Algorithmic handwriting analysis of Judah's military correspondence sheds light on composition of biblical texts

Shira Faigenbaum-Golovin[a,1,2], Arie Shaus[a,1,2], Barak Sober[a,1,2], David Levin[a], Nadav Na'aman[b], Benjamin Sass[c], Eli Turkel[a], Eli Piasetzky[d], and Israel Finkelstein[c]

[a]Department of Applied Mathematics, Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel; [b]Department of Jewish History, Tel Aviv University, Tel Aviv 69978, Israel; [c]Jacob M. Alkow Department of Archaeology and Ancient Near Eastern Civilizations, Tel Aviv University, Tel Aviv 69978, Israel; and [d]School of Physics and Astronomy, Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel

Edited by Klara Kedem, Ben-Gurion University, Be'er Sheva, Israel, and accepted by the Editorial Board March 3, 2016 (received for review November 17, 2015)

The relationship between the expansion of literacy in Judah and composition of biblical texts has attracted scholarly attention for over a century. Information on this issue can be deduced from Hebrew inscriptions from the final phase of the first Temple period. We report our investigation of 16 inscriptions from the Judahite desert fortress of Arad, dated *ca*. 600 BCE—the eve of Nebuchadnezzar's destruction of Jerusalem. The inquiry is based on new methods for image processing and document analysis, as well as machine learning algorithms. These techniques enable identification of the minimal number of authors in a given group of inscriptions. Our algorithmic analysis, complemented by the textual information, reveals a minimum of six authors within the examined inscriptions. The results indicate that in this remote fort literacy had spread throughout the military hierarchy, down to the quartermaster and probably even below that rank. This implies that an educational infrastructure that could support the composition of literary texts in Judah already existed before the destruction of the first Temple. A similar level of literacy in this area is attested again only 400 y later, *ca*. 200 BCE.

biblical exegesis | literacy level | Arad ostraca | document analysis | machine learning

Based on biblical exegesis and historical considerations scholars debate whether the first major phase of compilation of biblical texts in Jerusalem took place before or after the destruction of the city by the Babylonians in 586 BCE (e.g., ref. 1). A related—and also disputed—issue is the level of literacy, that is, the basic ability to communicate in writing, especially in the Hebrew kingdoms of Israel and Judah (2). The best way to answer this question is to look at the material evidence: the corpus of inscriptions that originated from archaeological excavations (e.g., ref. 3). Inscriptions citing biblical texts, or related to them, are rarely found (for two Jerusalem amulets possibly dating to this period, echoing the priestly blessing in Numbers 6:23–26, see refs. 4 and 5), probably because papyrus and parchment are not well preserved in the climate of the region. However, ostraca (inscriptions in ink on ceramic sherds) that deal with more mundane issues can also shed light on the volume and quality of writing and on the recognition of the power of the written word in the society.

To explore the degree of literacy and stage setting for compilation of literary texts in monarchic Judah, we turned to Hebrew ostraca from the final days of the kingdom, before its destruction by Nebuchadnezzar in 586 BCE and the deportation of its elite to Babylonia. Several corpora of inscriptions exist for this period. We focused on the corpus of over 100 Hebrew ostraca found at the fortress of Arad, located in arid southern Judah, on the border of the kingdom with Edom (see ref. 6 and Fig. 1). The inscriptions contain military commands regarding movement of troops and provision of supplies (wine, oil, and flour) set against the background of the stormy events of the final years before the fall of Judah. They include orders that came to the fortress of Arad from higher echelons in the Judahite military system, as well as correspondence with neighboring forts. One of the inscriptions mentions "the King of Judah" and another "the house of YHWH," referring to the Temple in Jerusalem. Most of the provision orders that mention the *Kittiyim*—apparently a Greek mercenary unit (7)—were found on the floor of a single room. They are addressed to a person named Eliashib, the quartermaster in the fortress. It has been suggested that most of Eliashib's letters involve the registration of about one month's expenses (8).

Of all of the corpora of Hebrew inscriptions, Arad provides the best set of data for exploring the question of literacy at the end of the first Temple period: (*i*) The lion's share of the corpus represents a short time span of a few years *ca*. 600 BCE; (*ii*) it comes from a remote region of the kingdom, where the spread of literacy is more significant than its dissemination in the capital; and (*iii*) it is connected to Judah's military administration and hence bureaucratic apparatus. Identifying the number of "hands" (i.e., authors) involved in this corpus can shed light on the

Significance

Scholars debate whether the first major phase of compilation of biblical texts took place before or after the destruction of Jerusalem in 586 BCE. Proliferation of literacy is considered a precondition for the creation of such texts. Ancient inscriptions provide important evidence of the proliferation of literacy. This paper focuses on 16 ink inscriptions found in the desert fortress of Arad, written *ca*. 600 BCE. By using novel image processing and machine learning algorithms we deduce the presence of at least six authors in this corpus. This indicates a high degree of literacy in the Judahite administrative apparatus and provides a possible stage setting for compilation of biblical texts. After the kingdom's demise, a similar literacy level reemerges only *ca*. 200 BCE.

Author contributions: S.F.-G., A.S., and B. Sober designed research; S.F.-G., A.S., and B. Sober performed research; S.F.-G., A.S., and B. Sober contributed new reagents/analytic tools; D.L. and E.T. supervised the development of the algorithms; N.N., B. Sass, and I.F. provided archaeological and epigraphical analysis and historical reconstruction; E.P. supervised the development of the algorithms; S.F.-G., A.S., and B. Sober analyzed data; S.F.-G., A.S., B. Sober, D.L., N.N., B. Sass, E.T., E.P., and I.F. wrote the paper; and E.P. and I.F. headed the research team.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. K.K. is a guest editor invited by the Editorial Board.

Data deposition: Two datasets are provided on our institutional website, with free and open access: www-nuclear.tau.ac.il/~eip/ostraca/DataSets/Modern_Hebrew.zip and www-nuclear.tau.ac.il/~eip/ostraca/DataSets/Arad_Ancient_Hebrew.zip.

[1]S.F.-G., A.S., and B. Sober contributed equally to this work.

[2]To whom correspondence may be addressed. Email: shirafaigen@gmail.com, ashaus@post.tau.ac.il, or baraksov@post.tau.ac.il.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1522200113/-/DCSupplemental.

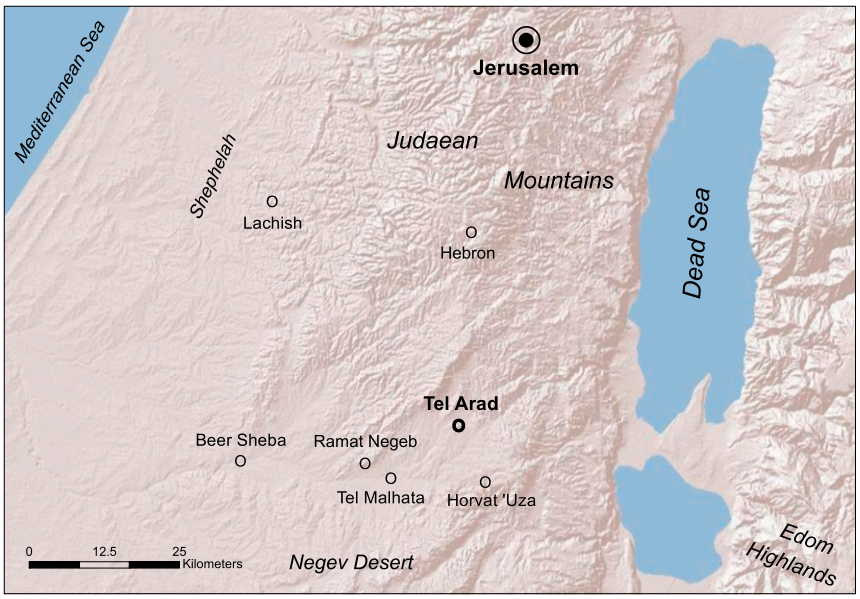

Fig. 1. Main towns in Judah and sites in the Beer Sheba Valley mentioned in the article.

dissemination of writing, and consequently on the spread of literacy in Judah.

Algorithmic Apparatus

One might try to use existing computerized algorithms for automatic handwriting comparison purposes. However, an algorithmic analysis of the Arad corpus via readily available means is hampered by several factors. First, the poor state of preservation of the ostraca (Fig. 2) could not be remedied by existing image acquisition methods (9, 10). Second, the imperfect digital images present a challenge for image segmentation and enhancement methods (11, 12). Finally, recognizing hands via document analysis algorithms is a tantalizing problem even in a modern writing setting (13). Consequently, we developed new methods for image processing and document analysis, as well as machine learning algorithms. These techniques allow us to identify the minimal number of authors represented in a given group of ostraca.

Our algorithmic sequence consisted of three consecutive stages, operating on digital images of the ostraca (see *Supporting Information*). All of the stages are fully automatic, with the exception of the first, which is a semiautomatic step.

i) Restoring characters (see example in Fig. 3; also see *Supporting Information* and ref. 14)

ii) Extraction of characters' features, describing their different aspects (e.g., angles between strokes and character profiles), and measuring the similarity ("distances") between the characters' feature vectors.

iii) 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 $P \leq 0.2$, we reject H_0 and accept the competing hypothesis of two different authors; otherwise, we remain undecided.

