Acta Palaeobot. 35 (1): 167–173, 1995

ATTEMPTS TO IDENTIFY 8000-BP SORGHUM USING IMAGE-ANALYSIS, INFRARED SPECTROSCOPY, AND BIOTECHNOLOGICAL PROCEDURES*

JEFFERY A. DAHLBERG [1], JOHN EVANS [2], ELIZABETH JOHNSON [1], MARIAN HYMAN [3], ED BIEHL [4], and FRED WENDORF [5]

[1] USDA-ARS-TARS, P. O. Box 70, Mayagüez, Puerto Rico 00681–0070
[2] Polytechnic of East London, Romford Road, London, E154LZ, United Kingdom
[3] Department of Chemistry, Texas A & M University, College Station, Texas 77843 USA
[4] Department of Chemistry, Southern Methodist University, Dallas, Texas 75275 USA
[5] Department of Anthropology, Southern Methodist University, Dallas, Texas 75275 USA

ABSTRACT: Sorghum (*Sorghum bicolor* (L.) Moench) with consistent radiocarbon dates of 8000 yrs bp have been excavated at an early Holocene archaeological site (E-75-6) at Nabta Playa near the Egyptian-Sudanese border. The objective of this research effort is to classify these sorghums within the known wild or domesticated races or working groups through the use of infrared spectroscopy, and biotechnological and image-analysis procedures. Preliminary results are discussed and evaluated. While intriguing, all these results must be regarded as highly speculative until a more detailed study of modern sorghum is completed. Of the three methods described, image-analysis coupled with infrared spectroscopy may offer the best hope of identifying the material in a non-destructive manner.

KEY WORDS: sorghum, Nabta Playa, 8000-bp, infrared spectroscopy, image-analysis, DNA isolation, biotechnology, archaeology

INTRODUCTION

Sorghum (*Sorghum bicolor* (L.) Moench) is the fifth most important cereal crop in the world after wheat (*Triticum* sp.), rice (*Oryza sativa* L., *O. glaberrima*), maize (*Zea mays* L.), and barley (*Hordeum vulgare*). In 1991, it ranked fifth among cereal crops in total area planted and metric tons harvested (FAO 1992) with 80% of the area devoted to sorghum lying within Africa and Asia, where average yields were 718 and 1023 kg ha^{-1}, respectively. These figures are 5.5 and 3.8 times lower than the average yield within the United States (FAO 1992). Production areas have remained constant within Asia and Africa but have decreased within the Americas in the last 10 years.

In the United States, sorghum grain is used primarily for livestock feed and stems and foliage for green chop, hay, silage, and pasture (House, 1985). In Africa and India, it is an important part of the diet in the form of unleavened bread, boiled porridge or gruel, and specialty foods such as popped grain and beer. Syrup is made from sweet sorghum (Doggett 1988, House 1985).

Mann et al. (1983) provide an excellent review of current hypotheses dealing with the origin and domestication of sorghum. In their review, they summarized that the origin and early domestication of sorghum took place in northeastern Africa north of the Equator and east of 10° E lat. approximately 5000 yrs ago. New evidence, however, may place its origin and domestication at 8000 yrs bp, 3000 years earlier than previously thought and 10–15° lat. further north than had been previously reported. Wendorf et al. (1992) have reported that carbonized seeds of sorghum with consistent radiocarbon dates of 8000 yrs bp have been excavated at an early Holocene archaeological site (E-75-6) at Nabta Playa near the Egyptian-Sudanese border. This new evidence may change our concepts of the origins of not only sorghum but of agriculture in general in Africa. It has been argued that the peoples of Afri-

* Presented at Archaeobotany of North Africa: I International Workshop Mogilany, May 27–31, 1994

ca had no agriculture until it was introduced from Asia Minor about 7000 bp (Mann et al. 1983). Wendorf's evidence is 1000 yrs older than that and indicates that the specimens found, though morphologically close to wild taxa, show some resemblance to domesticated sorghum with respect to the lipid fraction of the grains (Wasylikowa et al. 1993). Their findings should have a significant impact on our thinking of when and where sorghum was domesticated. The research objectives within should shed some light on the controversy surrounding the theory that the first truly domestic sorghum came from the Durras of India. According to this hypothesis, the early Bicolors were transported from Africa along the Sind-Punjab trade routes to India around 3000 bp, and the Durras later came in from India through the Middle East and down the Nile (Harlan & Stemler 1976, Mann et al. 1983).

