grindstones from South Africa.

South Afr Humanit 18:71-91

- Lentfer CJ (2009) Building a Comparative Starch Reference Collection for Indonesia and its Application to Palaeoenvironmental and Archaeological Research. *Terra Australis* 30:80-101
- Louderback LA, Herzog NM, Pavlik BM (2016) A new approach for identifying starch granules of wild food plants from arid western North America. *Starch-Starke* 68:1-7
- Maundu P, Tengnas B (2005) Useful Trees and Shrubs for Kenya. World Agroforestry Centre, Kenya
- Matsushima R, Yamashita J, Kariyama S, Enomoto T, Sakamoto W (2013) A Phylogenetic Re-evaluation of Morphological Variations of Starch Grains among Poaceae Species. *J Appl Glycosci* 60:37-44
- Mercader J, Akeju T, Brown M, Bundala M, Collins MJ, Copeland L, Crowther A, Dunfield P, Henry A, Inwood J, Itambu M, Kimb J-J, Larter, S, Longo L, Oldenburg T, Patalano R, Sammynaiken R, Soto M, Tyler R, Xhauflair H (2018) Exaggerated expectations in ancient starch research and the need for new taphonomic and authenticity criteria. <https://doi/10.17605/OSF.IO/S3CM4>
- Meyer A (1895) Untersuchungen Über die Stärkekorner. Fisher, Jena
- Musaubach M, Plos A, Bobot P (2013) Differentiation of archaeological maize (*Zea mays* L.) from native wild grasses based on starch grain morphology. Cases from the Central Pampas of Argentina. *J Archaeol Sci* 40:1186-1193
- Nageli KW, Cramer KE, Wartmann B (1858) Die Stärkekörner: Morphologische, Physiologische, Chemisch-Physicalische und Systematisch-Botanische. F. Schulthess, Zurich

- Parr J (2002) The Identification of Xanthorrhoea Resins by Starch Morphology: Prospects for Archaeological and Taxonomic Applications. *Econ Bot* 56:260-270
- Reichert E.T (1913) The Differentiation and Specificity of Starches in Relation to Genera, Species, Etc. Carnegie Institution of Washington, Washington
- Shapter FM, Henry RJ, Lee LS (2008) Endosperm and starch granule morphology in wild cereal relatives. *Plant Genetic Resources* 6:85-97
- Thompson SK (1992) Sampling. John Wiley & Sons, New York
- Torrence R, Wright R, Conway R (2004) Identification of starch granules using image analysis and multivariate techniques. *J Archaeol Sci* 31:519-532
- Van Wyk B, Gericke N (2000) People's Plants: A Guide to Useful Plants of Southern Africa. Briza Publications, Pretoria
- White F (1983) The Vegetation of Africa: A Descriptive Memoir to Accompany the UNESCO/AETFAT/UNSO Vegetation Map of Africa. UNESCO, Paris
- Williamson J (2005) Useful Plants of Malawi. Momfort Press, Limbe
- Wilson J, Hardy K, Allen R, Copeland L, Wrangham R, Collins M (2010) Automated classification of starch granules using supervised pattern recognition of morphological properties. *J Archaeol Sci* 37:594-604
- Yang X, Zhang J, Perry K, Ma Z, Wan, Li M, Diao X, Lu H (2012) From the modern to the archaeological: starch grains from millets and their wild relatives in China. *J Archaeol Sci* 39:247-254

Captions to Figures in Text

Figure 1: Family phylogenetic grouping of studied samples classified from derived to primitive (top to bottom)

Figure 2: a) Number of morphometric identifiers per family, b) Percent of granules displaying unique morphology from the total count per family

Captions to Tables in Text

Table 1: Sample provenance

Table 2: Starch granule typology

Table 3: Lead morphotypes and richness per species

Table 4: Dioscoreaceae morphotypes

Table 5: Fabaceae morphotypes

Table 6: Presence/absence of unique identifiers per family