

Pitfalls of the Geographic Population Structure (GPS) Approach Applied to Human Genetic History: A Case Study of Ashkenazi Jews

Pavel Flegontov^{1,2,3,*†}, Alexei Kassian^{4,5,*†}, Mark G. Thomas⁶, Valentina Fedchenko⁷, Piya Changmai¹, and George Starostin^{5,8}

¹Department of Biology and Ecology, Faculty of Science, University of Ostrava, Czech Republic

²Institute of Parasitology, Biology Centre, Czech Academy of Sciences, Ceske Budejovice, Czech Republic

³A.A. Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia

⁴Institute of Linguistics, Russian Academy of Sciences, Moscow, Russia

⁵Russian Presidential Academy of National Economy and Public Administration (RANEPA), Moscow, Russia

⁶Research Department of Genetics, Evolution and Environment, University College London, United Kingdom

⁷Saint Petersburg State University, Russia

⁸Russian State University for the Humanities, Moscow, Russia

*Corresponding author: E-mail: pavel.flegontov@osu.cz, a.kassian@gmail.com.

†These authors contributed equally to this work.

Accepted: June 27, 2016

Abstract

In a recent interdisciplinary study, Das et al. have attempted to trace the homeland of Ashkenazi Jews and of their historical language, Yiddish (Das et al. 2016. Localizing Ashkenazic Jews to Primeval Villages in the Ancient Iranian Lands of Ashkenaz. *Genome Biol Evol.* 8:1132–1149). Das et al. applied the geographic population structure (GPS) method to autosomal genotyping data and inferred geographic coordinates of populations supposedly ancestral to Ashkenazi Jews, placing them in Eastern Turkey. They argued that this unexpected genetic result goes against the widely accepted notion of Ashkenazi origin in the Levant, and speculated that Yiddish was originally a Slavic language strongly influenced by Iranian and Turkic languages, and later remodeled completely under Germanic influence. In our view, there are major conceptual problems with both the genetic and linguistic parts of the work. We argue that GPS is a provenancing tool suited to inferring the geographic region where a modern and recently unadmixed genome is most likely to arise, but is hardly suitable for admixed populations and for tracing ancestry up to 1,000 years before present, as its authors have previously claimed. Moreover, all methods of historical linguistics concur that Yiddish is a Germanic language, with no reliable evidence for Slavic, Iranian, or Turkic substrata.

Key words: population genetics, Ashkenazi Jews, Yiddish, geographic population structure, relexification.

Das et al. (2016) have recently presented an unorthodox interpretation of the history of Ashkenazi Jews, suggesting that they originated from a “Slavo–Iranian confederation”. This claim is essentially based on application of the geographic population structure (GPS) approach (Elhaik et al. 2014) to high-density autosomal single nucleotide polymorphism (SNP) data. GPS was used to infer geographic coordinates of populations supposedly ancestral to Ashkenazi Jews—

“primeval villages” (Das et al. 2016)—which were then interpreted at great length to support the Yiddish relexification hypothesis advanced earlier by one of the coauthors (Wexler 1991, 2002, 2010); an hypothesis that Yiddish was originally a Slavic language later remodeled completely under Germanic influence. In our view, there are major conceptual problems with both the genetic and linguistic components of the work. Below we argue that GPS is a provenancing tool, at best suited

© The Author 2016. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

to inferring the geographic region where a modern and recently unadmixed genome is most likely to arise, but is hardly suitable for admixed populations and for tracing ancestry up to 1,000 years before present, as its authors have previously claimed. Moreover, all methods of historical linguistics concur that Yiddish is a Germanic language, leaving no room for the Slavic relexification hypothesis and for the idea of significant early Yiddish–Persian contacts in Asia Minor. Thus, the authors' statement "Yiddish is a Slavic language created by Irano-Turko-Slavic Jewish merchants along the Silk Roads as a cryptic trade language, spoken only by its originators to gain an advantage in trade" (Das et al. 2016) remains an assertion in the realm of unsupported speculation.

