

Paleoethnobotanical Sampling Adequacy and Ubiquity

An Example from the American Southwest

Michael W. Diehl

One goal of modern archaeology is to assess how the problem of food provisioning was solved by the occupants of archaeological sites. To that end, identifying the kinds of food plants and assessing the relative frequency of their use has been a universal task in American archaeology since the late 1970s (Bohrer and Adams 1977; Ford 1979, 2003; Gremillion 1997; Minnis 1985, 2000; Popper and Hastorf 1988). Archaeologists are expected to analyze archaeobotanical remains in flotation samples. Despite the widespread recognition of that need, there are no published

standards that provide guidance about assessing sampling adequacy with respect to the number of analyzed samples and confidence in empirical measures of plant abundance. There are, in contrast, many studies that assess the impact of sample volume, processing techniques, identifying promising locations to sample within sites, the mechanics of sample collection, and formation processes (d'Alpoim Guedes and Spengler 2014; Lee 2012; Lennstrom and Hastorf 1995; Miksicek 1987; Minnis 1981; Pearsall 1988; van Roggen 2016; Wright 2003).

ABSTRACT

Archaeological researchers and compliance review officers need to know whether or not a research plan will yield sufficient information to meet research objectives. Despite the need, a key question is often not addressed in proposals or reports: how many flotation samples are sufficient to adequately characterize the food procurement practices at an archaeological site? This article reviews the relationship between ubiquity and statistical probability. By considering the relationship between theoretical ubiquity, measured ubiquity, population ubiquity, and statistical probability, archaeological researchers and compliance officers may assess how many samples must be analyzed in order to adequately characterize the paleoethnobotanical assemblage from a site. These considerations generally apply to any other archaeological data in which presence-absence measures are commonly used and are especially relevant to diet-breadth models in the interest arena of behavioral ecology.

Los investigadores arqueológicos y los oficiales de revisión de conformidad necesitan saber si un plan de investigación rendirá información suficiente para lograr las metas de la investigación. A pesar de esta necesidad, hay una pregunta clave que las propuestas o informes suelen dejar sin respuesta: «¿Cuántas muestras de flotación son suficientes para caracterizar adecuadamente las prácticas de obtención de comida en un sitio arqueológico?». Este artículo repasa la relación entre los conceptos de ubicuidad y probabilidad estadística. Considerando la relación entre la ubicuidad teórica, la ubicuidad medida, la ubicuidad de la población y la probabilidad estadística, los investigadores arqueológicos y los oficiales de conformidad pueden evaluar el número de muestras necesarias para caracterizar adecuadamente la colección paleoetnobotánica de un sitio. Estas consideraciones se pueden aplicar generalmente a casos arqueológicos en los que se utilizan medidas de presencia-ausencia y son especialmente relevantes para modelos de amplitud de dieta en ecología del comportamiento.

Advances in Archaeological Practice 5(2), 2017, pp. 196–205
Copyright 2017 © Society for American Archaeology
DOI:10.1017/aap.2017.5

This study asks how many samples should be analyzed to meet research objectives related to assessing prehistoric food selection, acquisition, and production, while retaining empirical confidence in the results. Answering that question allows archaeologists to structure research such that research objectives are met and will allow all those who evaluate research and research proposals, including compliance and granting officers, to assess whether analytical efforts are sufficient to meet stated research objectives and compliance standards.

In the first section, I review how food plant tissues wind up preserved in ways that allow researchers to evaluate past subsistence practices. I next discuss the “ubiquity” index, ways that formation processes relate to that commonly used measure of plants’ importance, and terms that clarify how accurate ubiquity scores may be achieved in light of research goals. In the third section, I use the archaeobotanical assemblage from the Las Capas site—a series of hamlets occupied by seasonally mobile forager-farmers around the tenth century B.C.—to illustrate the relationship between the number of analyzed samples and statistical confidence in ubiquity values. In the final section, I offer a suite of recommendations for sampling efforts and for compliance officers to assess the adequacy of both research designs and analytical efforts. The methods described here should be applicable to other studies of mobile or sedentary foragers or farmers.

FORMATION PROCESSES, BURNING, AND PRESERVATION

The desirability of ubiquity as a measure of plant importance in an assemblage is linked to formation processes that include sources of deposition, post-depositional environments, and the durability of plant tissues. Although it may seem counterintuitive, burning is arguably the best thing that can happen to a seed from an archaeological point of view; charring that does not destroy a seed renders it impervious to most forms of decomposition because charred plants are inedible to insects, animals, bacteria, and fungi (Miksicek 1987; Minnis 1981; Wright 2003).

Burned seeds in archaeological features are basically trashy remnants of food processing or consumption. Plants may be exposed to fire through parching and winnowing (to remove adhering nonfood tissues such as glumes), roasting in pits, spills that occur in hearths during cooking, disposal of small amounts of remnant or undesired food waste into hearths after meals, or bulk trash disposal, as when features are cleaned out for reuse. In the southwestern United States, food debris is usually found in extramural dumping areas, trash mounds, hearths, pits, or the interiors of abandoned houses.

Food waste is undesirable. I assume that, under ordinary food consumption circumstances, prehistoric people tried to avoid large spills, catastrophic spoilage, or great waste, even though those things must have occasionally happened. Unusual circumstances, such as sacrificial offerings or granary fires, may occur. Archaeologists fortunately do not mistake them for deposits associated with the routine food production-consumption-disposal cycle (Pauketat et al. 2002).

Since food waste is undesirable, the events that cause food to be burned may be viewed as minor accidents. They are the lesser spills, small debris-clearing instances and the like, that occur regularly in the ordinary business of food production and consumption. Those processes randomly sample foods that were consumed in the past in rough proportion to their importance in diets. The more often a food was cooked and consumed, the more likely it was to be burned, and the more commonly it should be observed in an assemblage.