The end product is a table containing the P for a comparison of each pair of ostraca. Before implementing our methodology on the Arad corpus, it was thoroughly tested on modern Hebrew handwritings and found solid (see *Supporting Information* for details).

Results

Using this computerized procedure we analyzed 16 inscriptions from the Arad fortress (namely, ostraca 1, 2, 3, 5, 7, 8, 16, 17, 18,

Fig. 2. Ostraca from Arad (see ref. 6): numbers 24 (*A*), 5 (*B*), and 40 (*C*). The poor state of preservation, including stains, erased characters, and blurred text, is evident. Images are courtesy of the Institute of Archaeology, Tel Aviv University, and of the Israel Antiquities Authority.

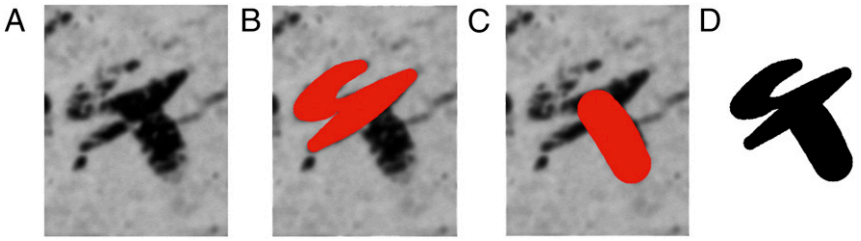

Fig. 3. Restoration of the character *waw* in Arad ostracon 24 (see ref. 14). (*A*) The original image. (*B* and *C*) reconstructed strokes. (*D*) The resulting character restoration (see *Supporting Information* for further details). Images are courtesy of the Institute of Archaeology, Tel Aviv University, and of the Israel Antiquities Authority.

21, 24, 31, 38, 39, 40, and 111), which are relatively legible and have a sufficient number of characters for examination. Two of the inscriptions (ostraca 17 and 39) are inscribed on both sides of the sherd, bringing the number of texts under investigation to 18. The results are summarized in Table 1. The ostraca numbers head the rows and columns of the table, with the intersection cells providing the comparisons' P. The cells with $P \leq 0.2$ are marked in red, indicating that the two ostraca are considered to be written by different authors. We reiterate that when $P > 0.2$ we cannot claim that they were written by a single author.

The results allow us to estimate the minimal number of writers in the tested inscriptions. For example, the examination of ostraca 7, 18, 24, and 40 reveals that their authors are pairwise distinct; in fact, six such "quadruplets" can be identified in Table 1, rendering the existence of at least four authors as highly likely; see *Supporting Information* for details. Therefore, based on the statistical analysis, it can be deduced that there are at least four unique hands in the tested corpus. Our algorithmic observations can be further supplemented by the textual and archaeological context of the ostraca, deliberately avoided until this point. In particular, the prosaic lists of names in ostraca 31 and 39* were most likely composed at Arad, as opposed to ostraca 7, 18, 24, and 40, which were probably dispatched from other locations.[†] As per the table, ostracon 31 differs from both sides of ostracon 39; we can thus conjecture an existence of two additional authors, totaling at least six distinct writers.

Discussion

Identifying the military ranks of the authors can provide information regarding the spread of literacy within the Judahite army. Our proposed reconstruction of the hierarchical relations between the signees and the addressees of the examined inscriptions is as follows[‡] (see Fig. 4):

i) The King of Judah: mentioned in ostracon 24 as dictating the overall military strategy
ii) An unnamed military commander: the author of ostracon 24
iii) Malkiyahu, the commander of the Arad fortress: mentioned in ostracon 24 and the recipient of ostracon 40[§]
iv) Eliashib, the quartermaster of the Arad fortress: the addressee of ostraca 1–16 and 18; mentioned in ostracon 17a; the writer of ostracon 31
v) Eliashib's subordinate: addressing Eliashib as "my lord" in ostracon 18

Following this reconstruction, it is reasonable to deduce the proliferation of literacy among the Judahite army ranks *ca*. 600 BCE. A contending claim that the ostraca were written by professional scribes can be dismissed with two arguments: the existence of two distinct writers in the tiny fortress of Arad (authors of ostraca 31 and 39) and the textual content of the inscriptions: Ostracon 1 orders the recipient (Eliashib) "write the name of the day," ostracon 7 commands "and write it before you. . .," and in ostracon 40 (reconstructions in refs. 6 and 18) the author mentions that he had written the letter. Thus, rather than implying the existence of scribes accompanying every Judahite official, the written evidence suggests a high degree of literacy in the entire Judahite chain of command.

The dissemination of writing within the Judahite army around 600 BCE is also confirmed by the existence of other military-related corpora of ostraca, at Horvat ʿUza (19) and Tel Malḥata (20) in the vicinity of Arad, and at Lachish[¶] in the Shephelah (summary in ref. 3)—all located on the borders of Judah (Fig. 1). We assume that in all these locations the situation was similar to Arad, with even the most mundane orders written down occasionally. In other words, the entire army apparatus, from high-ranking officials to humble vice-quartermasters of small desert outposts far from the center, was literate, in the sense of the ability to communicate in writing.

To support this bureaucratic apparatus, an appropriate educational system must have existed in Judah at the end of the first Temple period (2, 21–23). Additional evidence supporting writing awareness by the lowest echelons of society seems to come from the Meẓad Hashavyahu ostracon (24), which contains a complaint by a corvée worker against one of his overseers (most scholars agree that it was composed with the aid of a scribe).

Extrapolating the minimum of six authors in 16 Arad ostraca to the entire Arad corpus, to the whole military system in the southern Judahite frontier, to military posts in other sectors of the kingdom, to central administration towns such as Lachish, and to

*Contrary to the excavator's association of ostraca 31 and 39 with Stratum VII (ref. 6, also ref. 15) rather than VI where most of the examined ostraca were found, we agree with critics (16, 17) that these strata are in fact one and the same. Note that ostracon 31 was found in locus 779, alongside three seals of Eliashib (the addressee of ostraca 1–16 and 18, from Strata VI).

[†]Ostraca 5, 7, 17a, 18, and 24 were most probably written in other locations (6). Ostracon 40 may have been written by troop commanders Gemaryahu and Nehemyahu (see the following note) with some ties to Arad fortress; their names also appear at ostracon 31. This renders the common authorship of ostraca 31 and 40 unlikely. Furthermore, from Table 1, ostraca 40 and 39a have different authors.

[‡]We conjecture that the status of the officers who commanded the supplies to the Kittiyim (the Greek or Cypriot mercenary unit), who wrote ostraca 1–8 and 17a, was similar to that of Malkiyahu (the commander of the fortress at Arad), and in any case they were Eliashib's superiors. Also note that Gemaryahu and Nehemyahu (ostracon 40) are Malkiyahu's subordinates, whereas Hananyahu (author of ostracon 16, also mentioned in ostracon 3) is probably Eliashib's counterpart in Beer Sheba. The textual content of the ostraca also suggests differentiation between combatant and logistics-oriented officials (Fig. 4).

[§]Contrary to the excavator's dating of ostracon 40 to Stratum VIII of the late 8th century (ref. 6, also ref. 17), it should probably be placed a century later, along with ostracon 24 (see ref. 18 for details). Note that a conflict between the vassal kingdoms of Judah and Edom, seemingly hinted at in this inscription, is unlikely under the strong rule of the Assyrian empire in the region (*ca*. 730–630 BCE), especially along the vitally important Arabian trade routes.

[¶]In fact, Lachish ostracon 3, also containing military correspondence, represents the most unambiguous evidence of a writing officer. The author seems offended by a suggestion that he is assisted by a scribe. See detail, including discussion regarding the literacy of army personnel, in ref. 2.