The objective of this research effort is to classify the 8000-bp sorghums within the known wild or domesticated races or working groups through the use of infrared spectroscopy, and biotechnological and image-analysis procedures.

INFRARED SPECTROSCOPY

A relatively simple, inexpensive, and nondestructive method of identifying plant-remains by infrared spectroscopy has recently been developed (Harborne 1984, McLaren et al. 1990). The theoretical basis for it is that the seed is the product of a series of biochemical processes during its development, which are encoded in the genotype. The morphological changes involved in domestication result from a divergence of the gene pools, due both to selection by early cultivators and to natural processes. Thus, each seed variety is unique, as is its biochemistry, although it will share much of its identity with closely related forms.

The method involves the lipid fraction of the seed. Lipids are relatively stable molecules, are well understood chemically, and are easily extracted from seeds without damaging their morphology. They consist of various proportions of fatty acids, long-chain hydrocarbons, and long-chain mono-esters that vary according to species and, to some extent, time in the growth cycle. Lipids are also archaeologically stable and survive charring, which, in fact, protects them from decay and contamination. Infrared spectroscopy can be used to identify a charred seed by matching the overall pattern of the infrared spectrum from the specimen with a standard spectrum produced under identical conditions. The standard is selected on the basis of morphological criteria by which the seed's identity was narrowed to a few possibilities; the individual grain may then be chemically assigned to a particular species or subspecies.

BIOTECHNOLOGY

Several authors have reported success in isolating DNA from ancient specimens of both plants and animals. Cano et al. (1993) reported the successful isolation and amplification of DNA from a 120–135 million-year-old amber-fossilized weevil (Coleoptera) using polymerase chain reaction (PCR) techniques. Chloroplast DNA has been extracted and amplification of an 820-base pair DNA fragment of the chloroplast gene *rbc*L from a fossilized sample of *Magnolia* has been reported by Golenberg et al. (1990).

The major problem with fossilization is the oxidation of the pyrimidines, which can cause severe damage to the DNA. Intact DNA may be present in fossilized tissue if the process occurred in an anoxic environment such as at the bottom of a river bed, in amber, or in large tissues such as dinosaur bones. Archaeological maize specimens (4700 bp) and contemporary samples have had segments of the nuclear gene encoding alcohol dehydrogenase 2 (*Adh*2) isolated, amplified, and sequenced (Goloubinoff et al. 1993). The authors suggest that the gene pool of maize is millions of years old, and that several wild ancestral populations gave rise to domesticated races of maize.

IMAGE-ANALYSIS

Image-analysis is a nondestructive method for obtaining several different measurements of a sample. It provides a method by which one can acquire and examine images, process and enhance those images, and record a series of various image measurements. Van Zeist (1968) outlined key terminology for use in the description of ancient and charred cereal grains. These measurements include: (1) length (defined as the length of the grain as viewed from the ventral side from the lower to the upper end), (2) breadth, and (3) thickness. Two

types of index are also presented: (1) L:B index, which is a "measure of the degree of slenderness (high L:B indices) or plumpness of the grains" and (2) T:B index expressing relative height, with flat grains showing low values for this index. Image-analysis can provide additional measurements such as ArCircularity and ArRectangularity, which provide a mathematical expression of shape, and a series of measurements based on points, lines, and areas.

PROPOSED MATERIALS AND METHODS

INFRARED SPECTROSCOPY

We propose to compile a library of IR spectra of modern wild and domestic sorghums for comparison with the IR fingerprints of the archaeological sorghum. This will be supplemented by GC/MS analysis (gas chromatography interfaced with mass spectrometry). The GC/MS method will provide reliable identification since chromatographic elution patterns should be unique for each seed type. In addition, the identity and percent composition of each component in the extracts will be ascertained by mass spectrometric analysis of the bands in the GC elution patterns.

All of the seeds will be prepared in a manner similar to that described in the literature (McLaren et al. 1990), using hexane, chloroform, and isopropanol as extracting solvents. After removal of the solvent from each extraction (Soxhlet extractor), the residues will be subjected to gas chromatographic analyses using a diethylene glycol adipate-packed column, commonly used for lipid analyses. The chemical identity of each band in the chromatograms will be checked by GC/MS analysis. The IR spectrum of each extract will be obtained for comparison with those from other samples.