MEASURED UBIQUITY, POPULATION UBIQUITY, AND THEORETICAL UBIQUITY

The ubiquity of a plant is a measure of its presence or absence either in flotation samples (sample ubiquity) or in different archaeological features or locations (context ubiquity) in an assemblage of flotation samples collected from an archaeological site. Its value ranges from 0 (a plant is not observed in any sample or context) to 1.00 (a plant is present in every sample or feature). As originally crafted, ubiquity referred to the total number of samples in which a taxon was observed, relative to the total number of analyzed samples (Minnis 1981; Popper 1988). In this study, I am interested in something I call “context ubiquity” because spatial separation between sampled contexts is important. It is often the case that multiple flotation samples are obtained from the same depositional context, such as the fill of one large pit or of one abandoned house. Flotation samples from close proximity within a midden or trash-filled house may redundantly sample the same trash disposal event. Treating multiple flotation samples from the same context as discrete observations, as may happen when one focuses on sample ubiquity, may violate the general statistical requirement that each observation in a data set be independent of the others.

The measured ubiquity of a plant is given by the formula:

$$U_{\text{taxon}} = N_{\text{taxon}} / N_{\text{total}}$$

where U_{taxon} is the measured ubiquity of the plant, N_{taxon} is the total number of features or spatially discrete (including large vertical separation) contexts in which at least one specimen of the plant was observed, and N_{total} is the total number of analyzed features (or units) in the assemblage.

Obviously, archaeobotanists are better positioned to distinguish among important rather than unimportant food resources as the amount of data—the number of analyzed flotation samples or contexts—increases. This article directs attention to the degree of empirical confidence in our ubiquity indices, however, and as a prelude to that discussion, we need to consider other ways that the ubiquity index may be made to serve our research goals.

Another question that most archaeobotanists ask is: What is the “real” ubiquity of a plant for a prehistoric target population? Here, by “population,” I mean a unit of generalization. Most often the target population is an entire archaeological site, but it

could be one temporal component at a multicomponent site, or one temporal component in an entire geographic region. Since no standard term has been applied to the concept of an ultimate, objectively real ubiquity, I here introduce the term “population ubiquity” or, at times, $U_{\text{population}}$. Imagine traveling back in time with an array of cleverly hidden cameras and recording every episode of trash dumping or burning, so that one could really know just how often a plant was used and wound up charred and deposited in the ground. If one could record each instance, one would be able to calculate a population ubiquity value for each plant used by the occupants of a site.

Population ubiquity is usually unknowable because archaeologists rarely excavate entire archaeological sites. If one could excavate every feature, however, and analyze a flotation sample from each of them, then the population (which is the entire suite of features) is completely known through sampling. U_{measured} and $U_{\text{population}}$ for a particular plant will have identical values.

The final question most archaeobotanists ask is addressed by this research. How many samples do I need to collect to be satisfied that I know which plants were “reasonably important” for the purpose of my research questions? The term “theoretical ubiquity” or $U_{\text{theoretical}}$ is introduced to answer that question. It is an arbitrary value chosen by the researcher that is used to determine how many contexts need to be analyzed in order to have confidence in the measured ubiquity results.

The terms are arithmetically related. Confidence in measured ubiquity increases as the total number of contexts (N) analyzed approaches the number dictated by our research goal expressed as theoretical ubiquity; furthermore, measured ubiquity and population ubiquity will have identical values if 100 percent sampling of a site is achieved. Finally, if theoretical ubiquity thresholds are set very low (for example, on the order of 0.0001) and confidence demands are high (for example, a 95 percent chance of observing any plant with a ubiquity of 0.0001), then the number of samples (N) required to achieve that level of confidence will become very great.

Archaeological Applications of Ubiquity to Diet-Breadth Models

Ubiquity is important because it is ideal for use within certain research paradigms, especially behavioral ecology. Diet-breadth studies model changes in food acquisition strategies by examining when individual food items are included or excluded from food procurement activities (Kelly 1995; Simms 1987; Winterhalder and Kennett 2006). The presence or absence of food items among a suite of resources is central to such studies, and ways to improve our confidence in results are a subject of common interest.

The central assumption of diet-breadth models is that people on the whole, acting on the best information available, tend to use those resources that offer the greatest food value (expressed as calories or, in some studies, other critical nutritive components) for the energy (calories) expended to locate, acquire, transport, process, and consume them. Generally, if a food resource is inexpensive to acquire and offers high food value, it tends to be used first. It follows that if people “optimize” in the way that diet-

breadth models assume, then foods with the highest return rates should have the highest ubiquity values in an archaeobotanical assemblage; foods with very low return rates should have the lowest ubiquity values because they should not have been often consumed. The concordance between food resource return rates and measured ubiquity for plants observed in archaeological assemblages has been demonstrated for southern Arizona for the interval 1200 B.C. through A.D. 1150 (Diehl 1997, 2015; Diehl and Waters 2006).

Theoretical ubiquity, as discussed above, is useful for designing sampling strategies that integrate well with behavioral ecology models. Without knowing a priori the population ubiquity of every plant, researchers need to establish some sense of how important a plant needs to have been for it to be relevant for a particular research problem. By selecting $U_{\text{theoretical}}$ values and p values appropriate to the research questions asked, one can test diet-breadth models, or compare diet breadths of different groups, with reasonable confidence that observed differences in measured ubiquity are not an artifact of the sampling effort. Ubiquity also allows one to assess which patches of food on the landscape were used, and to track changes in landscape use over time. In the Tucson Basin, for example, increased dependence on agriculture through time was accompanied by an increase in the use of floodplain plants and a decline in the use of piedmont and montane plants (Diehl and Waters 2006).

The foregoing observations do not imply that the application of ubiquity is limited to behavioral ecology models. Any study, even a mere catalog list of the plants identified at an archaeological site, requires a sense of whether or not the presence or absence of a plant in an assemblage is empirically well supported (Bohrer and Adams 1977; Lee 2012; van Roggen 2016). Ubiquity is therefore central to any archaeological study that claims a basic concern with the kinds of foods consumed by the occupants of a site. For anthropological studies that embrace deep theory, and for simple descriptive studies alike, ubiquity and sample size are essential concerns.

HOW MANY SAMPLES DOES ONE NEED TO KNOW WHETHER A PLANT WAS USED?

How does a researcher know that the list of plants identified in an archaeological assemblage accurately reflects the range of foods consumed by the occupants of a site or a temporal component of that site? The answer to that question depends on the number of discrete contexts that are sampled (N), the level of confidence one desires in one’s results (p), and suppositions about how important a particular food plant *should be* in an assemblage for an archaeologist to want to know that it was used ($U_{\text{theoretical}}$). It is of course also dependent on the volume of flotation samples analyzed (Lee 2012; Miksicek 1987; van Roggen 2016), but for this study I assume that regional archaeobotanists already have a good idea of how large their samples need to be.

