

RESEARCH ARTICLE

Algorithmic handwriting analysis of the Samaria inscriptions illuminates bureaucratic apparatus in biblical Israel

Shira Faigenbaum-Golovin^{1*}, Arie Shaus^{1,2,3*}, Barak Sober^{4*}, Eli Turkel¹, Eli Piasetzky⁵, Israel Finkelstein²

1 Department of Applied Mathematics, Tel Aviv University, Tel Aviv, Israel, **2** Jacob M. Alkow Department of Archaeology and Ancient Near Eastern Civilizations, Tel Aviv University, Tel Aviv, Israel, **3** Department of Genetics, Harvard Medical School, Boston, MA, United States of America, **4** Department of Mathematics, Duke University, Durham, NC, United States of America, **5** School of Physics and Astronomy, Tel Aviv University, Tel Aviv, Israel

* shirafaigen@gmail.com (SF-G); ashaus@post.tau.ac.il (AS); barakino@math.duke.edu (BS)



OPEN ACCESS

Citation: Faigenbaum-Golovin S, Shaus A, Sober B, Turkel E, Piasetzky E, Finkelstein I (2020) Algorithmic handwriting analysis of the Samaria inscriptions illuminates bureaucratic apparatus in biblical Israel. PLoS ONE 15(1): e0227452. <https://doi.org/10.1371/journal.pone.0227452>

Editor: Paul Dille, University of Iowa, UNITED STATES

Received: July 10, 2019

Accepted: December 18, 2019

Published: January 22, 2020

Copyright: © 2020 Faigenbaum-Golovin et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript, its Supporting Information files, and uploaded to figshare: Samaria Ancient Hebrew dataset (<https://doi.org/10.6084/m9.figshare.10266206.v1>).

Funding: The research reported here received initial funding from European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC Grant Agreement 229418, and by an Early Israel

Abstract

Past excavations in Samaria, capital of biblical Israel, yielded a corpus of Hebrew ink on clay inscriptions (ostraca) that documents wine and oil shipments to the palace from surrounding localities. Many questions regarding these early 8th century BCE texts, in particular the location of their composition, have been debated. Authorship in countryside villages or estates would attest to widespread literacy in a relatively early phase of ancient Israel's history. Here we report an algorithmic investigation of 31 of the inscriptions. Our study establishes that they were most likely written by two scribes who recorded the shipments in Samaria. We achieved our results through a method comprised of image processing and newly developed statistical learning techniques. These outcomes contrast with our previous results, which indicated widespread literacy in the kingdom of Judah a century and half to two centuries later, ca. 600 BCE.

Introduction

The question of literacy in ancient (biblical) Israel is crucial for biblical exegesis and related fields. In a recent article [1] we dealt with the corpus of ostraca (ink inscriptions on clay sherds) unearthed at the desert fortress of Arad in Judah (the southern of the two Hebrew kingdoms), dated to ca. 600 BCE. We introduced an algorithmic framework that allowed us to estimate that the 18 inscriptions investigated were authored by at least four individuals (six if one adds information provided by the texts), representing different positions in the hierarchy of Judah's military system. This indicates significant dissemination of literacy in Judah in the years before its destruction by Babylonia. In the current study we turn to a corpus of ostraca written 150–200 years earlier, uncovered in Samaria, the capital of Israel (the Northern Kingdom) [2], the stronger and more prosperous of the two Hebrew kingdoms, which competed with Damascus for domination of the Levant in the 9th and 8th centuries BCE.

grant (New Horizons project), Tel Aviv University, as well as by ISF grant 2062/18. This study was supported by a generous donation from Mr. Jacques Chahine, made through the French Friends of Tel Aviv University. B. Sober is supported through Math+X grant 400837 from the Simons Foundation, as well as Duke University.

Competing interests: The authors have declared that no competing interests exist.

According to the biblical account (1 Kings 16:24), as well as the reference to Israel in Assyrian sources, King Omri established Samaria as the capital of the kingdom in the early 9th century BCE and founded its first strong dynasty, the “House of Omri.” A second period of prosperity in Israel took place in the first half of the 8th century. Excavations at the site, conducted at the beginning of the 20th century [3], revealed a monumental and rich metropolis that lasted until the Assyrian takeover of the kingdom in 722/720 BCE.

The Samaria excavations yielded one of the richest corpora of Iron Age Hebrew inscriptions [3–5]. Based on paleographic considerations combined with information on regnal years in the texts (more below), this assemblage, comprised of ca. 100 short administrative texts, mainly ostraca (Fig 1), most probably dates to the first half of the 8th century BCE [5–8]. The ostraca were found in a fill laid in preparation for the construction of a large building (see [8]; different view in [9]), labeled by the excavators [3] as the “Ostraca House.” The inscriptions record the delivery of wine and oil from villages or royal estates in the countryside to the capital. The telegraphic texts contain details such as regnal year of a king (citing the years 9, 10 and 15; see [6–8]), toponym, name of clan (which the Bible lists in the genealogy of the tribe of Manasseh—e.g., Joshua 17:2–3), commodity type (wine/oil), and personal name, probably the sender or the recipient (for details see Table D in S1 File, example on Fig 1, and map on Fig 2). Due to its size and provenance, the Samaria corpus is crucial for reconstructing the history of the Northern Kingdom and biblical research, as well as the study of ancient Hebrew language and script.

Despite many years of extensive research, several issues related to the Samaria corpus are still debated (e.g., [11–14], summary in [9]). In particular, it is not clear whether the ostraca were composed at various sites in the highlands around the capital and dispatched to Samaria along with the provisions mentioned in them, or whether they were written in the capital, possibly when the shipments arrived. The former option would indicate dissemination of writing, at least in the administrative echelon of the kingdom of Israel, while the latter would provide evidence for the royal bureaucracy in the capital. Related issues are the number of individuals who authored the inscriptions (theoretically, there could have been itinerant scribes who traveled between royal estates), and the meaning of the information in the texts (the regnal year of the king, identification of the toponym, clan system, and function of the individuals mentioned; see Table D in S1 File). Interrelation between writers and these specific categories of information may indicate specialization within the scribes’ milieus.