GPS is not Suitable for Inferring Ancestry

Briefly, GPS (Elhaik et al. 2014) works in the following way (fig. 1): 1) unsupervised ADMIXTURE analysis represents a worldwide panel, composed of modern individuals only, as a mixture of an optimal number (K) of hypothetical ancestral populations; 2) based on allele frequencies inferred by ADMIXTURE, K ancestral populations of size n are simulated and used in a supervised ADMIXTURE run (Alexander and Lange 2011); 3) populations from the original panel that supposedly have not migrated in the last few centuries form the reference panel, and test individuals are added; 4) genetic distances among the test individuals and reference populations are calculated based on the comparison of their admixture profiles obtained in the supervised ADMIXTURE run (genetic admixture distance is defined as "the minimal Euclidean distance between the admixture proportions of an individual to those of all individuals of a certain population," Das et al. 2016); 5) the test individuals are assigned to best-matching reference populations, based on genetic distances; and/or 6) geographic coordinates of the test individuals are predicted by averaging across reference populations' coordinates weighted according to their genetic distance from the test individuals (only ten best-matching references are used at this step). Unsupervised ADMIXTURE mode often overestimates small admixture coefficients, whereas the supervised mode mitigates this problem, provided that reference populations K are genetically homogeneous and truly ancestral for the other individuals in the dataset (Alexander and Lange 2011). This mode also allows easy analysis of new individuals without recomputing admixture coefficients for the whole dataset, and therefore is extensively used in commercial-oriented applications.

In practice, an unadmixed individual would be positioned by the GPS software on the map near its best-matching reference population, whereas an individual representing a two-way mixture of reference populations would be positioned on a line connecting coordinates of the mixture partners: "for an individual of mixed origins, the inferred coordinates represent

the mean geographical locations of their immediate ancestors" (Das et al. 2016). Thus, GPS essentially represents a clustering tool useful for inferring provenance of modern unadmixed genomes, provided a large reference panel is available. However, interpretation of GPS results offered by Das et al. (2016) goes much further: "GPS predictions should therefore be interpreted as the last place that admixture has occurred, termed here *geographical origin*." At another point Das et al. have provided a somewhat different interpretation: "GPS infers the geographical origins of an individual by averaging over the origins of all its ancestors." We would argue that these interpretations, which lie at the core of the paper discussed here, are simply wrong.

The first conceptual problem is that each individual has thousands of ancestors at the time depth of $\sim 1,000$ years (although the theoretical upper limit of distinct ancestors is much higher and equals 2^N , where N is the number of generations). Localizing all the ancestors to a single primeval village or to "the last place where admixture has occurred" is extremely difficult to imagine: even if we assume that ADMIXTURE profiles of most individual's ancestors 1,000 years ago are indistinguishable given the resolution of the GPS method, and that those ancestors were located in a limited geographical region, we cannot avoid the fact that mixture of genetically distinct populations was a widespread phenomenon in human history, both ancient and recent (Hellenthal et al. 2014). Obviously, averaged (modern!) coordinates of mixture partners cannot be interpreted as the last place where admixture has occurred or as averaged coordinates of all individual's ancestors, which is illustrated by simple diagrams showing movement of test and reference populations in time and space (fig. 2D). Second, as a clustering method operating on modern references, GPS has no way to trace population movements back in time (fig. 2B, C). Only studies of ancient genomes and their coordinates in space and time can approach locating ancestral homelands with enough precision (Allentoft et al. 2015; Haak et al. 2015). In summary, even if a dense sampling of modern reference populations is available, and if they have not moved for considerable time (an assumption made by Das et al. 2016), inferring correct ancestral locations of test populations that have admixed and/or moved over time is hardly possible with the GPS method.

Another fundamental problem lies in data reduction inherent in the GPS approach: genotypes at about 100,000 sites (Das et al. 2016) are not analyzed directly, but collapsed to just few variables, that is, admixture coefficients for nine hypothetical ancestral populations (Elhaik et al. 2014; Das et al. 2016). Genetic distances among individuals and their downstream analyses in Das et al. (2016) were essentially based on this extremely reduced set of variables; variables that are themselves biased by the particular sample coverage used to infer them (see fig. 1). On the contrary, disentangling recent migration and admixture events, for example, those within the