Table 1. Comparison between different Arad ostraca

No.	Ostraca content	1	2	3	5	7	8	16	17a	17b	18	21	24	31	38	39a	39b	40	111
1	Order to Eliashib, supply of provisions for the Kittiyim		0.64	0.50	0.91	0.30	0.64	0.51	0.98	0.78	0.53	0.24	0.003	0.10	0.27	0.41	0.06	0.23	0.79
2	Order to Eliashib, supply of provisions for the Kittiyim	0.64		1.00	1.00	0.72	1.00	0.39	0.85	0.78	0.31	0.75	0.79	0.06	0.38	0.98	0.70	0.11	0.96
3	Order to Eliashib mentioning Hananyahu, concerning provisions to Beer Sheba	0.50	1.00		0.23	0.06	0.55	0.36	1.00	0.77	0.27	0.94	0.72	0.16	0.61	0.96	0.84	0.22	0.79
5	Order to Eliashib, supply of provisions, probably for the Kittiyim	0.91	1.00	0.23		0.53	0.60	0.60	0.19	0.40	0.07	0.46	0.12	0.01	0.40	0.24	0.21	0.07	0.98
7	Order to Eliashib, supply of provisions for the Kittiyim	0.30	0.72	0.06	0.53		0.03	0.76	0.17	0.48	0.004	0.43	0.05	0.07	0.27	0.35	1.00	0.15	0.05
8	Order to Eliashib, supply of provisions for the Kittiyim	0.64	1.00	0.55	0.60	0.03		0.68	0.07	1.00	0.17	0.33	0.74	0.42	0.20	0.67	1.00	1.00	0.93
16	Letter to Eliashib from Hananyahu	0.51	0.39	0.36	0.60	0.76	0.68		0.33	1.00	0.03	0.80	0.13	0.38	0.38	0.41	0.40	0.72	0.68
17a	Order to Nahum to proceed to the house of Eliashib to collect provisions	0.98	0.85	1.00	0.19	0.17	0.07	0.33		1.00	0.92	0.36	0.13	0.41	1.00	0.68	1.00	0.17	0.68
17b	Note that Nahum provided provisions to the Kittiyim	0.78	0.78	0.77	0.40	0.48	1.00	1.00	1.00		1.00	0.35	0.40	0.47	1.00	1.00	0.33	0.20	0.40
18	Report to Eliashib from a subordinate fulfilling an order; mention of the Temple	0.53	0.31	0.27	0.07	0.004	0.17	0.03	0.92	1.00		3×10^{-4}	0.02	0.20	0.32	0.94	0.86	0.04	0.73
21	Letter to Gedalyahu from a subordinate, Yehokal	0.24	0.75	0.94	0.46	0.43	0.33	0.80	0.36	0.35	3×10^{-4}		0.35	0.04	0.23	0.71	0.21	0.31	0.90
24	A royal decree ordering the reinforcement of Ramat Negeb against Edom	0.003	0.79	0.72	0.12	0.05	0.74	0.13	0.13	0.40	0.02	0.35		0.01	0.05	0.73	0.38	0.002	0.92
31	List of names	0.10	0.06	0.16	0.01	0.07	0.42	0.38	0.41	0.47	0.20	0.04	0.01		0.33	0.16	0.11	0.35	0.57
38	List of names (including the son of Eliashib)	0.27	0.38	0.61	0.40	0.27	0.20	0.38	1.00	1.00	0.32	0.23	0.05	0.33		0.77	0.33	0.70	0.77
39a	List of names	0.41	0.98	0.96	0.24	0.35	0.67	0.41	0.68	1.00	0.94	0.71	0.73	0.16	0.77		1.00	0.04	0.75
39b	List of names	0.06	0.70	0.84	0.21	1.00	1.00	0.40	1.00	0.33	0.86	0.21	0.38	0.11	0.33	1.00		0.42	0.42
40	Gemaryahu and Nehemyahu report to Malkiyahu mentioning Edom and the king of Judah	0.23	0.11	0.22	0.07	0.15	1.00	0.72	0.17	0.20	0.04	0.31	0.002	0.35	0.70	0.04	0.42		0.67
111	Fragmentary, mentioning guard and horses	0.79	0.96	0.79	0.98	0.05	0.93	0.68	0.68	0.40	0.73	0.90	0.92	0.57	0.77	0.75	0.42	0.67	

A $P \leq 0.2$, highlighted in red, indicates rejection of "single writer" hypothesis, hence accepting a "two different authors" alternative. Note that ostraca 17 and 39 contain writing on both sides of the sherd (marked as "a" and "b").

the capital, Jerusalem, a significant number of literate individuals can be assumed to have lived in Judah *ca*. 600 BCE.

The spread of literacy in late-monarchic Judah provides a possible stage setting for the compilation of literary works. True, biblical texts could have been written by a few and kept in seclusion in the Jerusalem Temple, and the illiterate populace could have been informed about them in public readings and verbal messages by these few (e.g., 2 Kings 23:2, referring to the period discussed here). However, widespread literacy offers a better background for the composition of ambitious works such as the Book of Deuteronomy and the history of Ancient Israel in the Books of Joshua to Kings (known as the Deuteronomistic History), which formed the platform for Judahite ideology and theology (e.g., ref. 25). Ideally, to deduce from literacy on the composition of literary (to differ from mundane) texts, we should have conducted comparative research on the centuries after the destruction of Jerusalem, a period when other biblical texts were written in both Jerusalem and Babylonia according to current textual research (e.g., refs. 1 and 26). However, in the Babylonian, Persian, and early Hellenistic periods, Jerusalem and the southern highlands show almost no evidence in the form of Hebrew inscriptions. In fact, not a single securely dated Hebrew inscription has been found in this territory for the period between 586 and *ca*. 350 BCE[#]—not an ostracon or a seal, a seal impression, or a bulla [the little that we know of this period is in Aramaic, the script of the newly present Persian empire (27)]. This should come as no surprise, because the destruction of Judah brought about the collapse of the kingdom's bureaucracy and deportation of many of the literati. Still, for the centuries between *ca*. 600 and 200 BCE, the tension between current biblical exegesis (arguing for massive composition of texts) and the negative archaeological evidence remains unresolved.

Materials and Methods

This research was conducted on two datasets of written material. The main document assemblage was a corpus of 16 Hebrew ostraca inscriptions found at the Arad fortress (*ca*. 600 BCE). The research was performed on digital

[#]A few coins with Hebrew characters do appear between *ca*. 350 and 200 BCE.

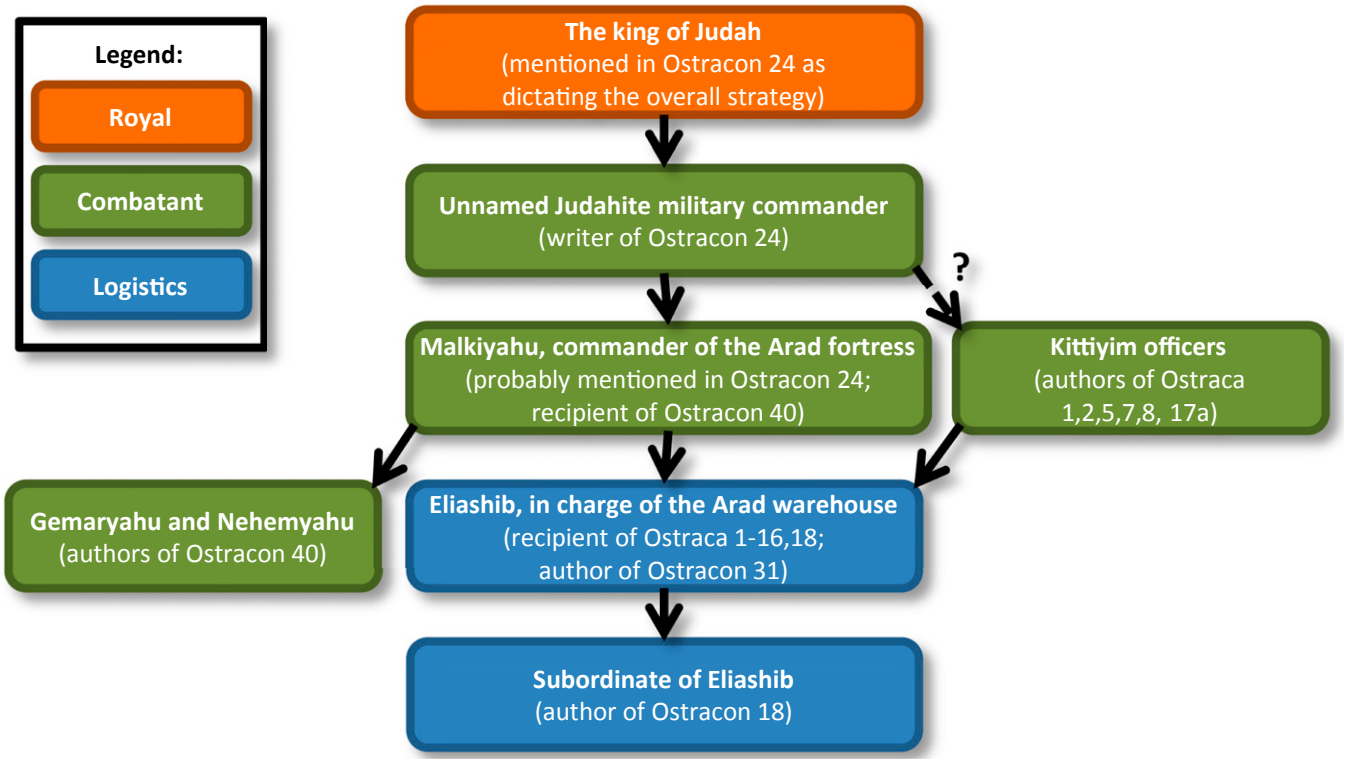

Fig. 4. Reconstruction of the hierarchical relations between authors and recipients in the examined Arad inscriptions; also indicated is the differentiation between combatant and logistics officials.

images of these inscriptions. A second dataset, used to validate the algorithm, contained handwriting samples collected from 18 present-day writers of Modern Hebrew.

The aim of our main algorithm was to differentiate between writers in a given set of texts. This algorithm consisted of several stages. In the first step, character restoration, the image of the inscription was segmented into (often noisy) characters that were restored via a semiautomatic reconstruction procedure. The method was based on the representation of a character as a union of individual strokes that were treated independently and later recombined. The purpose of stroke restoration was to imitate a reed pen's movement using several manually sampled key points. An optimization of the pen's trajectory was performed for all intermediate sampled points. The restoration was conducted via the minimization of image energy functional, which took into account the adherence to the original image, the smoothness of the stroke, as well as certain properties of the reed radius. The minimization problem was solved by performing gradient descent iterations on a cubic-spline representation of the stroke. The end product of the reconstruction was a binary image of the character, incorporating all its strokes (see Figs. S1 and S2).