In a previous article ([1], strengthened by results in [15,16]), which dealt with the corpus of ostraca from the fortress of Arad in the desert fringe of southern Judah, we introduced an algorithmic framework capable of detecting statistically significant “separations” of authors within pairs of inscriptions in the assemblage. Our former techniques allowed us to estimate the *minimal* number of writers within the investigated corpus. As will be seen below, applying our method to the Samaria ostraca yields a small number of separations, and thus a small estimate

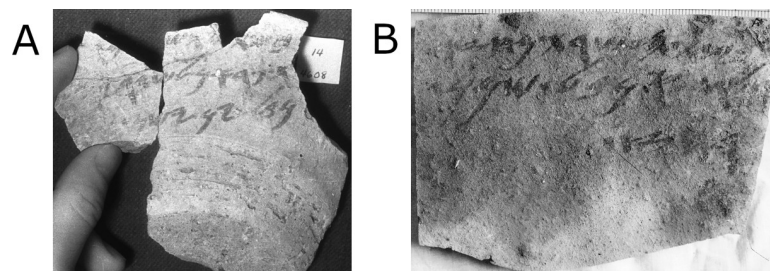


Fig 1. Examples of Samaria ostraca. (A) No. 14: “In the year ni[ne] from Az[. . .]t Par’an to Shemaryau jar of aged wine”; (B) No. 18: “In the year ten from Hazeroth to Gaddiyau jar of bath oil”.

<https://doi.org/10.1371/journal.pone.0227452.g001>

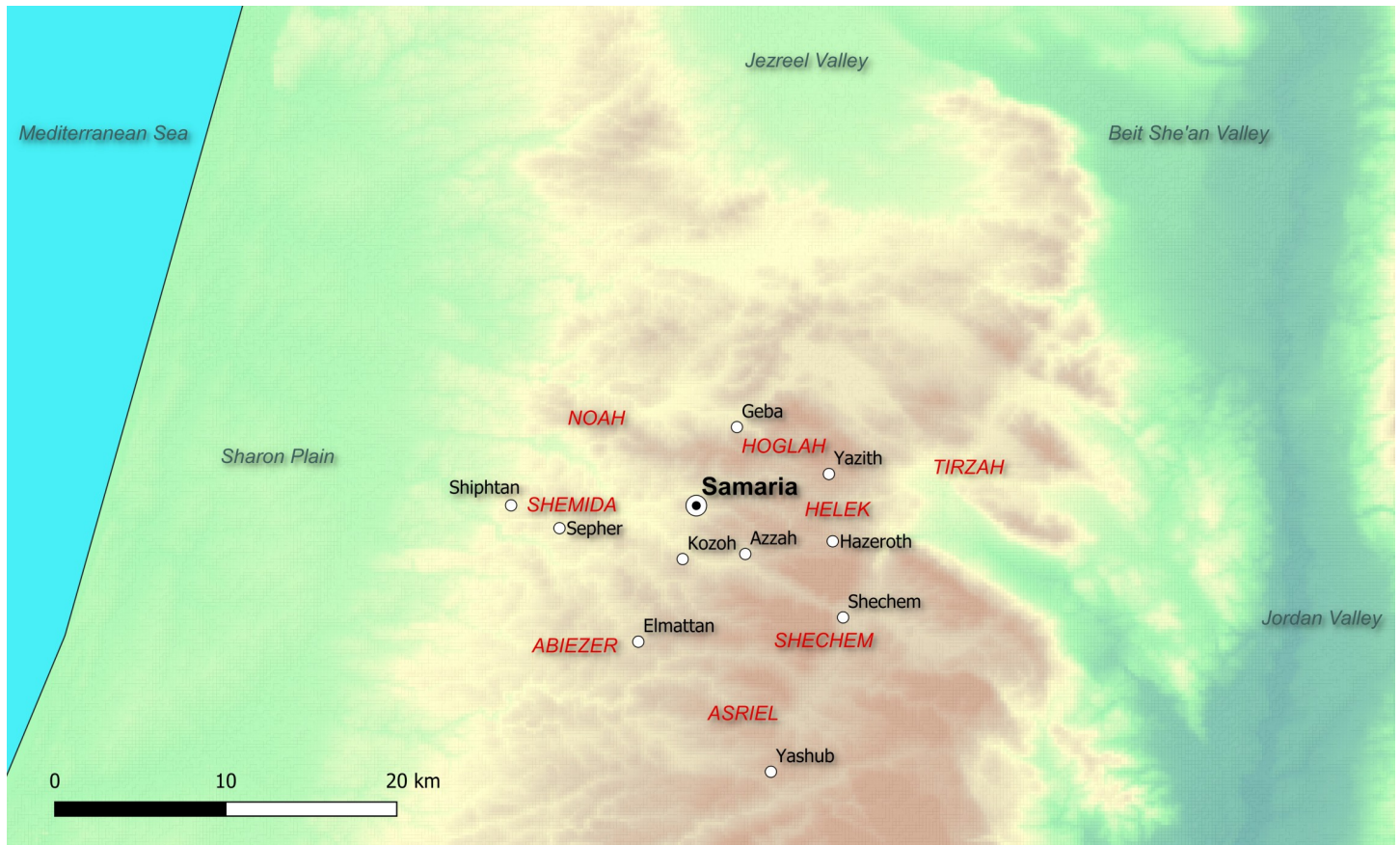


Fig 2. Map marking places and clans mentioned in the Samaria ostraca; based on [5,10]. In capital red are the names of the Manasseh tribal clans; in black are the names of identifiable localities.

<https://doi.org/10.1371/journal.pone.0227452.g002>

of the minimal number of scribes. In fact, theoretically, *all* the separations could have been obtained by chance, representing “false positives,” despite an underlying single author. In order to deal with this possibility, answering the research questions presented in the current paper necessitated a revision and enhancement of our algorithmic apparatus. The main goal of the present research is to deduce *the most likely* number of scribes in the Samaria corpus, taking into account the possibilities for both “false positive” and “false negative” writers’ separations. The most likely estimate is sufficient for shedding light on some of the fundamental issues under discussion. To the best of our knowledge, this is the first attempt to estimate the most likely number of writers at Samaria, or any other ancient corpus, via classical paleographic or computational means.

Materials and methods

Datasets

This research was conducted on two datasets of ancient written material. The main assemblage was a corpus of 39 Hebrew ostraca found in Samaria, probably dating to the early 8th century BCE. The study was performed on character reconstructions based on grayscale digital images of these inscriptions, scanned from negatives acquired by the Harvard expedition to Samaria [3] when they were unearthed at the beginning of the 20th century. All reconstructions, performed by methods specified in the [S1 File](#), Section 1, are available in [17].