Modern sorghum seeds will be selected for analysis and comparison. The analyses will be carried out at three different laboratories: Department of Environmental Science, University of East London (John Evans); Department of Chemistry, Texas A & M University (Marian Hyman); and Department of Chemistry, Southern Methodist University (Ed Biehl). John Evans has had considerable experience with the chemical characterization of seeds, and his methods will be used by the other two scientists.

It is proposed that the analyses be done in three different laboratories because each analysis must be based on a single seed; and, with so small a sample, there is a possibility of significant instrumental error within a single laboratory. The repetition will give much greater credibility to the results.

BIOTECHNOLOGY

The ultimate objective of extracting DNA from 8000-bp sorghum kernels for PCR amplification was to compare the DNA banding patterns of ancient sorghum to those of presently cultivated sorghum. Depending on the extent of similarity between the RAPD patterns generated, phylogenetic relationships and genetic distance could be determined between the ancient sorghum and presently cultivated sorghums. Thus, if a significant relationship exists, this would suggest some form of domestication and dispersal of sorghum much earlier than is presently believed.

From the standpoint of morphology (Rooney & Miller 1982), it was decided that the germ would be the most appropriate part of the sorghum kernel for DNA extraction. As a precaution, the pericarp of the kernel was sterilized to remove any inhabiting fungi, which could have provided a source of DNA contamination, before dissecting the kernel.

Fresh and oven-charred kernels from the working group were used for DNA extraction. Samples were charred by baking at 230° C for two hours in a forced-air oven. Fresh samples were surface sterilized with 95% ethanol, and the germ dissected from the fresh kernels. The germ could not be identified in the charred kernels, so the entire kernel was used for DNA extraction.

Several DNA extraction procedures were investigated to obtain DNA from fresh and charred kernels of sufficient purity for PCR amplification (modified Cano & Poinar (1993) RPM procedure; modified Afanador et al. (1993) CTAB procedure; modified Tai & Tanksley (unpubl. SDS procedure; and Goodwin & Steven (1993) microwave miniprep). Following extraction, several samples required further purification and two techniques, Instagen purification matrix (Bio-RAD, Hercules, CA)

and Sephadex G-50 prepared according to Sambrook et al. (1989) were utilized. Amplification was carried out using standard PCR protocols.

IMAGE-ANALYSIS

Modern sorghum kernels were viewed with a Fisher Stereomaster Zoom microscope and photographed with a Microage CA2063 Auto Shutter Camera. The image was captured on Imaging Technology's CFG Frame Grabber board and projected onto two monitors, a Sony Trinitron PVM1343MD and a Sunshine VGA monitor. Using a Sunshine 386 IBM compatible computer and the Optimas Boiscan software package, images were analyzed and data exported to Microsoft Excel® for storage. The software was calibrated by capturing an image of a ruler. Employing a drawing tool, the ruler was measured and calibration data stored within the software for each level of magnification. Six measurements, area, area breadth, area circularity, major axis length, area perimeter, and area rectangularity, were taken and stored on a data sheet. Data was analyzed using SAS.

Measurements are also being conducted on representatives of the 5 races of sorghum along with wild taxa currently available through the United States Department of Agriculture's National Collection of sorghum. At the same time, black and white photographs of the 8000-bp specimens (Fig. 1) are being scanned into the computer for analysis by the Optimas Bioscan program.

RESULTS TO DATE

INFRARED SPECTROSCOPY

This method has been successfully used at Abu Hureyra and Mureybit in Syria to recognize and distinguish wild and domestic rye, which had been previously identified as wild einkorn wheat (McLaren et al. 1990). However, it has not been used to distinguish wild and domestic sorghum, or to identify other African grasses.