The second stage of the algorithm, letter comparison, relied on features extracted from the characters' binary images, used to automatically compare characters from different texts. Several features were adapted, referring to 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. The features in use were SIFT (28), Zernike (29), DCT, K_d-tree (30), Image projections (31), L_1, and CMI (32). Additionally, for each feature, a respective distance was defined. Later on, all these distances were combined into a single, generalized feature vector. This vector described each character by the degree of its proximity to all of the characters, using all of the features. Finally, a distance between any two characters was calculated according to the Euclidean distance between their generalized feature vectors (see Table S1 for details concerning various features in use).

The final stage of the algorithm addressed the main question, What is the probability that two given texts were written by the same author? This was achieved by posing an alternative null hypothesis H_0 ("both texts were written by the same author") and attempting to reject it by conducting a relevant experiment. If its outcome was unlikely ($P \leq 0.2$), we rejected the H_0 and concluded that the documents were written by two individuals. Alternatively, if the occurrence of H_0 was probable ($P > 0.2$), we remained agnostic. The experiment testing the H_0 performed a clustering on a set of letters from the two tested inscriptions (of specific type, e.g., *alep*[‖]), disregarding their affiliation to either of the inscriptions. The clustering results should have resembled the original inscriptions if two different writers were present, while being random if this was not the case. Although this kind of test could have been performed on one specific letter, we could gain additional statistical significance if several different letters (e.g., *alep*, *he*, *waw*, etc.) were present in the compared documents. Subsequently, several independent experiments were conducted (one for each letter), and their P values were combined via the well-established Fisher's method (33). The combination represented the probability that H_0 was true based on all of the evidence at our disposal (see Fig. S3 for an illustration of the procedure's flow).

See *Supporting Information* for additional details regarding the methods in use and their results on both Ancient and Modern Hebrew datasets (available at www-nuclear.tau.ac.il/~eip/ostraca/DataSets/Arad_Ancient_Hebrew.zip and www-nuclear.tau.ac.il/~eip/ostraca/DataSets/Modern_Hebrew.zip, respectively). In particular, see Figs. S4 and S5 for samples taken from Modern and Ancient Hebrew datasets, respectively. Additionally, Table S2 summarizes the results of the Modern Hebrew experiment, while Table S3 provides statistics regarding the characters utilized in the Ancient Hebrew experiment.

ACKNOWLEDGMENTS. This research was made possible by the dedicated work of Ms. Ma'ayan Mor. The kind assistance of Dr. Shirly Ben-Dor Evian, Ms. Sivan Einhorn, Ms. Noa Evron, Dr. Anat Mendel, Ms. Myrna Pollak, Mr. Michael Cordonsky, and Mr. Assaf Kleiman is greatly appreciated. We also thank the PNAS editor and the reviewers for their helpful comments and suggestions. A.S. thanks the Azrieli Foundation for the award of an Azrieli Fellowship. Ostracon images are courtesy of the Institute of Archaeology, Tel Aviv University, and of the Israel Antiquities Authority. The research reported here received initial funding from the Israel Science Foundation – F.I.R.S.T. (Bikura) Individual Grant 644/08, as well as Israel Science Foundation Grant

[‖]The Latin transliteration of the letter names differs slightly between Modern and Ancient Hebrew. For Ancient Hebrew, several spellings can be found in the literature: *alep/aleph, bet, gimel, dalet, he, waw, zayin, het/ḥet, tet/ṭet, yod, kap/kaf, lamed, mem, nun, samek/samekh, ayin/ʿayin, pe, sade/ṣade, qop/qof, resh, shin, taw*. For Modern Hebrew, the Unicode standard names are *alef, bet, gimel, dalet, he, vav, zayin, het, tet, yod, kaf, lamed, mem, nun, samekh, ayin, pe, tsadi, qof, resh, shin, tav*. For simplicity's sake, in what follows, we use the first orthography (without the diacritics) for each letter.

1457/13. The research was also funded by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC Grant Agreement 229418, and by an Early Israel grant (New Horizons project), Tel Aviv University. This study was also supported by a generous donation from Mr. Jacques Chahine, made through the French Friends of Tel Aviv University.

1. Schmid K (2012) *The Old Testament: A Literary History* (Fortress, Minneapolis).
2. Rollston CA (2010) *Writing and Literacy in the World of Ancient Israel: Epigraphic Evidence from the Iron Age* (Society of Biblical Literature, Atlanta).
3. Aḥituv S (2008) *Echoes from the Past: Hebrew and Cognate Inscriptions from the Biblical Period* (Carta, Jerusalem).
4. Barkay G (1992) The priestly benediction on silver plaques from Ketef Hinnom in Jerusalem. *Tel Aviv* 19(2):139–192.
5. Barkay G, Vaughn AG, Lundberg MJ, Zuckerman B (2004) The amulets from Ketef Hinnom: A new edition and evaluation. *Bull Am Schools Orient Res* 334: 41–71.
6. Aharoni Y (1981) *Arad Inscriptions* (Israel Exploration Society, Jerusalem).
7. Na'aman N (2011) Textual and historical notes on the Eliashib archive from Arad. *Tel Aviv* 38(1):83–93.
8. Lemaire A (1977) *Inscriptions Hébraïques, Vol. 1: Les Ostraca*. Littératures anciennes du Proche-Orient 9 (Edicions du Cerf, Paris), pp 230–231.
9. Faigenbaum-Golovin S, et al. (2015) Computerized paleographic investigation of Hebrew Iron Age ostraca. *Radiocarbon* 57(2):317–325.
10. Faigenbaum S, et al. (2012) Multispectral images of ostraca: Acquisition and analysis. *J Archaeol Sci* 39(12):3581–3590.
11. Shaus A, Turkel E, Piasetzky E (2012) 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 (IEEE Computer Society, Los Alamitos, CA), pp 641–646.
12. Shaus A, Sober B, Turkel E, Piasetzky E (2013) Improving binarization via sparse methods. Proceedings of the 16th International Graphonomics Society Conference (Tokyo University of Agriculture and Technology Press, Tokyo), pp 163–166.
13. Louloudis G, Gatos B, Stamatopoulos N (2012) ICFHR 2012 competition on writer identification challenge 1: Latin/Greek documents. Proceedings of the 13th International Conference on Frontiers in Handwriting Recognition (IEEE Computer Society, Los Alamitos, CA), pp 829–834.
14. Sober B, Levin D (2016) Computer aided restoration of handwritten character strokes. arXiv:1602.07038.
15. Herzog Z (2002) The fortress mound at Tel Arad: An interim report. *Tel Aviv* 29(1): 3–109.
16. Mazar A, Netzer E (1986) On the Israelite fortress at Arad. *Bull Am Schools Orient Res* 263:87–91.
17. Ussishkin D (1988) The date of the Judean shrine at Arad, Israel. *Explor J* 38:142–157.
18. Na'aman N (2003) Ostracon 40 from Arad reconsidered. *Saxa loquentur. Studien zur Archäologie Palästinas/Israels. Festschrift für Volkmar Fritz zum 65 Geburtstag*, eds den Hertog CG, Hübner U, Münger S (Ugarit, Münster, Germany), pp 199–204.
19. Beit-Arieh I (2007) *Horvat 'Uza and Horvat Radum: Two Fortresses in the Biblical Negev*. Tel Aviv University Monograph Series 25 (Tel Aviv Univ, Tel Aviv).
20. Beit-Arieh I, Freud L (2015) *Tel Malḥata: A central city in the biblical Negev*. Tel Aviv University Monograph Series 32 (Tel Aviv Univ, Tel Aviv).
21. Rollston CA (1999) The script of Hebrew Ostraca of the Iron Age: 8th–6th centuries BCE. PhD thesis (Johns Hopkins Univ, Baltimore).
22. Rollston CA (2006) Scribal education in ancient Israel: The Old Hebrew epigraphic evidence. *Bull Am Schools Orient Res* 344:47–74.
23. Lemaire A (1981) *Les écoles et la formation de la Bible dans l'ancien Israël*. Orbis Biblicus et Orientalis 39 (Editions Universitaires, Fribourg, Switzerland).
24. Naveh J (1960) A Hebrew letter from the seventh century B.C. *Isr Explor J* 10(3): 129–139.
25. Na'aman N (2002) *The Past that Shapes the Present: The Creation of Biblical Historiography in the Late First Temple Period and After the Downfall* (Bialik Institute, Jerusalem). Hebrew.
26. Albertz R (2003) *Israel in Exile: The History and Literature of the Sixth Century B.C.E* (Society of Biblical Literature, Atlanta).
27. Lipschits O, Vanderhooft DS (2011) *The Yehud Stamp Impressions: A Corpus of Inscribed Impressions from the Persian and Hellenistic Periods in Judah* (Eisenbrauns, Winona Lake, IN).
28. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. *Int J Comput Vis* 60(2):91–110.
29. Tahmasbi A, Saki F, Shokouhi SB (2011) Classification of benign and malignant masses based on Zernike moments. *Comput Biol Med* 41(8):726–735.
30. Sexton A, Todman A, Woodward K (2000) Font recognition using shape-based quadtree and kd-tree decomposition, Proceedings of the 3rd International Conference on Computer Vision, Pattern Recognition and Image Processing (IEEE Computer Society, Los Alamitos, CA), pp 212–215.
31. Trier ØD, Jain AK, Taxt T (1996) Feature extraction methods for character recognition—A survey. *Pattern Recognit* 29(4):641–662.
32. Shaus A, Turkel E, Piasetzky E (2012) Quality evaluation of facsimiles of Hebrew First Temple period inscriptions. Proceedings of the 10th IAPR International Workshop on Document Analysis Systems (IEEE Computer Society, Los Alamitos, CA), pp 170–174.
33. Fisher RA (1925) *Statistical Methods for Research Workers* (Oliver and Boyd, Edinburgh).
34. Panagopoulos M, Papaodysseus C, Rousopoulos P, Dafi D, Tracy S (2009) Automatic writer identification of ancient Greek inscriptions. *IEEE Trans Pattern Anal Mach Intell* 31(8):1404–1414.
35. Mumford D, Shah J (1989) Optimal approximations by piecewise smooth functions and associated variational problems. *Commun Pure Appl Math* 42(5):577–685.
36. Kass M, Witkin A, Terzopoulos D (1988) Snakes: Active contour models. *Int J Comput Vis* 1(4):321–331.
37. Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci* 55(1):119–139.
38. Sivic J, Zisserman A (2003) Video Google: A text retrieval approach to object matching in videos. Proceedings of the 9th International Conference on Computer Vision (IEEE Computer Society, Los Alamitos, CA), pp 1470–1477.
39. Tahmasbi A (2012) Zernike moments. Available at www.mathworks.com/matlabcentral/fileexchange/38900-zernike-moments.
40. Armon S (2012) Descriptor for shapes and letters (feature extraction). Available at www.mathworks.com/matlabcentral/fileexchange/35038-descriptor-for-shapes-and-letters-feature-extraction.