Permission for research and publication of the results based on the negatives were obtained from the Harvard Semitic Museum. We refer to Texts # 2, 5, 6, 7, 8, 11, 12, 14, 15, 16a, 17a, 17b, 18, 19, 20, 21, 22, 24a, 29, 33, 34, 35, 36, 38, 40, 42, 43, 44, 45, 51, 52, 53, 54, 55, 56, 57, 59, 61 and 62, as well as the relatively short ostraca 11, 15, 17b, 33, 34, 40, 44 and 61, which were used for statistical enrichment of the algorithm. The registration numbers, and other details regarding the inscriptions, are provided in the [S1 File](#), Section 1. All necessary permits were obtained for the described study, which complied with all relevant regulations.

A second dataset, utilized for the confusion matrices estimation (see below), contained information from a corpus of 16 Hebrew ostraca found at the Arad fortress which date to ca. 600 BCE. This dataset was published in [13].

Algorithmic apparatus

Several factors can hamper an algorithmic analysis of ancient Hebrew ostraca via readily available means. First, the poor state of preservation of the ostraca ([Fig 1](#)) cannot be fully remedied by existing image acquisition methods [18–24]. Second, the imperfect digital images present a challenge for image segmentation and enhancement methods [25,26]. Third, although the task of identifying writers in handwritten texts has been addressed in previous literature (e.g., [27–32]), researchers presuppose a reference dataset with annotated authorship be used for training purposes, which is not present in our case. Additionally, the above publications do not aim directly at recognizing or distinguishing authors. Instead, they focus on finding a distance between inscriptions, which is useful only in limited scenarios (e.g., finding the *k*-most similar texts to a given one). In other words, the estimation of the number of writers in a given corpus has received little attention. In our previous publication [1], an algorithm yielding a *lower bound* for the number of authors in a group of inscriptions was introduced. Here we aim to provide the *maximum likelihood estimate* (MLE) for the number of hands in a corpus.

The algorithmic framework of this article consists of two consecutive stages. The goal of the first stage, presented in [1], is to establish separations between authors of every pair of inscriptions within the corpus. An improved version of this algorithm has been applied to the Samaria corpus. The purpose of the second, newly developed stage is to establish the most likely number of scribes within a given corpus of documents. The general idea is to obtain a number of *hands' separations* within the Samaria corpus, and then provide a statistical estimate for the number of *authors* that could have created such an observation, taking into account possible detections and non-detections errors. Below, we provide a concise description of the two stages; for further details see the [S1 File](#), Sections 1–2.

Stage I: Differentiating between authors. The algorithmic foundation of our approach was laid and verified on modern Hebrew handwriting [1]. Employing the algorithm on the ancient Arad ostraca [33] yielded a lower bound of four contemporaneous writers within a group of 18 inscriptions. The successful application of our approach on the noisy, deteriorated medium of the Arad ostraca indicated the potential for similar veins of research on other Hebrew Iron Age corpora.

The algorithm comprises a sequence of three consecutive sub-steps ([Fig 3A–3D](#)), assuming digital images of the ostraca as its input (see [S1 File](#), Section 1 for details, including certain enhancements with respect to [1], and up-to-date results on both modern handwriting and the Arad corpus; see also Tables B–C in [S1 File](#)), and operates on a character level. Throughout the article, by *character* we denote a particular instance of a given *letter* (e.g., there may be many characters, which are all occurrences of the letter *alep*).

A. **Restoring characters** (based on [34]; cf. [35,36]). The image is segmented into (often noisy) characters that are restored via a semi-automatic reconstruction procedure. The

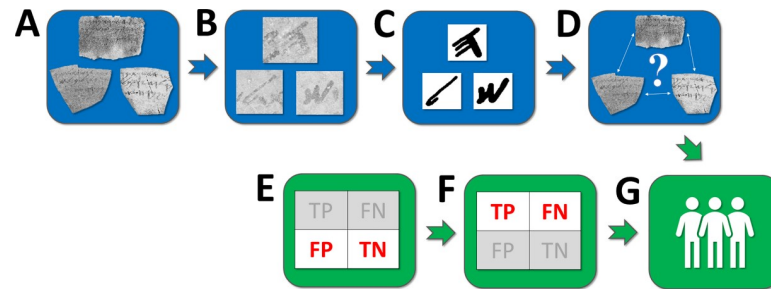


Fig 3. Basic algorithmic flow: Stage I: (A) scanned negatives of Samaria ostraca; (B) segmenting their characters; (C) restoring the characters and extracting features; (D) performing handwriting comparison. **Stage II:** (E) estimating False Positive and True Negative rates of the algorithm; (F) estimating True Positive and False Negative rates of the algorithm; (G) estimating the most likely number of scribes.

<https://doi.org/10.1371/journal.pone.0227452.g003>

purpose of the character restoration is to imitate a reed-pen's movement using several manually sampled key-points on the grayscale input images. An optimization of the pen's trajectory is performed for all intermediate sampled points. The restoration is conducted computationally via the minimization of an image energy functional, which considers the adherence to the original image, the smoothness of the stroke, as well as certain properties of the reed radius.

- B. Extraction of characters' features.** We adapted several well-known features, which describe aspects such as the character's overall shape, the angles between strokes, the character's center of gravity, as well as its horizontal and vertical projections (cf. [37]). The features, extracted from the reconstructed characters' binary images, are: SIFT [38,39] (maximal curvature points of the strokes); Zernike [40,41] (representation of the overall shape related to the human visual system); DCT [42] (representation of the overall shape related to frequency analysis); Kd-tree [42,43] (mapping based on iterative "center of gravity" calculations); image projections [44] (horizontal and vertical histograms); and L_1 and CMI [45,46,25] (both for character template matching). All these features were taken/adapted from the character recognition literature (however, not necessarily in historical documents setting), e.g., [44]. Subsequently, we exploit the pair-wise distances between all characters using all features, in order to define a vector representation of each character. As a result, the distance between two given characters is just a Euclidean distance between their vector representations.
- C. Testing the null hypothesis H_0** (for each pair of ostraca), that *two given inscriptions were written by the same author*. A corresponding p-value (P) is deduced, leveraging the data from the previous step. If $P \leq 0.1$, we reject H_0 and accept the competing hypothesis of *two different authors*; otherwise we remain undecided. The following procedure was conducted: given two inscriptions and a particular letter-type (e.g., *bet*), cluster the corresponding instances from the two inscriptions, based on their vector representations. Then, quantify the adherence of the clusters to the original inscriptions, and calculate the probability of achieving such (or better) observation. This practice is repeated for each letter-type, with the P 's combined into a single value through Fisher's method [47].