The number of samples required is given by the following equation:

$$p = (1 - U_{\text{theoretical}})^N$$

where p is the likelihood of NOT observing a given plant, $U_{\text{theoretical}}$ is the theoretical ubiquity, and N is the number of analyzed flotation samples from different features or discrete locations within trash middens. In that equation, $1 - U_{\text{theoretical}}$ is the chance of not observing a plant with a given theoretical ubiquity in one sample. The equation describes how the odds of failure to detect change across multiple samples of varying number (N).

Consider the following example. An archaeologist wishes to know whether or not 10 flotation samples will be sufficient in order to detect plants with a theoretical ubiquity of 0.10. Recall from our discussion of formation processes that we suppose food plant remains to essentially be randomly distributed in trash deposits in proportion to their overall use. So if a plant had a theoretical ubiquity of 0.10, then if we analyze only one sample, we will likely not see that plant. The archaeologist decides that 10 samples might be enough to capture a plant where $U_{\text{theoretical}} = 0.10$, but wants to be sure. The archaeologist can calculate the likelihood by setting $U_{\text{theoretical}}$ to 0.10, N to 10, and solving for p . In this example, $p = (1 - 0.10)^{10} = 0.35$.

Continuing with the example, the archaeologist sees that there is a 0.35 likelihood that a plant with $U_{\text{theoretical}} 0.10$ will be overlooked if only 10 features are analyzed. Concerned that this leaves too much room to miss important plants, the archaeologist asks how many features must be sampled in order to observe plants with $U_{\text{theoretical}} = 0.10$ and $p = 0.05$. The equation $0.05 = (1 - 0.10)^N$ shows that $N = 28$ yields $p = 0.052$ —acceptably accurate for the archaeologist's needs.

Continuing with the example, imagine that an ambitious archaeologist wants to know that measured ubiquities accurately reflect the real world, in the past, as it really happened. In other words, they want $U_{\text{theoretical}}$ to be very close to $U_{\text{population}}$. In this case, p is set very low, and $U_{\text{theoretical}}$ is set very low. When both p and $U_{\text{theoretical}}$ approach zero, U_{measured} and $U_{\text{population}}$ approach the same values for each taxon, and the number of required analyzed contexts (N) becomes very large.

Discussion

A researcher's choice of the desired $U_{\text{theoretical}}$ and p values are critical. How confident is "confident enough?" How important must a plant have been for the archaeologist to need to know about it? Those are decisions that must be made in light of the research questions to be answered. In my own research, I usually want to know that I am $p = 0.05$ likely to fail to observe a plant that was used and deposited five percent of the time ($U_{\text{theoretical}} = 0.05$). According to the equation, absent other mitigating circumstances, I need at least 59 sampled features.

DIET BREADTH, ERROR, AND SAMPLE SIZE AT THE LAS CAPAS SITE NEAR TUCSON ARIZONA

As statistical presentations go, the foregoing equation may sound good, and yet many reading this article may have doubts

about whether it will work in actual practice. The utility of the equation is demonstrated by drawing upon the paleoethnobotanical assemblage from an intensively studied prehistoric archaeological site, Las Capas (AZ AA:12:111 [ASM]; Diehl 2005; Mabry 2008; Sliva 2005; Vint 2015; Vint and Nials 2015; Whittlesey et al. 2010). The site lies on the east bank of the lower Santa Cruz River in Tucson, Arizona (Figure 1). Expansion of a wastewater treatment facility, local road improvements, and other developments have resulted in four substantial cultural resource management-based archaeological excavations at the site. Las Capas was occupied from roughly 2000 B.C. through 400 B.C., with the most extensive component occupied from B.C. 1200–600. Approximately 30,000 archaeological features, mostly extramural pits but also including pit-house floors, intramural pits, earth-berm bordered grid gardens, and associated irrigation canals, have been identified. Thousands of flotation samples have been analyzed and reported in the foregoing studies. Of those, 1,231 houses, intramural pits, and extramural pits contained charred food plant remains consistent with refuse from the food consumption-discard cycle.

The power of the predictive model is demonstrated by randomly sampling the database from Las Capas and plotting the number of identified taxa against the number of samples. To keep things simple, I randomly sampled the assemblage in five-sample increments, starting with five samples and going up to 65 samples. The ubiquities of the taxa observed in the randomly selected subsamples are presented in Table 1. The taxa (listed by common name) are presented in the order "detected" in the random sample. In this table, the measured ubiquities of each taxon are presented in columns for 5, 10, 15, etc. features, up to 65 features in the subsample. The final column is the theoretical ubiquity of each taxon in 1,231 sampled features. The table leads to two important observations.

First, the table shows how small sample sizes can affect measured ubiquity values, both by excluding important taxa and also by overstating the importance of other taxa. In the random sample, loco, sacaton grass, and tansy mustard are abundant enough in the first 10 samples that even at 40 samples they have relatively high ubiquities (greater than 0.10), as compared with the (more accurate) ubiquities recorded for all 1,231 sampled features in the right-hand column. At 60 samples, the measured ubiquities of loco and sacaton approach within 0.05 of the ubiquity observed using all 1,231 sampled features.

Figure 2 illustrates how increasing the number of samples brings the measured ubiquity close to the theoretical ubiquity. Our predictive equation suggests that, at 59 samples, plants should be represented by at least one specimen in an assemblage if they have a population ubiquity on the order of 0.05. If we consider that diet breadth is reasonably well described by observing most (95 of 100) plants that were eaten at one meal in 20 ($U_{\text{theoretical}} = .05$), then the goal has been met. Of course, if there are plants at a site with $U_{\text{population}} < 0.05$, then there is a greater likelihood than $p = 0.05$ that they will not be observed in 59 sampled features. If we wanted to be very confident that most plants with $U_{\text{population}} = 0.01$ were observed, then we would have to set $U_{\text{theoretical}} = 0.01$, set $p = 0.05$, and solve for N .