Supporting Information

Faigenbaum-Golovin et al. 10.1073/pnas.1522200113

Introduction

The main goal of the current research was to estimate the minimal number of authors involved in the scripting of the Arad corpus. To deal with this issue, we had to differentiate between authors of different inscriptions. Although relevant algorithms have been proposed in the past (e.g., ref. 34 for incised lapidary texts), our experience shows that most of the solutions are tailor-made for specific corpora. The poor state of preservation of the Arad First Temple period ostraca, and the high variance of their cursive texts of mundane nature, presented difficulties that none of the available methods could overcome (see Fig. 2). Therefore, novel image processing and machine learning tools had to be developed.

The input for our system is the digital images of the inscriptions. The algorithm involves two preparatory stages, leading to a third step that estimates the probability that two given inscriptions were written by the same author. All of the stages are fully automatic, with the exception of the first, semiautomatic, preparatory step. The basic steps of the algorithm are as follow:

i) Restoring characters via approximation of their composing strokes, represented as a spline-based structure, and estimated by an optimization procedure (for further details see *Description of the Algorithm*, *Character Restoration*).
ii) Feature extraction and distance calculation: creation of feature vectors describing the characters' various aspects (e.g., angles between strokes and character profiles); calculating the distance (similarity) between characters (see *Description of the Algorithm*, *Feature Extraction and Distance Calc*ulation).
iii) Testing the hypothesis that two given inscriptions were written by the same author. Upon obtaining a suitable P value (the significance level of the test, denoted as P), we reject the hypothesis of a single author and accept the competing proposition of two different authors; otherwise, we remain undecided (see *Description of the Algorithm*, *Hypothesis Testing*).

The next section will present an in-depth description of each of the stages. This will be followed by an experimental section that describes the application of our algorithm to both modern and ancient texts. We verify the validity of our approach by applying the algorithm to modern texts (with a number of contemporary texts written by individuals known to us).

Description of the Algorithm

Character Restoration. The state of preservation of most ostraca is poor at best. After more than two and a half millennia buried in the ground, the inscriptions are often blurry, partially erased, cracked, and stained. However, to analyze the script, clear black and white ("binary") images are required. Theoretically, such depictions of the inscriptions do exist, in the form of manually created facsimiles (drawings of the ostraca), created by epigraphic experts. However, these have been shown to be influenced by the prior knowledge and assumptions of the epigrapher (32). A potential solution for this problem could have been provided by automatic binarization procedures from the domain of image processing. Unfortunately, in our experimentations, various binarization methods produced unsatisfactory results (12).

We finally substituted these initial attempts with a semiautomatic approach of individual character restoration. Restoring a character is equivalent to reconstructing its strokes, which are the character's building blocks, and then combining them. Accordingly, henceforth we will discuss the problem of stroke restoration rather than complete character reconstruction. Stroke restoration aims at imitating the reed pen's movement using several manually sampled key points. An optimization of the pen's trajectory is performed for all intermediate sampled points, taking into account information from the noisy character image. A short mathematical description of the procedure follows; for more details and analysis see ref. 14.

A stroke could be referred to as a 2D piecewise smooth curve $(x(t),y(t))$, depending on the parameter $t\in[a,b]$. However, such a representation ignores the stroke's thickness, which is related to the stance of the writing pen toward the document (in our case, a potshard) and to the characteristics of the pen itself. In the case of Iron Age Hebrew, it is well accepted that the scribes used reed pens, which have a flat, rather than pointed, top. This fact makes the writing thickness even more essential to the process of stroke restoration. Therefore, we denote the stroke as a set-valued function:

$$S(t)=\left\{(p,q)|(p-x(t))^2+(q-y(t))^2\le r(t)^2\right\}\quad t\in[a,b],$$

where $x(t)$ and $y(t)$ represent the coordinates of the center of the pen at t, and $r(t)$ stands for the radius of the pen at t (Fig. S1). The corresponding stroke curve is thus

$$\gamma(t)=(x(t),\,y(t),r(t))\;t\in[a,b],$$

whereas the skeleton of the stroke will accordingly be the curve

$$\beta(t)=(x(t),y(t))\quad t\in[a,b]\,.$$

We note that our model of a written stroke is an approximation, because in reality the top of the reed pen was not necessarily a perfect circle.

Borrowing the idea of minimizing an energy functional (35, 36), we produce an analytic reconstruction of a stroke with respect to a given image $I(p,q)$ $((p,q)\in[1,N]\times[1,M])$. This reconstructed stroke $S^*(t)$ is defined as corresponding to the stroke curve $\gamma^*(t)$, minimizing the following functional:

$$F[\gamma(t)]=c_1\int_a^b\frac{G_I(t)}{r(t)^2}dt+c_2\int_a^b\frac{1}{\sqrt{r(t)}}dt+c_3\sum_{j=0}^{J-1}\int_{t_j+\varepsilon}^{t_{j+1}-\varepsilon}|K(\dot{x},\dot{y},\ddot{x},\ddot{y})|\,dt$$

$$\gamma^*(t)=\underset{\gamma(t)}{\operatorname{argmin}}F[\gamma(t)],$$

where $G_I(t)=\sum_{(p,q)\in S(t)}I(p,q)$ is the sum of the gray level values of the image I inside the disk $S(t)$; $\gamma(t_j)=(x(t_j),y(t_j),r(t_j))\;j=0,\ldots,J$ are manually sampled points on the stroke curve $\gamma(t)$, with respect to the natural parameter t; $\dot{x},\ddot{x}$ and $\dot{y},\ddot{y}$ denote the first and second derivatives of x and y; $K(\dot{x},\dot{y},\ddot{x},\ddot{y})=(\dot{x}\ddot{y}-\dot{y}\ddot{x})/(\dot{x}^2+\dot{y}^2)^{3/2}$ stands for the curvature of the skeleton of the stroke $\beta(t)$; $0<c_1,c_2,c_3,\varepsilon\in\mathbb{R}$ are parameters, set to $c_1=2,c_2=2{,}000,c_3=50,\varepsilon=0.01$ in our experiments.

The reconstruction is subject to initial and boundary conditions at (*a*) the beginning and end of strokes; (*b*) intersections of strokes; (*c*) significant extremal points of the curvature; and (*d*) points with no traces of ink. These conditions are supplied by manual sampling.

The energy minimization problem described above is solved by performing gradient descent iterations on a cubic-spline

representation of the stroke (for more details see ref. 14). The end product of the reconstruction is a binary image of the character, incorporating all its strokes.

Fig. S2 presents a restoration of an entire character, stroke by stroke. It can be seen that although the original character image contains several erosions (Fig. S2*A*), the reconstructed strokes (Fig. S2*C*) look both smooth and complete, and their union results in a clear letter, adhering to the character image (Fig. S2*D*).

Feature Extraction and Distance Calculation. Commonly, automatic comparison of characters relies upon features extracted from the characters' binary images. In this study, we adapted several well-established features from the domains of computer vision and document analysis. These features refer to 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. Some of these features correspond to characteristics commonly used in traditional paleography (21).

The feature extraction process includes a preliminary step of the characters' standardization. The steps involve rotating the characters according to their line inclination, resizing them according to a predefined scale, and fitting the results into a padded (at least 10% on each side) square of size $a_L \times a_L$ (with $L = 1, ..., 22$ the index of the alphabet letter under consideration). On average, the resized characters were 300 × 300 pixels.

Subsequently, the proximity of two characters can be measured using each of the extracted features, representing various aspects of the characters. For each feature, a different distance function is defined (to be combined at a later stage; discussed below).

Table S1 provides a list of the features and distances we use, along with a description of their implementation details. Some of the adjustments (e.g., replacement of the L_2 norm with the L_1 norm) were required due to the large amount of noise present in our medium.