The end product of Stage I is a table containing the P for all the pair-wise comparisons of ostraca. Naturally, although our algorithm declares $P \leq 0.1$ as a case of two distinct hands, it may also signify a false positive result. Empirically, this situation occurs with probability of *much less* than the expected 0.1 (i.e., our algorithm is "conservative"). Nevertheless, false separations do exist and might even dominate the results, e.g., if the number of writers is small. As

will be seen in the Results section below, the Samaria corpus is characterized by a small number of separations (Table 1). Hence, an inevitable question arises: Is it plausible to assume that most of these separations are in fact false ones, with only a handful of writers present? Conversely, if some of the separations are true, *what is the most likely number of scribes in the Samaria corpus?*

Stage II: Estimating the most likely number of authors. The estimation of the most likely number of scribes in the Samaria corpus necessitated the development of a new statistical framework. Its sub-steps are as follows (see Fig 3E–3G; for further details, see the S1 File, Section 2):

- A. **Estimating True Negative (TN) and False Positive (FP) rates** via “same writer” simulations.
- B. **Estimating True Positive (TP) and False Negative (FN) rates** via “different writer” simulations.
- C. **Estimating the most likely number of writers** within the corpus (via a Maximum Likelihood procedure).

Steps A and B assess the empirical probabilities for TP, TN, FP, and FN separations, i.e., the *confusion matrices*, in different configurations (see the S1 File, Section 2). Naturally, the simulations estimating the confusion matrices for a given corpus (in our case, Samaria) require an independent set of documents, preferably from approximately the same period, medium, language, and script. In order to be able to evaluate the percentage of false/true detections/mis-detections of separate hands (i.e., TP, TN, FP, and FN) through Monte Carlo simulations, these inscriptions should be accompanied by pre-established separations between their authors. In this study, we consider the abovementioned Arad corpus [1,33], the richest among Hebrew Iron Age corpora, and the separations presented in the S1 File, Section 1, as the most suitable basis for our simulations. (Indeed, although at least a century and a half younger than the Samaria inscriptions, the Arad documents were written in the same script, using the same language, and utilizing the same medium—ink on clay; in fact, even the contexts of these corpora, mainly recording a supply of commodities, are rather similar). Thereafter, Step C estimates the empirical distributions of *separations*, considering scenarios of different *number of writers*. For instance, assuming that all the inscriptions were created by a single scribe, we conduct a Monte Carlo simulation in order to assess its corresponding conditional Probability Density Function (PDF), i.e., the probabilities of obtaining no separations for the whole corpus, a single separation, two separations, etc. Subsequently, assuming two writers in the corpus, we estimate another conditional PDF, and so on. In total, we estimate the conditional PDF's for scenarios ranging from a single writer to a number of writers equaling the number of inscriptions. A scenario maximizing the PDF value at the *observed number of separations*, provides us with the *MLE for the number of writers*. Moreover, under a confidence level of $1-\alpha = 0.95$, there may be other possible estimates for the number of writers.

Results

Our algorithm was applied to 31 legible Samaria texts with sufficient textual information and a low curvature of text lines (namely, texts 2, 5, 6, 7, 8, 12, 14, 16a, 17a, 18, 19, 20, 21, 22, 24a, 29, 35, 36, 38, 42, 43, 45, 51, 52, 53, 54, 55, 56, 57, 59, and 62). In addition, eight texts were used for enriching the features' statistics (11, 15, 17b, 33, 34, 40, 44, and 61; see the S1 File, Section 1). Note the double-sided ostraca 16, 17 and 24, with the *recto* denoted as “a,” and the *verso* denoted as “b.” All available letters with sufficient quantities were utilized: the Hebrew *bet*,

Table 1. Comparison between different Samaria ostraca using Stage I of the algorithmic framework.

Text	2	5	6	7	8	12	14	16a	17a	18	19	20	21	22	24a	29	35	36	38	42	43	45	51	52	53	54	55	56	57	59	62						
2			0.07	0.4		0.35	0.08	0.4	0.23	0.71	0.4		1		0.4		0.4							0.4		0.4	0.4										
5						0.6							0.4	0.16				1			1	0.95															
6	0.07				0.4	0.61	0.09	0.07	0.12	0.23			0.2								1		0.19		0.71	0.4	0.4										
7	0.4						0.03		0.6	0.27			0.16										0.27														
8			0.4			1	1						1									1		1													
12	0.35	0.6	0.61		1		0.9	0.62	0.7	0.94	0.45		0.48	0.41			0.4	1			0.91		0.6	0.35	0.39	0.91						0.6					
14	0.08		0.09	0.03	1	0.9		0.6	0.28	0.86	0.42		0.08	0.16	0.42		0.42			0.27	1	0.6	0.03	0.67	0.16	0.36	0.03				0.4						
16a	0.4		0.07			0.62	0.6		0.4	0.09			0.37	0.81								0.6		0.13		0.4											
17a	0.23		0.12	0.6	0.7	0.28	0.4			0.19	0.4		0.74	0.81	0.4	0.4	0.4				1	0.4	0.84	0.4	0.4	0.12	0.27										
18	0.71		0.23	0.27	0.94	0.86	0.09	0.19	0.19		0.58		0.25		1	1	1				0.6	0.6	0.15	0.91	0.61	0.6											
19	0.4				0.45	0.42			0.4	0.58			0.5	0.81									0.58														
20													0.91														0.4										
21	1	0.4	0.2	0.16	1	0.48	0.08	0.37	0.74	0.25	0.5	0.91		0.86	0.42	1	0.47		0.4	0.27		0.27	0.42		0.69	0.5	0.83	0.42									
22		0.16				0.41	0.16	0.81	0.81		0.81		0.86				0.42		0.16						0.81	0.81	0.81										
24a	0.4						0.42		0.4	1			0.42											1													
29													1											0.4			0.4										
35	0.4		1			0.4	0.42		0.4	1			0.47	0.42																							
36																																					
38						1							0.4	0.16																							
42													0.27														0.07										
43			1				0.27															0.27															
45			0.95		1	0.91	1	0.6	1	0.6			0.27							0.27			0.6		0.23												
51	0.4						0.03		0.4	0.6			0.42										0.6		0.6												
52						0.6																															
53	0.71		0.19	0.27	1	0.35	0.67	0.13	0.84	0.15	0.58		0.69	0.81	1	0.4				0.6	0.23	0.6			0.77	0.93	0.27										
54	0.4					0.39	0.16		0.4	0.91			0.5	0.81										0.77													
55	0.4		0.91	0.36	0.4	0.12	0.61		0.4	0.83	0.81	0.4	0.83	0.81						0.07			0.93		0.93												
56	0.4						0.03		0.27	0.6			0.42												0.27												
57			1																																		
59			1			0.6	0.4															0.4			1												
62			0.6																			0.4			1												

A $P \leq 0.1$, highlighted in red, indicates writers' separation established by our algorithm.