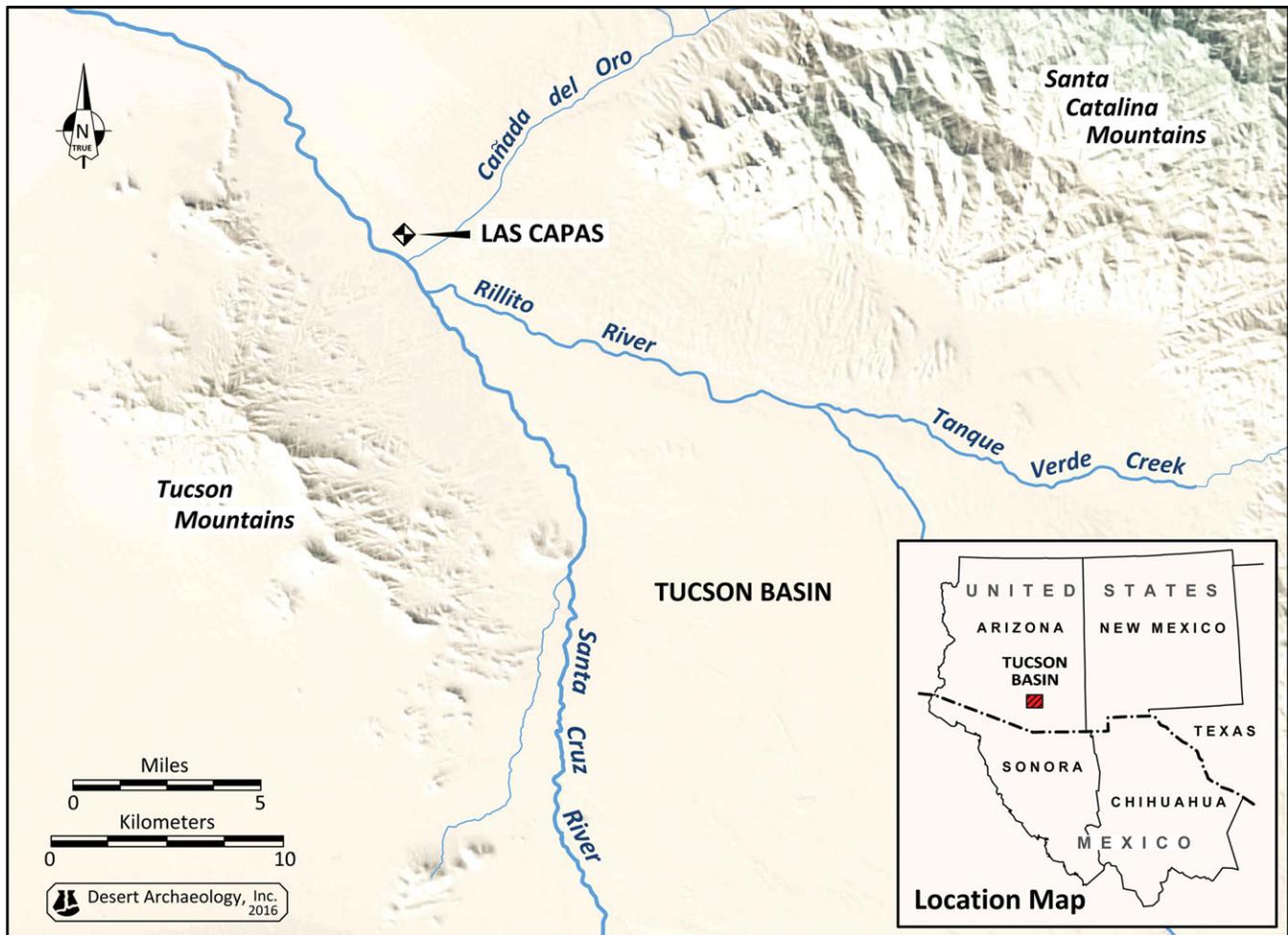


FIGURE 1. The site of Las Capas in the Tucson Basin of southern Arizona (illustration by C. Gilman).

The equation is imperfect because it is probabilistic rather than absolute. Depending on the luck of the draw in a random sample, some plants with the desired $U_{\text{theoretical}}$ value may be overlooked. Figure 2 plots the number of taxa whose measured ubiquity in a subsample of the data differs by more than 0.05 from the ubiquity in 1,231 features, in five-sample increments. The hatched lines represent the best linear fit for the distribution and the 95 percent confidence bars around the best fit is shown. This particular random subsample seems luckily to have generated measured ubiquities within 0.05 of $U_{\text{theoretical}}$ at 50 samples. The linear fit suggests that if some other subset of the 1,231 analyzed features were chosen, all of the plants with ubiquities greater than or equal to 0.05 would be observed in 64 samples.

Fifty-Nine Sampled Features Represents a Minimum Effort, More May Be Needed

At any site where food consumption, processing, or trash discard activities are spatially segregated, feature fill might contain the remnants of very different kinds of food-related activities. Spe-

cialized contexts are not the best contexts for characterizing the general subsistence pattern of a group of people. To the extent that different features contain “de facto refuse” specific to the preparation of specific foods (Schiffer 1987), such features are undesirable for empirically measuring the ubiquities of specific taxa or the diet breadth of the occupants of a site.

Most of the paleobotanical samples from Las Capas come from residential areas, and there is very little internal heterogeneity evident within those areas. That may be in part the consequence of occupation by people who were seasonally mobile (Diehl and Davis 2016). More sedentary people with specific activity areas such as platform mounds, ceremonial areas, granaries, or spatially discrete food processing areas will yield assemblages skewed in the direction of whichever kinds of activity areas are most represented in excavated contexts. Fifty-nine features may be adequate for sampling a site where most of the excavated features are full of generalized refuse from ordinary and commonplace food preparation, consumption, and trash disposal events. For sites with strong spatial segregation in activities, sampling 59 features may be insufficient. Likewise, if a site has multiple temporal components and one wishes to compare

TABLE 1. Measured Ubiquities, by Number of Analyzed Features, of Plant Taxa Observed in the Las Capas Paleoethnobotanical Assemblage in 65 Randomly Selected Features.