The first IR analyses of sorghum were made in 1992 by Dr. John Evans of the University of East London, using the 8000-bp seeds from the site E-75-6 at Nabta Playa. IR comparisons

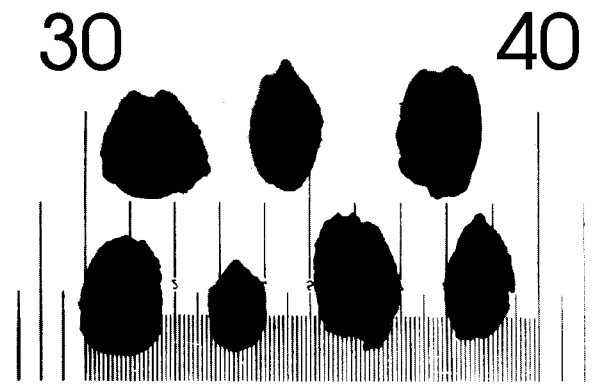

Fig. 1. Samples taken from Nabta Playa site E-75–6 to be scanned into computer and analyzed using imaging software

were made between five archaeological sorghum grains, six cultivated races of *Sorghum bicolor* (race *bicolor*, race *guinea*, small-seeded; race *guinea*, large-seeded; race *caudatum*; race *kafir*; race *durra*), and four wild species of sorghum (*S. arundinaceum*, *S. aethiopicum*, *S. verticilliflorum*, and *S. virgatum*). The modern samples were selected by Dr. Jack Harlan from the Herbarium collections at Kew.

The spectra obtained from the archaeological specimens were all very similar, which suggests that they belong to the same species. Three of the domestic specimens gave poor results (probably because of adhesive or fungicides used on the Herbarium collection). The spectra from the uncontaminated seeds, however, have a marked general similarity to those of the ancient specimens. This confirms that the latter are, indeed, sorghum. Though morphologically close to wild taxa, the 8000-bp samples show some resemblance to domesticated sorghum with respect to the lipid fraction of the grains (Wasylikowa et al. 1993).

Dr. E. Biehl analyzed nine extant sorghum samples sent to him by Dr. Fred Miller (Texas A & M University) using gas chromatography comparisons of lipids. Using Average Linkage Cluster Analysis, estimates made by the cluster provide a fairly close fit to the race information provided by Dr. Miller and also his own grouping of the material (personal communication).

The material falls basically into four of the major races as described by Harlan and de Wet (1972). Their placement of the working groups in the different races follows very closely to the breakdown of Murty and Govil (1967)

in their paper concerning the working groups. Murty and Govil placed the working group into sub-series, which Harlan and de Wet describe as races. Harlan and de Wet went a step further by dividing Murty and Govil's sub-series into the intermediate races such as *Kafir-Durra, Bicolor-Kafir*, etc., and by splitting the sub-series *Caffra* into two separate races, *Kafir* and *Caudatum*. The nine sorghums fall into four of the five major races described by Harlan and de Wet or three of the five sub-series of Murty and Govil (Tab. 1).

Using this criteria, it seems that all the entries fall into two major groups: Group 1 containing 1, 2, 3, 4, 5, 6, 7, and 9 which contain *Caudatum* in their background, and Group 2, which contains the only *Bicolor*, No. 8. The tree diagram from the cluster analysis suggests three major groups: Group 1, consisting of 1, 2, 3, 5, 6, 7, and 9; Group 2 with No. 8, and Group 3 having No. 4. The cluster analysis further breaks the Group 1 into three clusters: 1 and 6; 5 and 7; and 2, 3, and 9. These breakdowns are similar to those that Dr. Miller suggested. The only true odd character is RT x 2752, which, for some reason, is treated as a separate cluster in the analysis. In terms of lipid fractions, however, it was the most diverse. Unfortunately, the material chosen for this analysis was composed of germplasm that has been worked for several years in Dr. Miller's breeding programs and its true identity in terms of race or working group classification has become muddled. While intriguing, all these results must be regarded as highly speculative until a more detailed study of modern sorghum is completed. Currently, several proposals have been submitted to continue this avenue of research.

Table 1. Breakdown of nine sorghum samples using gas chromatography comparisons of lipids and average linkage cluster analysis

Pedigree	Race or Sub-Series
RTx430 (1)	*Caudatum* or *Caffra*
RTx2783 (2)	*Bicolor, Caudatum, Durra* or *Bicoloria, Caffra, Durra*
RTx436 (3)	*Caudatum, Durra* or *Caffra, Durra*
RTx2752 (4)	*Caudatum, Kafir* or *Caffra*
RTx623 (5)	*Caudatum, Kafir* or *Caffra*
BTx399 (6)	*Caudatum, Durra* or *Caffra, Durra*
BAZ9504 (7)	*Caudatum, Durra* or *Caffra, Durra*
Greenleaf (8)	*Bicolor* or *Bicoloria*
Hegari (9)	*Caudatum* or *Caffra*

BIOTECHNOLOGY

To achieve the objectives of this research problem, a number of factors had to be considered. The most important of these were the problems that could be encountered when extracting DNA from ancient plant material. One serious drawback with the sorghum kernels used in this study was not only their minuscule size, which gave little protection against DNA oxidative processes, but also the preparation of the kernels.