<https://doi.org/10.1371/journal.pone.0227452.t001>

yod, lamed, mem, nun, resh, shin, and tau. In total, 293 legible characters were restored, based upon computerized images of the inscriptions.

The complete results of Stage I of our framework, applied to the Samaria corpus, are summarized in Table 1. The ostraca numbers head the rows and columns of the table, and the intersection cells provide the comparisons' p-values. The cells with $P \leq 0.1$ are marked in red, indicating that the two ostraca are considered to be written by different authors. We reiterate that when $P > 0.1$, we cannot claim that they were written by a single author; in such a case our algorithm remains agnostic.

As seen in Table 1, most of the Samaria ostraca pairs could not have been compared (gray cells), due to insufficient letter statistics. This is caused by the brevity of the inscriptions, containing a very low number of legible restored characters for each text (9.5 characters on average). Nevertheless, 138 comparisons were performed, yielding **10 separations**.

In order to assess the most likely number of writers, Stage II of the algorithm was applied. A summary of the results can be seen in Fig 4. It depicts the conditional PDFs for different number of writers' scenarios, ranging from a single to up to five scribes (our simulations included scenarios of 1–31 writers; only graphs for 1–5 writers are presented). It can be seen that provided the previously obtained 10 separations, **the most likely estimate for the number of writers in Samaria is two**. Furthermore, out of all tested scenarios, two is the only valid estimate for the number of writers under a confidence level of $1 - \alpha = 0.95$.

The confusion matrices of Stage II, Step B were based on the separation statistics, stemming from the application of Stage I on the Arad corpus. This application (utilizing a threshold of $P = 0.1$) may have theoretically focused on data producing "excessively" significant results, while restricting the number of potentially false outcomes. In order to rule out this possibility, another simulation, in a less restrictive mode (employing a threshold of $P = 0.2$ for Arad only), was conducted. Our main results, including the MLE, were upheld even in this setting (see the S1 File, Section 2, as well as Fig A in S1 File and Table G in S1 File, for additional details).

Discussion

The maximum likelihood estimate of two writers, obtained in this research, seems to shed light on the administrative apparatus in the kingdom of Israel. As mentioned above, the 31

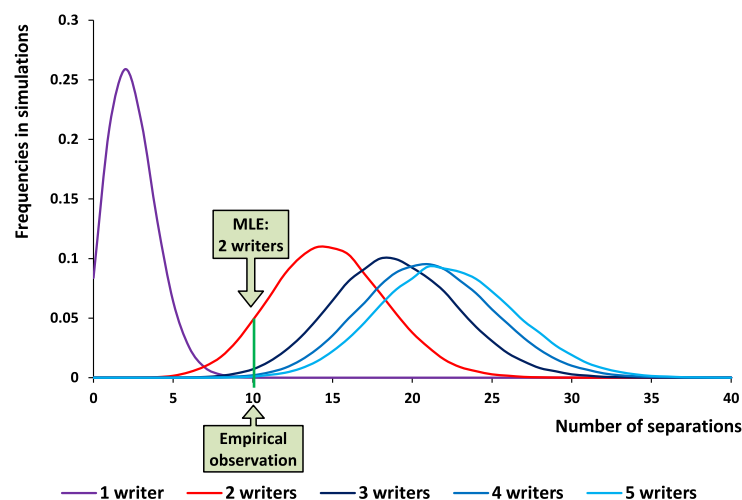


Fig 4. Conditional PDF for the number of separations in our ostraca sample, with different simulated number of writers. The MLE is two authors, based on 10 observed separations.

<https://doi.org/10.1371/journal.pone.0227452.g004>

tested Samaria ostraca, spanning a maximum of seven years (assuming a single monarch; cf. 8 which calculates one year, more below), contain various documentation characteristics (i.e., year, commodity, name of person, clan and toponym). We tried to find an interrelation between the two writers and these characteristics. Various clustering algorithms (e.g., SVM-clustering, max-cut, k-means, PCA-based methods) were applied on both the feature vector representations and final p-value tables. However, the clustering results were inconsistent and did not create even an approximate division according to any particular characteristic (e.g., year, type of commodity, etc.). Moreover, a close examination of the characteristics detailed in Table D in [S1 File](#) vs. the separations obtained in [Table 1](#), reveals that a clear-cut division of the ostraca is impossible for any characteristic (e.g., Ostraca 6 and 14 are separated by the algorithm despite mentioning the same commodity, wine; Ostraca 16 and 18 are separated although both mention oil). In other words, a hypothetical split of the inscription into two groups according to any particular characteristic cannot be achieved. As a result, and in light of the short span of regnal years mentioned (9–15, or a single year if one follows [8]), we suggest that **the two writers were contemporaneous, and performed similar duties.**

Furthermore, although the ostraca originated from various locations in the highlands around the capital and mention different clans in the region (see [Fig 2](#)), the fact that they were written by only two individuals seems to indicate that **the scribes were located in Samaria** rather than in the countryside. While we cannot rule out the possibility that the two scribes were traversing the countryside, documenting shipments on demand (cf. the “wandering scribes” of the Amarna tablets, offering their services to rulers of city-states of Late Bronze Canaan; see [48]) we see such a possibility as less plausible in the case of mundane activities (shipment of agricultural goods to differ from diplomatic correspondence) and an organized and well-governed kingdom. Note that the only contemporaneous corpus—that of Kuntillet Ajrud in the remote northern Sinai Desert [49]—is also related to the royal administration of the kingdom of Israel. This seems to attest to the existence of a centralized bureaucratic apparatus in Samaria in particular and in the Northern Kingdom in general.