Plant Common Name	Number of Analyzed Features													1,231 Measured Ubiquity in All Sampled Features
	5	10	15	20	25	30	35	40	45	50	55	60	65	
Cheno-ams	1.00	0.90	0.93	0.90	0.88	0.77	0.66	0.65	0.62	0.58	0.56	0.52	0.49	0.50
Maize (<i>Zea mays</i>)	0.80	0.80	0.87	0.85	0.84	0.83	0.83	0.78	0.76	0.78	0.78	0.80	0.80	0.81
Loco (<i>Astragalus</i>)	0.60	0.30	0.27	0.20	0.16	0.13	0.11	0.10	0.08	0.05	0.07	0.07	0.06	0.02
Dock (<i>Rumex</i>)	0.40	0.40	0.33	0.30	0.28	0.23	0.23	0.23	0.20	0.20	0.20	0.18	0.18	0.09
False purslane (<i>Trianthema</i>)	0.40	0.20	0.13	0.15	0.20	0.17	0.17	0.20	0.18	0.20	0.20	0.18	0.18	0.19
Mesquite (<i>Prosopis</i>)	0.40	0.30	0.27	0.25	0.24	0.20	0.17	0.18	0.18	0.18	0.16	0.17	0.17	0.21
Mint family (Labiatae)	0.40	0.30	0.20	0.25	0.20	0.17	0.14	0.15	0.13	0.12	0.13	0.12	0.11	0.08
Sacaton grass (<i>Sporobolus</i>)	0.40	0.20	0.13	0.15	0.16	0.13	0.11	0.10	0.09	0.08	0.07	0.07	0.06	0.07
Tansy mustard (<i>Descurainia</i>)	0.40	0.40	0.47	0.40	0.32	0.27	0.23	0.20	0.17	0.16	0.15	0.13	0.12	0.20
Cholla cactus (<i>Cylindropuntia</i>)	0.20	0.20	0.13	0.10	0.08	0.07	0.06	0.05	0.04	0.04	0.04	0.03	0.03	0.01
Grass family (Gramineae)	0.20	0.10	0.07	0.05	0.08	0.10	0.09	0.08	0.07	0.06	0.05	0.05	0.05	0.05
Panicgrass (<i>Panicum</i>)	0.20	0.10	0.07	0.05	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.01
Purslane (<i>Portulaca</i>)	0.20	0.10	0.20	0.15	0.12	0.13	0.11	0.10	0.11	0.10	0.09	0.08	0.08	0.08
Spiderling (<i>Boerhaavia</i>)	0.20	0.30	0.20	0.20	0.16	0.13	0.11	0.10	0.09	0.08	0.07	0.07	0.06	0.03
Spurge (<i>Euphorbia</i>)	0.20	0.10	0.07	0.05	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.01
Stinkgrass (<i>Eragrostis</i>)	0.20	0.20	0.20	0.15	0.12	0.10	0.09	0.10	0.09	0.08	0.07	0.07	0.06	0.04
Saguaro cactus (<i>Carnegiea</i>)		0.20	0.13	0.15	0.12	0.10	0.14	0.15	0.16	0.16	0.15	0.13	0.14	0.12
Cactus family (Cactaceae)		0.10	0.07	0.05	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.01
Clammyweed (<i>Polanisia</i>)		0.10	0.27	0.20	0.16	0.13	0.14	0.15	0.13	0.12	0.13	0.12	0.11	0.06
Juniper (<i>Juniperus</i>)		0.10	0.07	0.05	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	<0.01
Ricegrass (<i>Oryzopsis</i>)		0.10	0.07	0.05	0.04	0.03	0.11	0.03	0.02	0.02	0.02	0.02	0.02	<0.01
Globemallow (<i>Sphaeralcea</i>)					0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.01
Goosefoot (<i>Chenopodium</i>)					0.04	0.07	0.14	0.20	0.27	0.30	0.33	0.32	0.35	0.30
Prickly pear cactus (<i>Platyopuntia</i>)					0.04	0.03	0.06	0.05	0.04	0.04	0.04	0.03	0.03	0.04
Walnut (<i>Juglans</i>)					0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	<0.01
Seepweed (<i>Suaeda</i>)						0.10	0.09	0.10	0.09	0.08	0.07	0.07	0.08	0.04
Hedgehog cactus (<i>Echinocactus</i>)							0.03	0.03	0.02	0.04	0.07	0.06	0.06	0.06
Pigweed (<i>Amaranthus</i>)							0.09	0.10	0.11	0.12	0.15	0.13	0.14	0.11
Composite family (Compositae)								0.03	0.02	0.02	0.02	0.02	0.02	0.03
Caltrop (<i>Kallstroemia</i>)									0.02	0.02	0.02	0.02	0.02	<0.01
Saltbush (<i>Atriplex</i>)										0.02	0.02	0.02	0.02	0.01
Sedge family (Cyperaceae)											0.02	0.02	0.02	0.03
Bunchgrass (<i>Bouteloua</i>)											0.02	0.02	0.02	<0.01

ubiquities between the two components, each component would need to be represented by at least 59 samples.

Another consideration is that not all food plant remains will be preserved in rough concordance with the frequency with which they are used (Miksicek 1987; Pearsall 1988). The 26 taxa represented in our 65-feature sample from Las Capas contain the most durable seeds and tissues (maize cupules) in the Las Capas assemblage (Diehl 2015). There were 29 other seed taxa in the complete Las Capas assemblage that were not identified by the 65-feature subsample. All 29 had measured ubiquities of less

than 0.05. Some of these might have been consumed more frequently than their ubiquities indicate because some seeds are porous, or prepared in ways that do not facilitate charring (Miksicek 1987; Wright 2003).