After the features are extracted, and the distances between the features are measured, there arises a challenge of combining the various distances. Several combination techniques [e.g., AdaBoost (37) and Bag of Features (38)] were considered. Unfortunately, boosting-related methods are unsuitable due to the lack of training statistics, and the Bag of Features performed poorly in preliminary experiments using a modern handwritten character dataset (details regarding this dataset are given below). Hence, we developed a different approach for combining the distances.

Our main idea was to consider the distances of a given character from all of the other characters, with respect to all of the features under consideration (i.e., two characters closely resembling each other ought to have similar distances from all other characters). Namely, they will both have small distances from similar characters and large distances from dissimilar characters. This observation leads to a notion of a generalized feature vector (defined here for the first time to our knowledge).

The generalized feature vector is defined by the following procedure (for each letter $L = 1, ..., 22$ in the alphabet). First, we define a distance matrix for each feature. For example, the SIFT distance matrix is

$$U_{SIFT} = \begin{pmatrix} D_{SIFT}(1,1) & \cdots & D_{SIFT}(1,J_L) \\ \vdots & \ddots & \vdots \\ D_{SIFT}(J_L,1) & \cdots & D_{SIFT}(J_L,J_L) \end{pmatrix} = \begin{pmatrix} - & \vec{u}^{1}_{SIFT} & - \\ & \vdots & \\ - & \vec{u}^{J_L}_{SIFT} & - \end{pmatrix},$$

where J_L represents the total number of characters, $D_{SIFT}(i,j)$ is the SIFT distance between characters i and j, and $\vec{u}^{i}_{SIFT} = (D_{SIFT}(i,1) \cdots D_{SIFT}(i,J_L))$ is the vector of SIFT distances between the character i and all of the others.

In addition, we denote the SD of the elements of the matrix U_{SIFT} by $\sigma_{SIFT} = std\{D_{SIFT}(i,j) | (i,j) \in \{1,...,J_L\} \times \{1,...,J_L\}\}$. Matrices of all of the other features ($U_{Zernike}$,U_{DCT}, and so forth) and their respective SDs ($\sigma_{Zernike}$, σ_{DCT}, etc.) are calculated in a similar fashion.

Therefore, each character k is represented by the following vector (of size $7 \cdot J_L$), concatenating the respective normalized row vectors of the distance matrices:

$$\vec{u}_k = \left(\frac{\vec{u}^{k}_{SIFT}}{\sigma_{SIFT}} || \frac{\vec{u}^{k}_{Zernike}}{\sigma_{Zernike}} || \frac{\vec{u}^{k}_{DCT}}{\sigma_{DCT}} || \frac{\vec{u}^{k}_{Kd-tree}}{\sigma_{Kd-tree}} || \frac{\vec{u}^{k}_{Proj}}{\sigma_{Proj}} || \frac{\vec{u}^{k}_{L1}}{\sigma_{L1}} || \frac{\vec{u}^{k}_{CMI}}{\sigma_{CMI}} \right) \in \mathbb{R}^{7 \cdot J_L}.$$

In this fashion, each character is described by the degree of its kinship to all of the characters, using all of the various features.

Finally, the distance between characters i and j is calculated according to the Euclidean distance between their generalized feature vectors:

$$chardist(i,j) = \left\| \vec{u}_i - \vec{u}_j \right\|_2.$$

The main purpose of this distance is to serve as a basis for clustering at the next stage of the analysis.

Hypothesis Testing. At this stage we address the main question raised above: What is the probability that two given texts were written by the same author? Commonly, similar questions are addressed by posing an alternative null hypothesis H_0 and attempting to reject it. In our case, for each pair of ostraca, the H_0 is both texts were written by the same author. This is performed by conducting an experiment (detailed below) and calculating the probability ($P \in [0,1]$) of an affirmative answer to H_0. If this event is unlikely ($P \leq 0.2$), we conclude that the documents were written by two different individuals (i.e., reject H_0). However, if the occurrence of H_0 is probable ($P > 0.2$), we remain agnostic. We reiterate that in the latter case we cannot conclude that the two texts were in fact written by a single author.

The experiment, which is designed to test H_0, is composed of several substeps (illustrated in Fig. S3):

i) Initialization: We begin with two sets of characters of the same letter type (e.g., *alep*), denoted A and B, originating from two different texts (Fig. S3*A*).

ii) Character clustering: The union $A \cup B$ is a new, unlabeled set (Fig. S3*B*). This set is clustered into two classes, labeled I and II, using a brute-force (and not heuristic) implementation of k-means ($k = 2$). The clustering uses the generalized feature vectors of the characters, and the distance *chardist*, defined above (Fig. S3*C*).

iii) Cluster labels consistency: If $|I| > |II|$, their labels are swapped.

iv) Similarity to cluster I: For each of the two original sets, A and B, the maximal proportion of their elements in class I (their "similarity" to class I) is defined as

$$MP_I = \max\left\{ \frac{|A \cap I|}{|A|}, \frac{|B \cap I|}{|B|} \right\}.$$

v) Counting valid combinations: We consider all of the possible divisions of $A \cup B$ into two classes i and ii, s.t. $|i| = |I|$. The number of such valid combinations is denoted by NC.

vi) Significance level calculation: The P value is calculated as

$$P = \frac{|\{i \,|\, MP_i \geq MP_I\}|}{NC}.$$

That is, P is the proportion of valid combinations with at least the same observational MP. This is analogous to integrating over a tail of a probability density function.

The rationale behind this calculation is based on the scenario of two authors (negation of H_0). In such a case, we expect the k-means clustering to provide a sound separation of their characters (Fig. S3*D*), that is, *I* and *II* would closely resemble *A* and *B* (or *B* and *A*). This would result in MP_I being close to 1. Furthermore, the proportion of valid combinations with $MP_i \geq MP_I$ will be meager, resulting in a low *P*. In such a case, the H_0 hypothesis would be justifiably rejected.

In the opposite scenario of a single author:

- If a sufficient number of characters is present, there is an arbitrary low probability of receiving clustering results resembling *A* and *B*. In a common case, the MP_I will be low, which will result in high *P*.
- Alternatively, if the number of characters is low, the clustering may result in a high MP_I by chance. However, in this case *NC* would be low, and the *P* will remain high.

Either way, in this scenario, we will not be able to reject the H_0 hypothesis.

Notes:

- We assume that each given text was written by a single author. If multiple authors wrote the text, both H_0 and its negation should be altered. We do not cover such a case.
- In substep *iii*, the swapping is performed for regularization purposes, because the measurement on substep *iv* is not symmetric. Substep *iii* verifies that *I* is a minority class, and thus the value of $MP_I = 1$ is achieved only if the clustering resembles the original sets *A* and *B*.
- In cases where $|I| = |II|$ (substep *iii*), the results of substeps *iv*–*vi* can be affected by swapping the classes. To avoid such infrequent inconsistencies, we perform the calculations for both alternatives, and choose the lower *P*.
- Note that in any case, the definition of *P* in substep *vi* results in $P > 0$.
- Not every text provides a sufficient amount of characters for every type of letter in the alphabet. In our case, we do not perform comparisons for sets *A* and *B* such that: $|A| = 1$ & $|B| \leq 6$ or $|B| = 1$ & $|A| \leq 6$ or $|A| = 2$ & $|B| = 2$.

As specified, substeps *i*–*vi* are applied to one specific letter of the alphabet (e.g., *alep*) present (in sufficient quantities) in the pair of texts under comparison. However, we can often gain additional statistical significance if several different letters (e.g., *alep*, *he*, *waw*, etc.) are present in the compared documents. In such circumstances, several independent experiments are conducted (one for each letter), resulting in corresponding *P*s. We combine the different values into a single *P* via the well-established Fisher method (ref. 33; in case no comparison can be conducted for any letter in the alphabet, we assign $P = 1$). This end product represents the probability that H_0 is true based on all of the evidence at our disposal.

Experiment Details and Results

Our experiments were conducted on two large datasets. The first is a set of samples collected from contemporary writers of Modern Hebrew (www-nuclear.tau.ac.il/~eip/ostraca/DataSets/Modern_Hebrew.zip). This dataset allowed us to test the soundness of our algorithm. It was not used for parameter-tuning purposes, however, because the algorithm was kept as parameter-free as possible. The second dataset contained information from various Arad Ancient Hebrew ostraca, dated to *ca*. 600 BCE, described in detail in the main text (www-nuclear.tau.ac.il/~eip/ostraca/DataSets/Arad_Ancient_Hebrew.zip). Following are the specifications and the results of our experiments for both datasets.

Modern Hebrew Experiment. The handwritings of 18 individuals $i = 1, ..., 18$ were sampled. Each individual filled in a Modern Hebrew alphabet table consisting of 10 occurrences of each letter, out of the 22 letters in the alphabet (the number of letters and their names are the same as in Ancient Hebrew; see Fig. S4 for a table example). These tables were scanned and their characters were segmented. For a complete dataset of the characters, see www-nuclear.tau.ac.il/~eip/ostraca/DataSets/Modern_Hebrew.zip.