The 8000-bp sorghum grains were discovered in the hearths, pot-holes, and in different parts of the house fill. All kernels were preserved in charred condition and thus they were subjected to temperature extremes prior to being covered by playa silts (Wasylikowa et al. 1995). Oven-charred kernels for PCR amplification, therefore, were considered appropriate to compare the DNA banding patterns from fresh and charred kernels of the same variety. The results would indicate whether PCR-amplifiable DNA could be extracted from charred kernels in addition to giving some indication of the degree of pattern divergence caused by charring.

DNA extraction for PCR amplification from both fresh and charred kernels raised other questions such as the level of contaminating substances present in the kernel, a storage organ, and the part of the kernel that would yield the most DNA with the least amount of plant metabolite contaminants.

DNA was obtained from leaf material with the CTAB Protocol, and the germ of the fresh kernel and charred kernels with the CTAB and SDS Protocols. Though the leaf DNA could be amplified by the PCR without further purification the kernel DNA could not be amplified. On purification on sephadex G-50 columns, no DNA was recovered from the charred kernel DNA samples.

Problems were encountered with filter clogging in the RPM extraction Procedure. The fresh kernel from race 33 yielded a mucilaginous supernatant that could not go through the filter. As a result, the DNA could not be eluted from the filter as the TE could not pass through it. In addition, the charred material had too much debris which clogged the filter, even after double centrifugation of the supernatant. As a result, no DNA was obtained using this extraction procedure.

DNA was extracted from the members of the working group using the CTAB extraction procedure, purified on sephadex G-50 columns, quantified by fluorometry and stored at 4°C. However, this DNA was not of sufficient purity for PCR amplification. It is believed that less DNA should have been loaded onto the column or the DNA could be further purified by passing it through a sephadex G-50 column again, since PCR amplification of DNA extracted by both the CTAB and SDS DNA extraction procedures was obtained after purifying the DNA on sephadex G-50 columns.

Currently, we are investigating other means of DNA purification. We will be working closely with the Sorghum Biotechnological Group at Texas A & M University to try to solve this problem. Until its resolution, no attempt to extract DNA from the charred remains will be made.

IMAGE-ANALYSIS

Measurements have been recorded on the working group samples used in this study. However, principle component analysis, coupled with cluster analysis has not resulted in any significant findings. Other statistical approaches are being studied and it is hoped that this conference will provide some direction as to additional types of measurement that could be used for analysis in conjunction with the imaging system. This system coupled with infrared spectroscopy may offer the best hope of identifying the material in a non-destructive manner.

The earliest previously known sorghum in Africa, identified from impressions in pottery, is from Kadero, some 50 km north of Khartoum. The sorghum is ca. 5000 years old and is morphologically wild (Stemler 1990). However, domestic sorghum occurred in Arabia at around 4800 bp and possibly a millennium earlier (Potts in press). Unless one assumes that wild sorghums were carried to Arabia and domesticated there, cultivated sorghum must have existed in Africa well before that date. Resolution of this problem would have enormous potential for elucidating early African agricultural origins and the role these developments may have played in the Neolithic of the Sahel.

REFERENCES

AFANADOR K. L., HALEY S. D. & KELLY J. D. 1993. Adoption of a "mini-prep" DNA extraction method for RAPD marker analysis in common bean (*Phaseolus vulgaris* L.). Ann. Rep. Bean. Improv. Coop., 36: 10–11.

CANO J. R. & POINAR H. N. 1993. Rapid isolation of DNA from fossil and museum specimens suitable for PCR. Biotechniques, 15: 432–436.

DOGGETT H. 1988. Sorghum. Longman Scientific & Technical, John Wiley & Sons, Inc., New York.

FAO (Food and Agriculture Organization of the United Nations) 1992. FAO Production Yearbook. FAO, Rome.