One such candidate in most prehistoric archaeological sites from the American Southwest is the common bean (*Phaseolus vulgaris*). Beans in the American Southwest were primarily pot-boiled. At Hohokam sites from the eleventh century A.D. where they are present, they rarely occur with measured ubiquities of 0.05 or greater. Beans are part of the agricultural triumvirate of

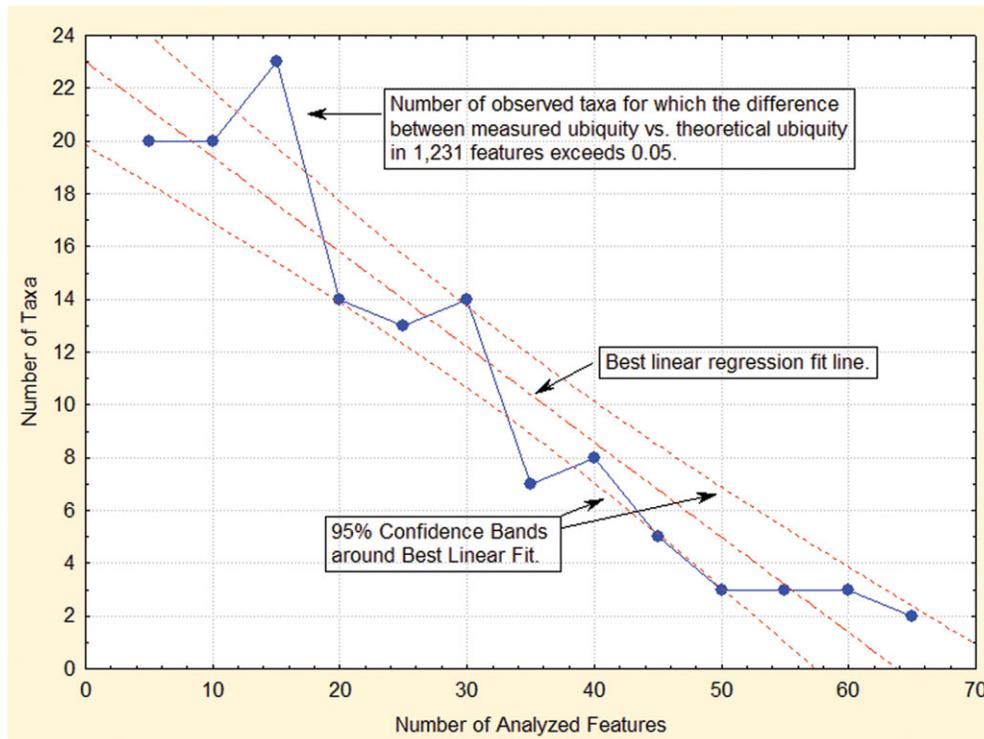


FIGURE 2. Scatterplot of the number of plant taxa for which the difference between measured and theoretical ubiquity exceeds 0.05, by number of analyzed features.

maize, beans, and squash associated with sedentary farmers throughout the American Southwest after the sixth century A.D, so 0.05 is probably something close to a “best circumstances” recovery rate from sites, despite the fact that beans were invariably a staple crop (Ford 1981; Moerman 1998).

At Las Capas, small cotyledon fragments identified as “cf. *Phaseolus* sp.” have been identified on three occasions. They may be domesticated common beans, or instead a wild species of *Phaseolus* that occurs in southern Arizona. The first compelling example of a domesticated bean comes from a 400 B.C.–A.D 50 component at the Santa Cruz Bend site (Huckell 1998). Was common bean used at Las Capas? So far there is no conclusive evidence that it was. But then food preparation at Las Capas must have been very different from food preparation at Hohokam sites because the people living at Las Capas had no ceramic vessels suitable for food preparation or storage (Heidke 1999, 2005). If beans were consumed at Las Capas, they were stone boiled in bags or baskets. Stone boiling was achieved by heating rocks in a fire, and moving the hot rocks elsewhere to drop them in baskets or bags. Food spills from that process wound up on the ground, rather than in a hearth where charring might promote preservation.

Another example is provided by a potentially important plant from the nightshade family that was observed in the full 1,231-feature assemblage from Las Capas, but was not observed in the 65-feature subsample. “*Physalis/Solanum*” type seeds are small, flat, porous, and lack a hard seed coat. Tomatillos (*Physalis*) are

a common ingredient in southwestern Native American cuisine. The observed remnants would have been overlooked in less intensively sampled sites because of the low measured ubiquity in all 1,231 samples ($U_{\text{Physalis}} = 0.003$). According to our equation, in order to have a 0.95 probability of detecting a plant with a $U_{\text{theoretical}} = 0.003$, we would expect to analyze samples from 990 features.

The point here is not to suggest that tomatillos or common beans were important resources at Las Capas; the empirical evidence does not support that finding. The point is that the formation processes associated with food processing, preparation, consumption, discard, and post-depositional effects are destructive processes. If one wants to know whether or not especially fragile food plant tissues were consumed at a site, 59 features is an insufficient number to detect that plant’s presence in an assemblage.

One should also consider that the equation in this text has been tested in circumstances where preservation is nearly ideal. The Las Capas site is located in the American Southwest, where preservation is very good (Adams 2001). It is also located on an alluvial floodplain and capped by several meters of fill. Evidence of chronic bioturbation, such as insect exoskeleton fragments and rodent droppings, is scant. Sites located in contexts where preservation processes are more adverse will likely require additional sampling. Archaeobotanists may use extant databases to correct the equation presented here to account for local formation processes.

Finally, there is the consideration that most southwestern United States archaeological research designs involve diachronic comparisons of multiple components at a site, and synchronic comparisons of multiple social units (households, house compounds, discrete clusters of compounds within larger archaeological sites, and so forth). The equation suggests that 59 samples is the reasonable minimum sampling effort *per household, temporal component, or other unit of comparison*.

What If a Site Does Not Have 59 Features?

Entire sites are almost never excavated. Most archaeological fieldwork in the United States falls within the realm of cultural resource management conducted for state or federal agencies; in such instances, excavated portions of sites are typically constrained to an "Area of Potential Effect." Access to portions of a site are therefore defined by constraints imposed by the landowner or the contracting party. In other circumstances, archaeological sites may be quite small. In such instances, it may be that fewer than 59 features are available to be excavated. Does that mean that one should give up on flotation analysis altogether because we cannot sample enough features to meet the sample requirements dictated by our desired theoretical ubiquity? The answer is no.

Reviewing our measured and population ubiquities, recall that their values approach identity as sampling efforts approach 100 percent. When sites are small, a 100 percent sample defines the population. If samples from every feature are analyzed, then one has complete confidence in the sampling effort. In such circumstances, the only way to know that the paleobotanical assemblage adequately represents the site from which it was derived is to analyze at least one sample from every feature.

When sites are very large but areas of potential effect are not, 59 features may not be available to sample. In that event, all available features should be sampled. It is often the case that sites are revisited in future cultural resource management projects as new areas of potential effect are created and mitigated. Archaeological sites located near urban areas, critical facilities, and roads have a tendency to be repeatedly impacted by development. The repeated sampling that occurs in association with small projects eventually produces assemblages that are large enough to provide empirically robust information about plant ubiquities and diet breadth.