From this raw data, a series of "simulated" inscriptions were created. Owing to the need to test both same-writer and different-writer scenarios, the data for each writer were split. Furthermore, to imitate a common situation in the Arad corpus, where the scarcity of data is prevalent (Table S3), each simulated inscription used only three letters (i.e., 15 characters, 5 characters for each letter). In total, 252 inscriptions were "simulated" in the following manner:

- All of the letters of the alphabet except for *yod* (because it is too small to be considered by some of the features) were split randomly into seven groups (three letters in each group) $g = 1, ..., 7$: *gimel*, *het*, *resh*; *bet*, *samek*, *shin*; *dalet*, *zayin*, *ayin*; *tet*, *lamed*, *mem*; *nun*, *sade*, *taw*; *he*, *pe*, *qop*; *alep*, *waw*, *kap*.
- For each writer *i*, and each letter belonging to group *g*, five characters were assigned into simulated inscription $S_{i,g,1}$, with the rest assigned to $S_{i,g,2}$.

In this fashion, for constant *i* and *g*, we can test whether our algorithm arrives at wrong rejection of H_0 for $S_{i,g,1}$ and $S_{i,g,2}$ (FP indicates "false-positive" error; 18 writers and 7 groups producing 126 tests in total). Additionally, for constant *g*, $1 \leq i \neq j \leq 18$, and $b, c \in \{1,2\}$, we can test whether our algorithm fails to correctly reject H_0 for $S_{i,g,b}$ and $S_{j,g,c}$ (FN indicates "false-negative" error $[(18 \times 17)/2] \times 7 \times 2 \times 2 = 4{,}284$ tests in total).

The results of the Modern Hebrew experiment are summarized in Table S2. It can be seen that in modern context the algorithm yields reliable results in ~98% of the cases (about 2% of both FP and FN errors). These results signify the soundness of our algorithmic sequence. The successful and significant results on the Modern Hebrew dataset paved the way for the algorithm's application on the Arad Ancient Hebrew corpus.

Arad Ancient Hebrew Experiment. As specified in the main text, the core experiment addresses ostraca from the Arad fortress, located on the southern frontier of the kingdom of Judah. These inscriptions belong to a short time span of a few years, *ca*. 600 BCE, and are composed of army correspondence and documentation.

The texts under examination are 16 ostraca: 1, 2, 3, 5, 7, 8, 16, 17, 18, 21, 24, 31, 38, 39, 40, and 111. Ostraca 17 and 39 contain writing on both sides of the potshard and were treated as separate texts (17a and 17b and 39a and 39b), resulting in 18 texts under examination. As stated in the algorithm description, we assume that each text was written by a single author. A short summary of the content of the texts can be seen in Table 1.

The seven letters we used were *alep, he, waw, yod, lamed, shin*, and *taw*, because they were the most prominent and simple to restore. In the abovementioned ostraca, out of the 670 deciphered characters of these types in the original publication (6), 501 legible characters were restored, based upon computerized images of the inscriptions. These images were obtained by scanning the negatives taken by the Arad expedition (courtesy of the Israel Antiquities Authority and the Institute of Archaeology of Tel Aviv University). After performing a manual quality assurance procedure (verifying the adherence of the restored characters to the original image; Fig. S2*D*), 427 restored characters remained. The resulting letters' statistics for each text are summarized in Table S3. For a complete dataset of the characters, see www-nuclear.tau.ac.il/~eip/ostraca/DataSets/Arad_Ancient_Hebrew.zip. In addition, a comparison between several specimens of the letter *lamed* is provided in Fig. S5.

We reiterate that our algorithm requires a minimal number of characters to compare a pair of texts. For example, when we compared ostraca 31 and 38, the letters in use were *he* (7:1 characters), *waw* (6:2 characters), and *yod* (4:2 characters). The three independent tests respectively yielded $P = 0.125$, $P = 0.25$, and $P = 1$. Their combination through Fisher's method resulted in the final value of $P = 0.327$, not passing the preestablished threshold. Therefore, in this case, we remain agnostic with respect to the question of common authorship. However, the comparison of texts 1 and 24 used all possible letters, *alep*, *he*, *waw*, *yod*, *lamed*, *shin*, and *taw*, resulting in *P*s of 0.559, 0.00366, 0.375, 0.119, 0.0286, 0.429, and 0.0769, respectively. The combined result was $P = 0.003$, passing the threshold of 0.2. Therefore, in the latter case, we reject the H_0 hypothesis and conclude that these texts were written by two different individuals.

The complete comparison results are summarized in Table 1. We can observe six pairwise distinct "quadruplets" of texts: (*i*) 7, 17a, 24, and 40; (*ii*) 5, 17a, 24, and 40; (*iii*) 7, 18, 24, and 40; (*iv*) 5, 18, 24, and 40; (*v*) 7, 18, 24, and 31; and (*vi*) 5, 18, 24, and 31. The existence of no less than six such combinations indicates the high probability that the corpus indeed contains at least four different authors. As specified in the main text, additional (contextual) considerations can raise this number up to at least six distinct writers. Among these, the different authors of the prosaic lists of names in ostraca 31 and 39 were most likely located at the tiny fort of Arad, implying the composition by authors who were not professional scribes. For the full implications of our results, see the main text.

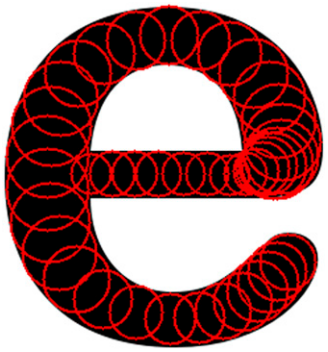

Fig. S1. The Latin character "e" as unification of discs. The discs painted in red over the character were created using the stroke restoration algorithm.

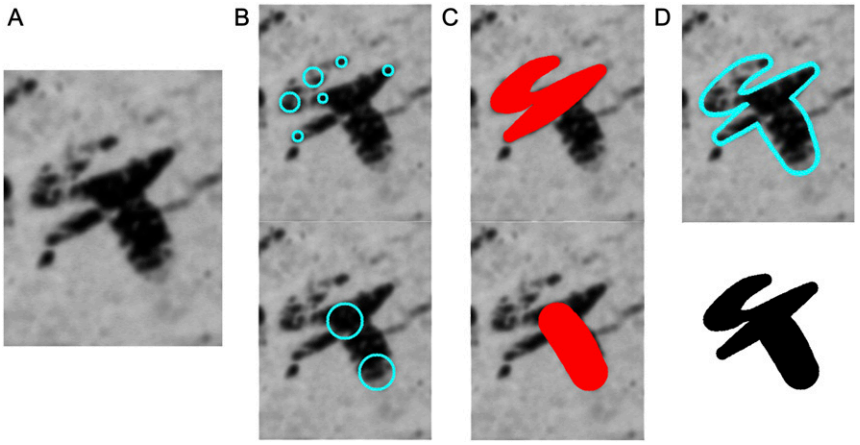

Fig. S2. Example of a semiautomatic stroke restoration of the character *waw* from Arad ostracon 24. (*A*) Image of the character to be reconstructed. (*B*) Manually sampled key points (of top and bottom strokes, respectively). (*C*) The semiautomatic stroke restorations (of top and bottom strokes, respectively). (*D*) The reconstructed character (*Top*: the contour of the reconstructed character overlaid on top of the original image; *Bottom*: the binary image of the restored character). Images are courtesy of the Institute of Archaeology, Tel Aviv University, and of the Israel Antiquities Authority.

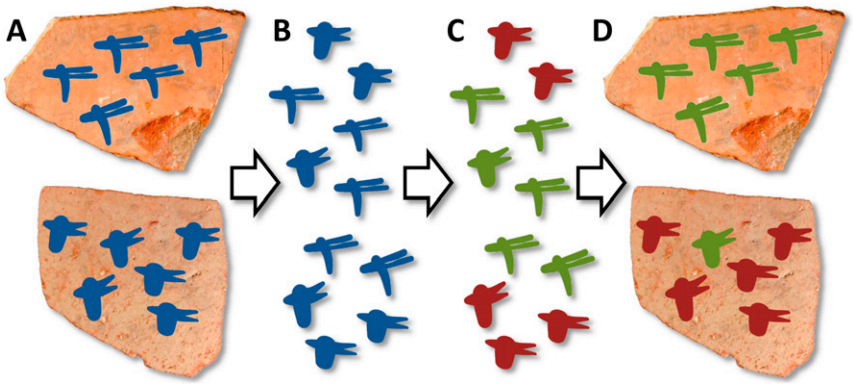

Fig. S3. Artificial illustration of H_0 rejection experiment (containing only *alep* letters). (*A*) Two compared documents. (*B*) Unifying their sets of characters. (*C*) Automatic clustering. (*D*) The clustering results vs. the original documents. Images are courtesy of the Institute of Archaeology, Tel Aviv University, and of the Israel Antiquities Authority.