One may ask why the Samaria ostraca do not refer to earlier or later years of Israelite monarchs. To answer, we first note that the ostraca were found out of stratigraphic context, in a fill prepared for the construction of the “Ostraca House” [8] and possibly even beyond [9], meaning that there is no way to trace their original provenance. Theoretically, they may represent one collection of records, while other collections may have been disposed of in other places. In any event, the dating of the Samaria ostraca can be narrowed down by taking the year 15 mentioned in some of them as being attributed to the rule of a specific king [6]. According to the biblical account, correlated with Assyrian records, only five kings ruled in Samaria for periods of 15 years or longer (Ahab 871/873–852 BCE; Jehu 842–814 BCE; Jehoahaz 817–800 BCE; Joash 800–784 BCE; and Jeroboam II 788–747 BCE (dates of Jehoahaz and Jeroboam II include coregencies). Since proto-Canaanite script (which antedates the Hebrew script) still appears in the 9th century [50,51], the days of Joash and Jeroboam II are the most likely. Judging from the prosperity of the kingdom, the latter option is the most plausible, though the fact that the latest year referred to is 15 may point to the former (Jeroboam II ruled for over 40 years, while his father Joash reigned for 16 years). One may hypothesize that the Samaria ostraca represent an important phase in the bureaucracy of the kingdom. During the days of Joash or the first years of Jeroboam II, Hebrew writing had already been sufficiently developed to enable recording on ostraca; a while later, during the peak prosperity of the kingdom, the system could have changed to a more efficient recording system, perhaps using papyri.

As mentioned above (see also [52]), there is a high level of standardization in the format of the Samaria texts, which may support the proposed bureaucratic apparatus, perhaps even administrative centralization. Nevertheless, at some point between the years 10 and 15 a

change in the documentation formula occurred: ostraca belonging to year 9 or 10 contain commodity type and neglect the clan feature, whereas ostraca belonging to year 15 neglect the commodity type and contain the clan name. As already noted, we could not associate years 9–10 with one scribe and 15 with the other. Therefore, since there are two contemporaneous scribes, this change may be attributed to a new/different administration directive, rather than to the scribes' preferences. In addition, there is a noticeable increase in the number of inscriptions pertaining to year 15. Explicitly, there are 9 inscriptions bearing the year 9; 14 inscriptions bearing the year 10; and 29 inscriptions bearing the year 15. If the sample we have is representative, this may indicate increased activity during the later years of the given monarch/s.

Another notable consequence of the current research pertains to the field of paleography. Apart from the two Samaria scribes, the only significant pieces of evidence for writing with ink on clay-sherds in the Northern Kingdom are two inscriptions from Beit-Shean (20 km from Samaria; see [53]) and the inscriptions from Kuntillet Ajrud in the Sinai Desert (~250 km from Samaria; see [49]). The latter site is short-lived and small (a single building), therefore, the number of scribes there must have been restricted. Archaeologically (ceramic evidence for Beth-Shean and Kuntillet Ajrud) and paleographically (Samaria), these inscriptions all date to the same period. They represent a handful of scribes and hence—contra to conventional wisdom—it is doubtful if one can construct a reliable paleographic system based on these finds.

Contrasting the epigraphic evidence presented above, covering the territory of the kingdom of Israel in the early 8th century BCE, with the level of literacy in the kingdom of Judah ca. 600 BCE based on our study of the Arad ostraca ([1], acknowledging the different natures of the two corpora—military correspondence versus receipts of shipments of agricultural commodities), an interesting tendency appears. On the one hand, we observe just two scribes within the large Samaria ink ostraca corpus of the flourishing Northern Kingdom's capital, with very little supporting evidence of writing skills from other sites in the realm. This may hint that during this period literacy was, to some extent, restricted to the royal court (note that Kuntillet Ajrud is apparently also related to the kingdom's administration). On the other hand, in a small desert outpost far from the center of Judah of the late 7th century, at least six contemporaneous writers are attested [1]. Indeed, thriving scribal activity in Judah is demonstrated by other, contemporaneous corpora, such as those from Lachish [54], Horvat 'Uza, Horvat Radum [55] and Tel Malhata [56], as well as individual ostraca scattered far and wide across the kingdom, e.g., the Ophel in Jerusalem [57], Meza Hashavyahu on the coast [58], and Nahal Yarmut in the Shephelah [59]. In other words, over the course of the century and a half or two centuries that separate the two corpora, we observe development from a writing milieu centered mainly around the royal court to a broad proliferation of literacy.

Supporting information

S1 File. Supplementary material.
(PDF)

Acknowledgments

This research was made possible by the dedicated reconstruction work of Ms. Ma'ayan Mor. The kind assistance of Dr. Adam J. Aja, Ms. Anna Axenov Shaus, Dr. Tatjana Barazon, Ms. Sivan Einhorn, Ms. Noa Evron, Dr. Joseph A. Greene, Dr. Assaf Kleiman, Dr. Eythan Levy, Dr. Anat Mendel-Geberovich, Ms. Myrna Pollak, Prof. Christopher A. Rollston and Prof. David Steinberg is greatly appreciated. We also thank the PLOS ONE editor and the

anonymous reviewers for their helpful comments and suggestions. Ostraca images are courtesy of the Semitic Museum, Harvard University.

Author Contributions

Conceptualization: Shira Faigenbaum-Golovin, Arie Shaus, Barak Sober, Eli Piasezky, Israel Finkelstein.

Formal analysis: Shira Faigenbaum-Golovin, Arie Shaus, Barak Sober.

Funding acquisition: Eli Piasezky, Israel Finkelstein.

Investigation: Shira Faigenbaum-Golovin, Arie Shaus, Barak Sober.

Methodology: Shira Faigenbaum-Golovin, Arie Shaus, Barak Sober.

Resources: Eli Piasezky, Israel Finkelstein.

Software: Shira Faigenbaum-Golovin, Arie Shaus, Barak Sober.

Supervision: Eli Turkel, Eli Piasezky, Israel Finkelstein.

Writing – original draft: Shira Faigenbaum-Golovin, Arie Shaus, Barak Sober.

Writing – review & editing: Shira Faigenbaum-Golovin, Arie Shaus, Barak Sober, Eli Turkel, Eli Piasezky, Israel Finkelstein.