CONCLUSIONS

Archaeologists have been concerned with sampling efforts and statistical confidence for decades (Drennan 1996; Shennan 1988). By drawing attention to the issue of minimum acceptable effort, and ways to assess whether an effort has been sufficient to achieve research goals, this article provides means by which archaeologists and compliance officers in public governance and agencies can assess the adequacy of research efforts in archaeological research and reporting efforts.

In this article, I have shown how the formation processes of the archaeobotanical record interact with theoretical models and measures of diet breadth in quantifying the importance of plants using the ubiquity index. Researchers may use the information

presented here to structure analyses a priori to have the best chance of answering their questions about the food plants consumed at archaeological sites. The example of the Las Capas site in southern Arizona shows that, under relatively ideal preservation circumstances, and using flotation samples that exceed four liters in volume, a reasonable confidence (95 percent chance) that plants with ubiquities greater than or equal to 0.05 would be spotted in 59 analyzed features. Sites with relatively inferior preservation, or from which small (volume) samples were collected, would require that more features be analyzed.

Toward that effort, I introduced the terms measured ubiquity, theoretical ubiquity, and population ubiquity. I have shown that measured and theoretical ubiquity are useful terms for guiding analytical efforts because they allow one to determine how many samples should be analyzed when one cannot analyze samples from all features at an archaeological site. The term population ubiquity describes what one would observe if all features at a site are sampled and their contents analyzed. When sites are small, confidence in the sampling effort requires 100 percent sampling.

I have touched upon but not visited in detail the subject of differential preservation of seeds of different plants or of different tissues from the same plant. As noted earlier, some plant remains are inherently soft, or low-density, or the edible portions are prepared in ways that inhibit archaeological preservation. It is beyond the scope of this article to provide guidance for how to handle specific plants, because they are many and varied, and because depositional environments vary from region to region. But for hard tissues such as maize cupules, small seeds, and wood, where preservation is reasonable, the concepts explored in this research are broadly applicable to archaeological analyses from most arid landscapes, or from other contexts such as out-house pits or anaerobic environments. Archaeologists working in different preservation environments should be able to apply these concepts in a modified form to account for local preservation for the purpose of structuring data recovery efforts.

Acknowledgments

Archaeological research at Las Capas and the preparation of this article were made possible by funding provided by Pima County, Arizona. The project was funded and administered by the Pima County Regional Wastewater Reclamation Department and the Pima County Office of Sustainability and Conservation, Cultural Resources and Historical Preservation Division (Arizona Antiquities Act Project Specific Permit Number 2008–118). The author is grateful to Jim Heidke and Paul Minnis for reviewing early drafts of this article.

Data Availability Statement

Macrobotanical data from Las Capas are available in Diehl (2005), Vint (2015), and Whittlesey et al. (2010).

REFERENCES CITED

- Adams, Karen R.
2001 Looking Back through Time: Southwestern U.S. Archaeobotany at the New Millennium. In *Ethnobiology at the Millennium: Past Promise and*