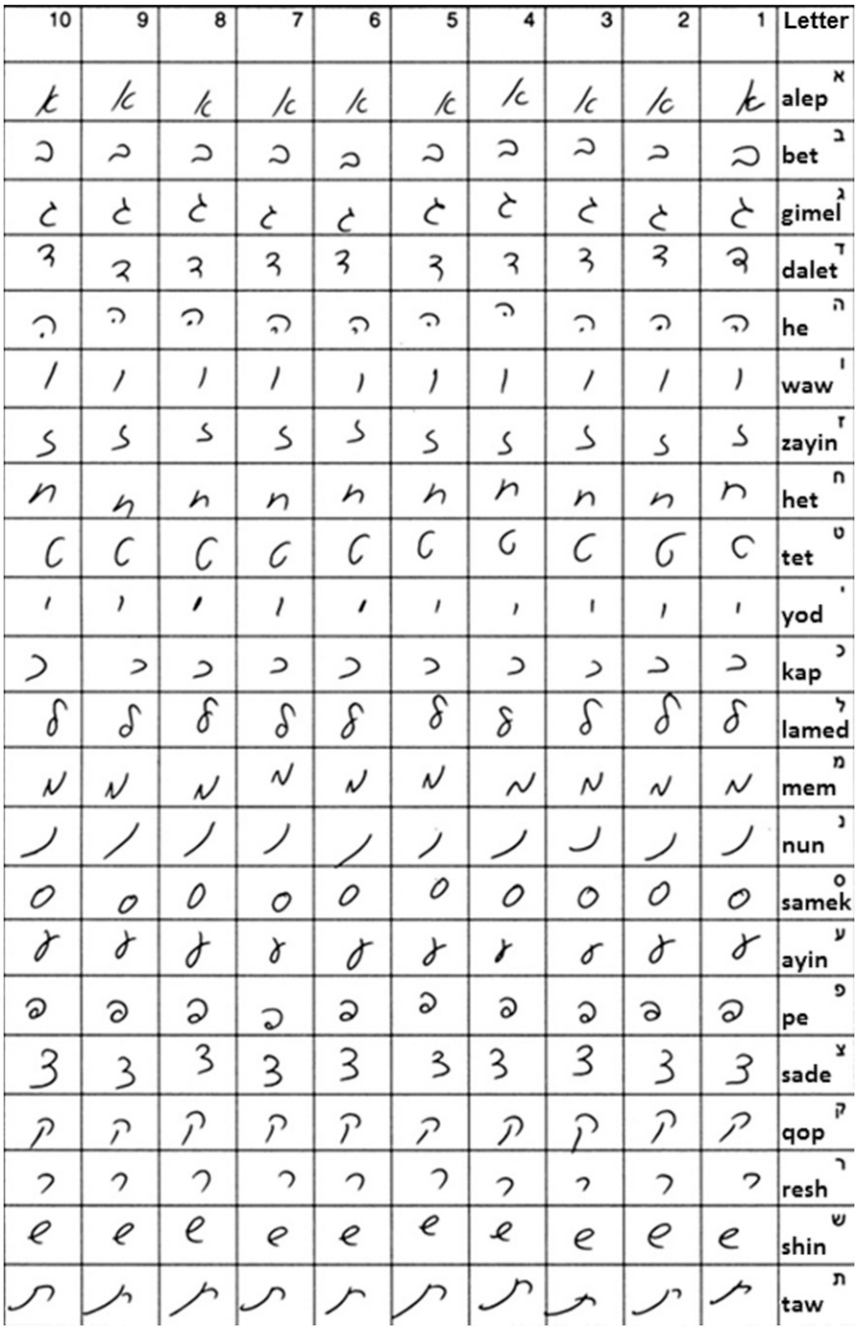

Fig. S4. An example of a Modern Hebrew alphabet table, produced by a single writer (with 10 samples of each letter).

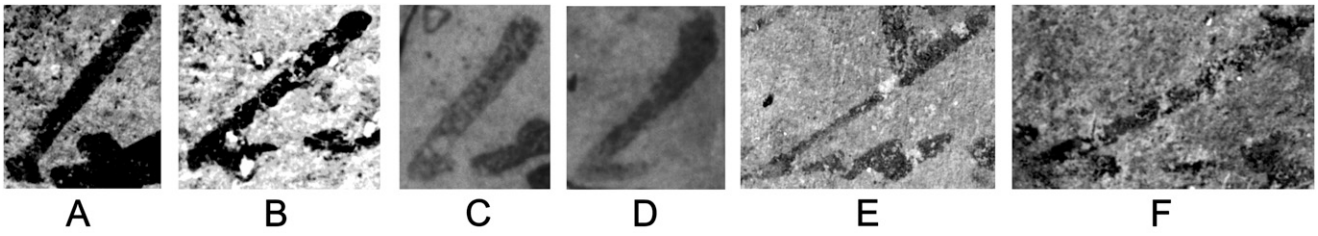

Fig. S5. Comparison between several specimens of the letter *lamed*, stemming from Arad 1 (*A* and *B*), Arad 7 (*C* and *D*), and Arad 18 (*E* and *F*). Note that our algorithm cannot distinguish between the author of Arad 1 and the author of Arad 7, or the authors of Arad 1 and Arad 18. However, Arad 7 and Arad 18 were probably written by different authors ($P = 0.015$ for the letter *lamed* and $P = 0.004$ for the whole inscription, combining information from different letters). Images are courtesy of the Institute of Archaeology, Tel Aviv University, and of the Israel Antiquities Authority.

Table S1. Features and distances used in our algorithm

Feature (ref.)	Feature implementation details	Distance implementation details
SIFT (28)	For each character j, we use the normalized SIFT descriptors $\vec{d}_i \in \mathbb{R}^{128}$ (with $\|\vec{d}_i\|_2 = 1$) and the spatial locators $\vec{l}_i \in [1, a_L]^2$ for at most 40 significant key points $k_i = (\vec{d}_i, \vec{l}_i)$, according to the original SIFT implementation. The resulting feature is a set $f_j^{SIFT} = \{k_i\}_{i=1}^{40}$.	The distance between f_1^{SIFT} and f_2^{SIFT} is determined as follows: *i*) For each key point $k_i^1 \in f_1^{SIFT}$, find a matching key point $m_i^2 \in f_2^{SIFT}$ s. t. $m_i^2 = \underset{(d_j^2, l_j^2) \in f_2^{SIFT}}{\arg\min}\, dist(k_i^1, k_j^2)$; where $dist(k_i^1, k_j^2) = \arccos(\langle d_i^1, d_j^2 \rangle) \cdot \|l_i^1 - l_j^2\|_2^2$. Thus, our definition augments the original SIFT distance by adding spatial information. *ii*) The one-sided distance is $D_{SIFT}^{1,2} = \underset{i}{median}\{dist(k_i^1, m_i^2)\}$. *iii*) The final distance is $D_{SIFT}(1,2) = \frac{D_{SIFT}^{1,2} + D_{SIFT}^{2,1}}{2}$.
Zernike (29)	An off-the-shelf (39) implementation was used. Zernike moments up to the fifth order were calculated.	$D_{Zernike}$ is the L_1 distance between the Zernike feature vectors.
DCT	MATLAB (R2009a) default implementation was used.	D_{DCT} is the L_1 distance between the DCT feature vectors.
K_d-tree (30)	An off-the-shelf (40) implementation was used. Both orders of partitioning are used (first height, then width, and vice versa)	$D_{Kd-tree}$ is the L_1 distance between the K_d-tree feature vectors.
Image projections (31)	The implementation results in cumulative distribution functions of the histogram on both axes.	D_{Proj} is the L_1 distance between the projections' feature vectors; this is similar to the Cramér–von Mises criterion (which uses L_2 distance).
L1	Existing character binarizations.	D_{L1} is the L_1 distance between the character images.
CMI (32)	Existing character binarizations, with values in $\{0,1\}$.	The CMI computes a difference between the averages of the foreground and the background pixels of $\Im$, marked by a binary mask M, $CMI(M, \Im) = \mu_1 - \mu_0$, where $\mu_k = mean\{\Im(p,q) | M(p,q) = k\}$ $k = 0,1$ In our case, given character binarizations B_1, B_2, the one-sided distance is $D_{CMI}^{1,2} = 1 - CMI(B_1, B_2)$. The final distance is $D_{CMI}(1,2) = \frac{D_{CMI}^{1,2} + D_{CMI}^{2,1}}{2}$.

Table S2. Results of the Modern Hebrew experiment

Group of letters (corresponding to g-index of simulated inscriptions)	False positive (FP out of all same-writer comparisons)	False negative (FN out of all different-writer comparisons)	False positive, % (FP out of all same-writer comparisons)	False negative, % (FN out of all different-writer comparisons)
Gimel, het, resh	0/18	8/612	0	1.31
Bet, samek, shin	1/18	5/612	5.56	0.82
Dalet, zayin, ayin	1/18	18/612	5.56	2.94
Tet, lamed, mem	0/18	22/612	0	3.59
Nun, sade, taw	0/18	3/612	0	0.49
He, pe, qop	0/18	16/612	0	2.61
Alep, waw, kap	1/18	11/612	5.56	1.80
Total	3/126	83/4,284	2.38	1.94

The percentages of false-positive and false-negative errors are about 2% each.

Table S3. Letter statistics for each text under comparison

	Alphabet letters						
Text	Alep	He	Waw	Yod	Lamed	Shin	Taw
1	4	5	3	7	3	3	8
2	6	3	3	5	3	1	7
3	2	4	5	4	4	3	3
5	5	3	1	3	4	2	4
7	1	2	1	4	6	8	5
8	2	1	2	1	4	4	2
16	6	3	9	5	10	3	2
17a	2	4	2	2	2	1	2
17b		1		2	1	1	2
18	2	4	4	5	6	6	3
21	5	4	6	6	12	5	2
24	9	10	5	8	4	4	7
31	3	7	6	4	1	1	
38	1	1	2	2	2	1	
39a	3	3	3	5	2	1	1
39b	3	1	1	4	1		
40	4	5	3	4		3	2
111	4	3	3	3	1	3	2