References

1. Faigenbaum-Golovin S, Shaus A, Sober B, Levin D, Na'aman N, Sass B, et al. Algorithmic handwriting analysis of Judah's military correspondence sheds light on composition of biblical texts. *Proc Natl Acad Sci USA*. 2016; 113: 4664–4669. <https://doi.org/10.1073/pnas.1522200113> PMID: 27071103
2. Finkelstein I. *The Forgotten Kingdom. The Archaeology and History of Northern Israel*. Atlanta: Society of Biblical Literature. 2013.
3. Reisner GA, Fischer CS, Lyon DG. *Harvard Excavations at Samaria*. Cambridge, MA: Harvard University Press. 1924.
4. Ahituv S. *Echoes from the Past: Hebrew and Cognate Inscriptions from the Biblical Period*. Jerusalem: Carta. 2008.
5. Aharoni Y. *The Land of the Bible: A Historical Geography*. Louisville, KY: Westminster John Knox Press. 1979.
6. Lemaire A. *Les ostraca hébreux de l'époque royale Israélite*. Paris, 1973.
7. Lemaire A. *Inscriptions hébraïques, Vol. 1: Les ostraca*. *Littératures anciennes du Proche-Orient* 9. Paris: Cerf. 1977.
8. Rainey AF. Toward a precise date for the Samaria ostraca. *Bull Am Schools Orient Res*. 1988; 272: 69–74.
9. Tappy RE. *The Archaeology of the Ostraca House at Israelite Samaria: Epigraphic Discoveries in Complicated Contexts*. ASOR 70. Boston: American Schools of Oriental Research. 2016.
10. USGS EROS (Earth Resources Observatory and Science (EROS) Center) (public domain): <https://www.usgs.gov/centers/eros>
11. Yadin Y. Recipients or owners? A note on the Samaria ostraca. *Isr Explor J*. 1959; 9: 184–187.
12. Rendsburg G. *Diglossia in Ancient Hebrew*. American Oriental Series 72. New Haven, CT: American Oriental Society. 1990.
13. Kaufman IT. Samaria ostraca. In Freedman DN, editor. *The Anchor Bible Dictionary*, Vol. 5. New York, NY: Doubleday. 1992. pp 921–926.
14. Niemann HM. A new look at the Samaria ostraca: The king-clan relationship. *Tel Aviv*. 2008; 35: 249–266.
15. Shaus A, Turkel E. Writer identification in modern and historical documents via binary pixel patterns, Kolmogorov-Smirnov test and Fisher's method. *J Imaging Sci Technol*. 2017; 61: 010404–1–010404–9. <https://doi.org/10.2352/J.ImagingSci.Technol.2017.61.1.010404>

16. Faigenbaum-Golovin S, Levin D, Piasetzky E, Finkelstein I. Writer characterization and identification of short modern and historical documents: reconsidering paleographic tables. *Proceedings of the 19th ACM Symposium on Document Engineering (DocEng 2019)*, 23:1–23:4.
17. Samaria Ancient Hebrew dataset, figshare (CC BY 4.0 license): <https://doi.org/10.6084/m9.figshare.10266206.v1>
18. Faigenbaum S, Sober B, Shaus A, Moinester M, Piasetzky E, Bearman G, et al. Multispectral images of ostraca: Acquisition and analysis. *J Archaeol Sci*. 2012; 39: 3581–3590. <https://doi.org/10.1016/j.jas.2012.06.013>
19. Sober B, Faigenbaum S, Beit-Arieh I, Finkelstein I, Moinester M, Piasetzky E, et al. Multispectral imaging as a tool for enhancing the reading of ostraca. *Palestine Exploration Quarterly*. 2004; 146: 185–197. <https://doi.org/10.1179/0031032814Z.000000000101>
20. Faigenbaum-Golovin S, Mendel-Geberovich A, Shaus A, Sober B, Cordonsky M, Levin D, et al. Multispectral imaging reveals biblical-period inscription unnoticed for half a century. *PLOS ONE*. 2017; 12: e0178400. <https://doi.org/10.1371/journal.pone.0178400>
21. Mendel-Geberovich A, Shaus A, Faigenbaum-Golovin S, Sober B, Cordonsky M, Piasetzky E, et al. A brand new old inscription: Arad ostracon 16 rediscovered via multispectral imaging. *Bull Am Schools Orient Res*. 2017; 378: 113–125. <https://doi.org/10.5615/bullamerschoorie.378.0113>
22. Faigenbaum S, Sober B, Finkelstein I, Moinester M, Piasetzky E, Shaus A, et al. Multispectral imaging of two Hieratic inscriptions from Qubur el-Walaydah. *Egypt and the Levant*. 2014; 24: 349–353. <https://doi.org/10.1553/s349>
23. Faigenbaum S, Sober B, Moinester M, Piasetzky E, Bearman G. Multispectral imaging of Tel Malhata ostraca. In Beit-Arieh I, Freud L, editors. *Tel Malhata: A Central City in the Biblical Negev*. Tel Aviv: Tel Aviv University. 2015. pp. 510–513.
24. Shaus A, Sober B, Tzang O, Ioffe Z, Cheshnovsky O, Finkelstein I. Raman binary mapping of Iron Age ostracon in unknown material composition and high fluorescence setting—A proof of concept, *Archaeometry*. 2019; 61, 459–469. <https://doi.org/10.1111/arcm.12419>
25. Shaus A, Turkel E, Piasetzky E, Binarization of First Temple period inscriptions—Performance of existing algorithms and a new registration based scheme. *Proceedings of the 13th International Conference on Frontiers in Handwriting Recognition (ICFHR 2012)*, 641–646. <https://doi.org/10.1109/ICFHR.2012.187>
26. Shaus A, Sober B, Turkel E, Piasetzky E. Improving binarization via sparse methods. *Proceedings of the 16th International Graphonomics Society Conference (IGS 2013)*, 163–166.
27. Bar-Yosef I, Beckman I, Kedem K, Dinstein I. Binarization, character extraction, and writer identification of historical Hebrew calligraphy documents. *Int J Doc Anal Recognit*. 2007; 9: 89–99.
28. Bulacu M, Schomaker L. Text-independent writer identification and verification using textural and allo-graphic features. *IEEE Trans Pattern Anal Mach Intell*. 2007; 29: 701–717. <https://doi.org/10.1109/TPAMI.2007.1009> PMID: 17299226
29. Panagopoulos M, Papaodysseus C, Rousopoulos P, Dafi D, Tracy S. Automatic writer identification of ancient Greek inscriptions. *IEEE Trans Pattern Anal Mach Intell*. 2009; 31: 1404–1414. <https://doi.org/10.1109/TPAMI.2008.201> PMID: 19542575
30. Fecker D, Asit A, Märgner V, El-Sana J, Fingscheidt T, Writer identification for historical Arabic documents. *Proceedings of the 22nd International Conference on Pattern Recognition (ICPR 2014)*, 3050–3055.
31. Dhali M, He S, Popovic M, Tigchelaar E, Schomaker L. A digital palaeographic approach towards writer identification in the Dead Sea scrolls. *Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2017)*, 693–702.
32. Dahllöf M, Automatic scribe attribution for medieval manuscripts. *Digital Medievalist*. 2018; 11: 1–26.
33. Aharoni Y, *Arad Inscriptions*. Jerusalem: Israel Exploration Society. 1981.
34. Sober B, Levin D. Computer aided restoration of handwritten character strokes. *Computer-Aided Design*. 2017; 89: 12–24.
35. Mumford D, Shah J. Optimal approximations by piecewise smooth functions and associated variational problems. *Commun Pure Appl Math*. 1989; 42: 577–685.
36. Kass M, Witkin A, Terzopoulos D. Snakes: Active contour models. *Int J Comput Vis*. 1988; 1: 321–331.
37. Rollston CA. *The script of Hebrew Ostraca of the Iron Age: 8th–6th centuries BCE*, PhD thesis, Johns Hopkins University, Baltimore. 1999.
38. Lowe DG. Distinctive image features from scale-invariant keypoints. *Int J Comput Vis*. 2004; 60: 91–110.