- Future Prospects, edited by Richard I. Ford, pp. 49–100. Anthropological Papers, Museum of Anthropology, University of Michigan no. 91, Ann Arbor, Michigan.
- Bohrer, Vorsila L., and Karen R. Adams
1977 *Ethnobotanical Techniques and Approaches at Salmon Ruin, New Mexico*. Contributions in Anthropology, vol. 8, no. 1. Eastern New Mexico University, Portales.
- d'Alpoim Guedes, Jade, and Robert Spengler
2014 Sampling Strategies in Paleoethnobotanical Analysis. In *Method and Theory in Paleoethnobotany*, edited by Jade d'Alpoim Guedes and Christina Warner, pp. 77–94. University Press of Colorado, Boulder.
- Diehl, Michael W.
1997 Rational Behavior, the Adoption of Agriculture, and the Organization of Subsistence during the Late Archaic Period in the Greater Tucson Basin. In *Rediscovering Darwin: Evolutionary Theory and Archaeological Explanation*, edited by C. Michael Barton and Geoffrey A. Clark, pp. 251–266. Archaeological Papers No. 7. American Anthropological Association, Arlington, Virginia.
2015 Foraging Off the Gardens: Diet Breadth and Anthropogenic Landscape Change during the Early Agricultural Period in the Tucson Basin. In *Implements of Change: Tools, Subsistence, and the Built Environment of Las Capas, an Early Agricultural Community in Southern Arizona*, edited by James M. Vint, pp. 345–375. Anthropological Papers No. 51. Archaeology Southwest, Tucson, Arizona.
- Diehl, Michael W. (editor)
2005 *Subsistence and Resource Use Strategies of Early Agricultural Communities in Southern Arizona*. Anthropological Papers No. 34. Center for Desert Archaeology, Tucson, Arizona.
- Diehl, Michael W., and Owen K. Davis
2016 The Short, Unhappy Use Lives of Early Agricultural Period “Food Storage” Pits at the Las Capas Site, Southern Arizona. *American Antiquity* 81:333–344.
- Diehl, Michael W., and Jennifer A. Waters
2006 Aspects of Optimization and Risk during the Early Agricultural Period in Southeastern Arizona. In *Behavioral Ecology and the Transition to Agriculture*, edited by Douglas J. Kennett and Bruce Winterhalder, pp. 63–86. University of California Press, Berkeley.
- Drennan, Robert D.
1996 *Statistics for Archaeologists: A Commonsense Approach*. Plenum Press, New York.
- Ford, Richard I.
1979 Paleoethnobotany in American Archaeology. In *Advances in Archaeological Method and Theory*, vol. 2, edited by Michael B. Schiffer, pp. 285–336. Academic Press, New York.
1981 Gardening and Farming before A.D. 1000: Patterns of Prehistoric Cultivation North of Mexico. *Journal of Ethnobiology* 1:6–27.
2003 Foreword. In *People and Plants in Ancient Eastern North America*, edited by Paul E. Minnis, pp. vii–xvi. Smithsonian Institution, Washington, D.C.
- Gremillion, Kristen J.
1997 Introduction. In *People, Plants, and Landscapes: Studies in Paleoethnobotany*, edited by Kristen J. Gremillion, pp. 1–9. University of Alabama Press, Tuscaloosa.
- Heidke, James M.
1999 Cienega Phase Incipient Plain Ware from Southeastern Arizona. *Kiva* 64:311–228.
2005 Early Agricultural Period Pottery from Las Capas and Los Pozos. In *Material Cultures and Lifeways of Early Agricultural Communities in Southern Arizona*, edited by R. Jane Sliva, pp. 171–206. Anthropological Papers No. 35. Center for Desert Archaeology, Tucson, Arizona.
- Huckell, Lisa W.
1998 Macrobotanical Remains. In *Archaeological Investigations of Early Village Sites in the Middle Santa Cruz Valley: Analyses and Synthesis*, Part I, edited by Jonathan B. Mabry, pp. 57–148. Anthropological Papers No. 19. Center for Desert Archaeology, Tucson, Arizona.
- Kelly, Robert L.
1995 *The Foraging Spectrum: Diversity in Hunter-Gatherer Lifeways*. Smithsonian Institution Press, Washington, D.C., and London.
- Lee, Gyoung-Ah
2012 Taphonomy and Sample Size Estimation in Paleoethnobotany. *Journal of Archaeological Science* 39:648–655.
- Lennstrom, Heidi A., and Christine A. Hastorf
1995 Interpretation in Context: Sampling and Analysis in Paleoethnobotany. *American Antiquity* 60:701–721.
- Mabry, Jonathan B. (editor)
2008 *Las Capas: Early Irrigation and Sedentism in a Southwestern Floodplain*. Anthropological Papers No. 28. Center for Desert Archaeology, Tucson, Arizona.
- Miksicek, Charles H.
1987 Formation Processes of the Archaeobotanical Record. In *Advances in Archaeological Method and Theory*, vol. 10, edited by Michael B. Schiffer, pp. 211–247. Academic Press, New York.
- Minnis, Paul E.
1981 Seeds in Archaeological Sites: Sources and Some Interpretive Problems. *American Antiquity* 46:143–152.
1985 *Social Adaptation to Food Stress: A Prehistoric Southwestern Example*. University of Chicago Press, Chicago, Illinois.
2000 Introduction. In *Ethnobotany: A Reader*, edited by Paul E. Minnis, pp. 3–10. University of Oklahoma Press, Norman.
- Moerman, Daniel E.
1998 *Native American Ethnobotany*. Timber Press, Portland, Oregon.
- Pauketat, Timothy R., Lucretia S. Kelly, Gayle J. Fritz, Neal H. Lopinot, Scott Elias, and Eve Hargrave
2002 The Resides of Feasting and Public Ritual at Early Cahokia. *American Antiquity* 67:257–279.
- Pearsall, Deborah M.
1988 Interpreting the Meaning of Macroremain Abundance: The Impact of Source and Context. In *Current Paleoethnobotany: Analytical Methods and Cultural Interpretations of Archaeological Plant Remains*, edited by Christine A. Hastorf and Virginia S. Popper, pp. 97–118. University of Chicago Press, Chicago, Illinois.
- Popper, Virginia S.
1988 Selecting Quantitative Measures in Paleoethnobotany. In *Current Paleoethnobotany: Analytical Methods and Cultural Interpretations of Plant Remains*, edited by Christine A. Hastorf and Virginia S. Popper, pp. 53–71. University of Chicago Press, Chicago, Illinois.
- Popper, Virginia S., and Christine A. Hastorf
1988 Introduction. In *Current Paleoethnobotany: Analytical Methods and Cultural Interpretations of Archaeological Plant Remains*, edited by Christine A. Hastorf and Virginia S. Popper, pp. 1–16. University of Chicago Press, Chicago, Illinois.
- Schiffer, Michael B.
1987 *Formation Processes of the Archaeological Record*. University of New Mexico Press, Albuquerque.
- Shennan, Stephen
1988 *Quantifying Archaeology*. Edinburgh University Press, Edinburgh. Reprinted in 1990 by Academic Press, New York.
- Simms, Steven R.
1987 *Behavioral Ecology and Hunter-Gatherer Foraging: An Example from the Great Basin*. International Series No. 381. British Archaeological Reports, Oxford, England.
- Sliva, R. Jane (editor)
2005 *Material Cultures and Lifeways of Early Agricultural Communities in Southern Arizona*. Anthropological Papers No. 35. Center for Desert Archaeology, Tucson, Arizona.
- van Roggen, Judith
2016 Revisiting Sand Canyon Pueblo: Testing Archaeological Approaches for Sampling Ancient Plant Remains. Unpublished Ph.D. dissertation, College of Graduate Studies, University of British Columbia, Okenaga.
- Vint, James M. (editor)
2015 *Implements of Change: Tools, Subsistence, and the Built Environment of Las Capas, an Early Agricultural Irrigation Community in Southern Arizona*. Anthropological Papers No. 51. Center for Desert Archaeology, Tucson, Arizona.

- Vint, James M., and Fred L. Nials (editor)
2015 *The Anthropogenic Landscape of Las Capas, an Early Agricultural Irrigation Community in Southern Arizona*. Anthropological Papers No. 50. Center for Desert Archaeology, Tucson, Arizona.
- Whittlesey, Stephanie M., S. Jerome Hesse, and Michael S. Foster (editors)
2010 *Recurrent Sedentism and the Making of Place: Archaeological Investigations at Las Capas, a Preceramic Period Farming Community in the Tucson Basin, Southern Arizona*. Cultural Resource Report No. 07–556. SWCA Environmental Consultants, Tucson, Arizona.
- Winterhalder, Bruce, and Douglas J. Kennett
2006 Behavioral Ecology and the Transition from Hunting and Gathering to Agriculture. In *Behavioral Ecology and the Transition to Agriculture*,

- edited by Douglas J. Kennett and Bruce Winterhalder, pp. 1–21. University of California Press, Berkeley.
- Wright, Patti J.
2003 Preservation or Destruction of Plant Remains by Carbonization? *Journal of Archaeological Science* 30:577–583.

AUTHOR INFORMATION

Michael W. Diehl ■ Desert Archaeology, Inc. 3975 N. Tucson Blvd., Tucson, AZ 85716, USA (mdiehl@desert.com)