GOLENBERG E. M., GIANNASI E. D., CLEGG M. T., SMILEY C. J., DURBIN M., HENDERSON D. & ZURAWSKI G. 1990. Chloroplast DNA sequence from a Miocene *Magnolia* species. Nature, 344: 656–658.

GOLOUBINOFF P., PÄÄBO S. & WILSON A. C. 1993. Evolution of maize inferred from sequence diversity of an Adh2 gene segment from archaeological specimens. Proc. Natl. Acad. Sci., 90: 1997–2001.

GOODWIN D. C. & STEVEN L. B. 1993. Microwave miniprep of total genomic DNA from fungi, plants, protists and animals. Biotechniques, 15: 438–444.

HARBORNE J. B. 1984. Pylochemical methods; Guide to modern techniques of plant analyses, 2nd edition. Bhapman and Hall, London

HARLAN J. R. & STEMLER A. 1976. The races of sorghum in Africa. In: Harlan J. R., de Wet J. M. J. & Stemler A. B. L. (eds) Origins of African plant domestication. Mouton Publishers, The Hague: 465–478.

— & de WET J. M. J. 1972. A simplified classification of cultivated sorghum. Crop Science, 12: 172–176.

HOUSE L. R. 1985. A guide to sorghum breeding. 2nd ed. Patancheru, A. P. 502324, India: ICRISAT.

MANN J. A., KIMBER C. T. & MILLER F. R. 1983. The origin and early cultivation of sorghums in Africa. Texas A & M Univ., Texas Agric. Exp. Stn. Bull., 1454.

McLAREN F. S., EVANS J. & HILLMAN G. C. 1990. Identification of charred seeds from Epipaleolithic sites of S.W. Asia. In: Archaeometry '90. Birkhäuser Verlag. Basel: 797–806.

MURTY B. R. & GOVIL J. N. 1967. Description of 70 groups in genus sorghum based on a modified Snowden's classification. Indian J. Genet., 27: 75–91.

POTTS D. T. (in press). The late prehistoric, protohistoric, and early historic periods in Eastern Arabia (c. 5000–1200 B.C.) J. World Prehist., 5 (2).

ROONEY L. W. & MILLER F. R. 1982. Variation in the structure and kernel characteristics of sorghum. ICRISAT. Proceedings of the International Symposium on sorghum grain quality, 28–31 October 1981, Patancheru, A.P., India: 143–162.

SAMBROOK J., FRITSCH E. F. & MANIATIS T. 1989.

Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory. Cold Spring Harbor, NY.

STEMLER A. B. L. 1990. A scanning electron microscope analysis of plant impression in pottery from the sites of Kadero, El Zakiab, Um Direiwa and El Kadada. Archéol. du Nil Moyen., 4: 87–105.

TAI H. T. & TANKSLEY S. D. (unpubl.) A rapid and inexpensive method for isolation of total DNA from dehydrated plant tissue. Plant Molecular Biology Reporter.

VIERLING R. A., XIANG Z., JOSHI C. P., GILBERT M. L. & NGUYEN H. T. (in press). Genetic diversity among elite sorghum lines revealed by restriction fragment length polymorphisms and random amplified polymorphic DNAs.

WASYLIKOWA K., HARLAN J. R., EVANS J., WENDORF F., SCHILD R., CLOSE A. E. & KRÓLIK H. 1993. Examination of botanical remains from early neolithic houses at Nabta Playa, Western Desert, Egypt, with special reference to sorghum grains. In: Shaw T., Sinclair P., Andah B. & Okpoko A. (eds) The archaeology of Africa. Food, Metals and Towns. Routledge, London: 154–164.

—, SCHILD R., KRÓLIK H., WENDORF F., KUBIAK-MARTENS L. & HARLAN J. R. 1995. Archaeobotany of the early Neolithic site at Nabta Playa, Western Desert, South Egypt (Preliminary results). Acta Plaeobot. 35 (1): 133–155.

WENDORF F., CLOSE A. E., SCHILD R., WASYLIKOWA K., HOUSLEY R. A., HARLAN J. R. & KRÓLIK H. 1992. Saharan exploitation of plants 8000 years BP. Nature, 359: 721–724.

ZEIST W. van 1968. Prehistoric and early historic food plants in the Netherlands. Palaeohistoria, 14: 41–172.