39. Markus D, Sablatnig R. Recognition of degraded handwritten characters using local features. Proceedings of the 10th International Conference on Document Analysis and Recognition (ICDAR 2009), 221–225.
40. Tahmasbi A, Saki F, Shokouhi SB. Classification of benign and malignant masses based on Zernike moments. *Comput Biol Med* 41. 2011; 726–735. <https://doi.org/10.1016/j.compbiomed.2011.06.009> PMID: 21722886
41. Tahmasbi A. Zernike moments. 2012. Available at www.mathworks.com/matlabcentral/fileexchange/38900-zernike-moments.
42. Sexton A, Todman A, Woodward K. Font recognition using shape-based quadtree and kd-tree decomposition. Proceedings of the 3rd International Conference on Computer Vision, Pattern Recognition and Image Processing (CVPRIP 2000), 212–215.
43. Armon S. Descriptor for shapes and letters (feature extraction). 2012. Available at www.mathworks.com/matlabcentral/fileexchange/35038-descriptor-for-shapes-andletters-feature-extraction.
44. Trier ØD, Jain AK, Taxt T. Feature extraction methods for character recognition—A survey. *Pattern Recognit.* 1996; 29: 641–662.
45. Shaus A, Turkel E, Piasetzky E. Quality evaluation of facsimiles of Hebrew First Temple period inscriptions. Proceedings of the 10th IAPR International Workshop on Document Analysis Systems (DAS 2012), 170–174. <https://doi.org/10.1109/DAS.2012.70>
46. Shaus A, Faigenbaum-Golovin S, Sober B, Turkel E, Piasetzky E. Potential contrast—A new image quality measure. In Proceedings of the IS&T International Symposium on Electronic Imaging 2017, Image Quality and System Performance XIV Conference (IQSP 2017), 52–58. <https://doi.org/10.2352/ISSN.2470-1173.2017.12.IQSP-226>
47. Fisher RA. *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd. 1925.
48. Goren Y, Finkelstein I, Na'aman N. *Inscribed in Clay: Provenance Study of the Amarna Tablets and other Ancient Near Eastern Texts*. Tel Aviv University Monograph Series 28. Tel Aviv: Tel Aviv Univ. 2004.
49. Meshel Z. *Kuntilet 'Ajrud (Horvat Teman): An Iron Age II Religious Site on the Judah-Sinai Border*. Jerusalem: Israel Exploration Society. 2012.
50. Finkelstein I, Sass B. The West Semitic alphabetic inscriptions, Late Bronze II to Iron IIA: Archeological context, distribution and chronology. *Hebrew Bible and Ancient Israel*. 2013; 2: 149–220.
51. Sass B, Finkelstein I. The swan-song of Proto-Canaanite in the ninth century BCE in light of an alphabetic inscription from Megiddo. *Semitica et Classica*. 2016; 9: 19–42.
52. Mendel A. *Epigraphic lists in Israel and its neighbors in the First Temple period*, PhD thesis, Hebrew University, Jerusalem. 2014.
53. Mazar A. *Excavations at Tel Beth-Shean 1989–1996: Vol. I: From the Late Bronze Age IIB to the Medieval Period*. The Beth-Shean Valley Archaeological Project Publications 1. Jerusalem: Israel Exploration Society. 2006.
54. Torczyner H. *Lachish I: The Lachish Letters*. London and New York: Oxford University Press. 1938.
55. Beit-Arieh I. *Horvat 'Uza and Horvat Radum: Two Fortresses in the Biblical Negev*. Tel Aviv University Monograph Series 25. Tel Aviv: Tel Aviv University. 2007.
56. Beit-Arieh I, Freud L. *Tel Malhata: A Central City in the Biblical Negev*. Tel Aviv University Monograph Series 32. Tel Aviv: Tel Aviv University. 2015.
57. Faigenbaum-Golovin S, Rollston CA, Piasetzky E, Sober B, Finkelstein I. The Ophel (Jerusalem) ostracum in light of new multispectral images. *Semitica*. 2015; 57: 113–137.
58. Naveh J. A Hebrew letter from the seventh century B.C. *Isr Explor J*. 1960; 10: 129–139.
59. Mendel-Geberovich A, Faigenbaum-Golovin S, Shaus A, Sober B, Cordonsky M, Piasetzky E, et al. A renewed reading of Hebrew ostraca from Cave A-2 at Ramat Beit Shemesh (Nahal Yarmut), based on multispectral imaging. *Vetus Testamentum*. 2019; 69: 682–701. <https://doi.org/10.1163/15685330-00001370>