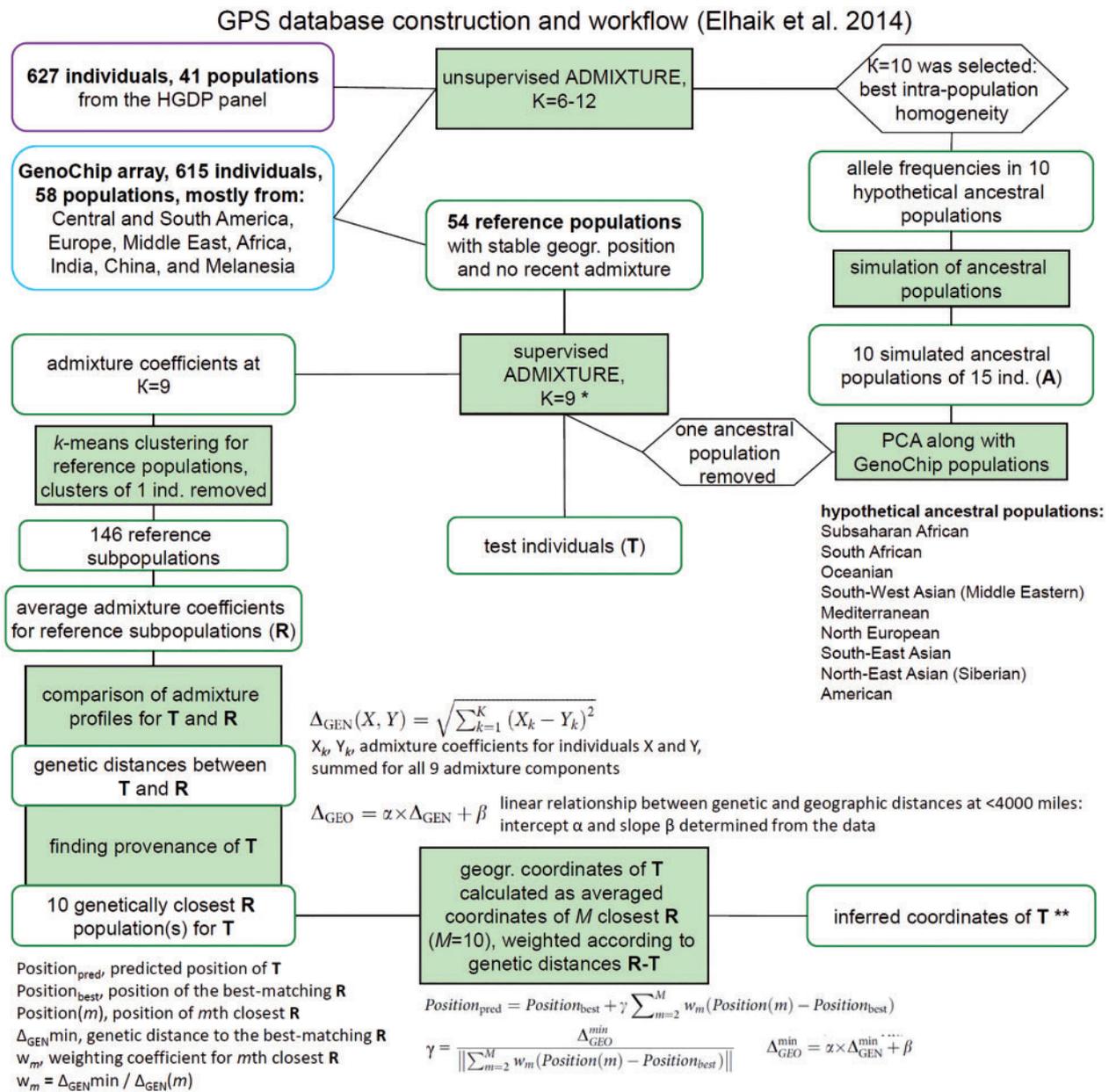


Fig. 1.—A diagram illustrating the GPS workflow. * Steps prior to the supervised ADMIXTURE run are used for the method optimization only, and a typical workflow starts with a supervised ADMIXTURE run. ** Coordinates of test individuals inferred by the GPS algorithm were interpreted by Das et al. (2016) as coordinates of ancestral populations (“primeval villages”) or as the last place where admixture has occurred.

1000-year window, often requires the most sophisticated methods available and the largest amounts of genotype data. Variations in frequencies of common SNP alleles, used in high-throughput genotyping arrays, tend to be small in recently diverged populations (Schiffels et al. 2016), and therefore all methods of clustering and admixture inference based on common genetic variants lack resolution in this time window. Extremely rare SNP alleles (with global frequency <1%) or autosomal haplotypes provide a much

better resolution (Leslie et al. 2015; Schiffels et al. 2016). The approach based on rare SNP alleles (rarecoal: Schiffels et al. 2016) requires whole-genome data, and autosomal haplotypes can be inferred from dense SNP array or whole-genome data, and analyzed with the ChromoPainter, fineSTRUCTURE (Lawson et al. 2012), and GLOBETROTTER tools (Hellenthal et al. 2014).

Ten Sardinian villages analyzed in the original GPS publication (Elhaik et al. 2014) make a good example of this data

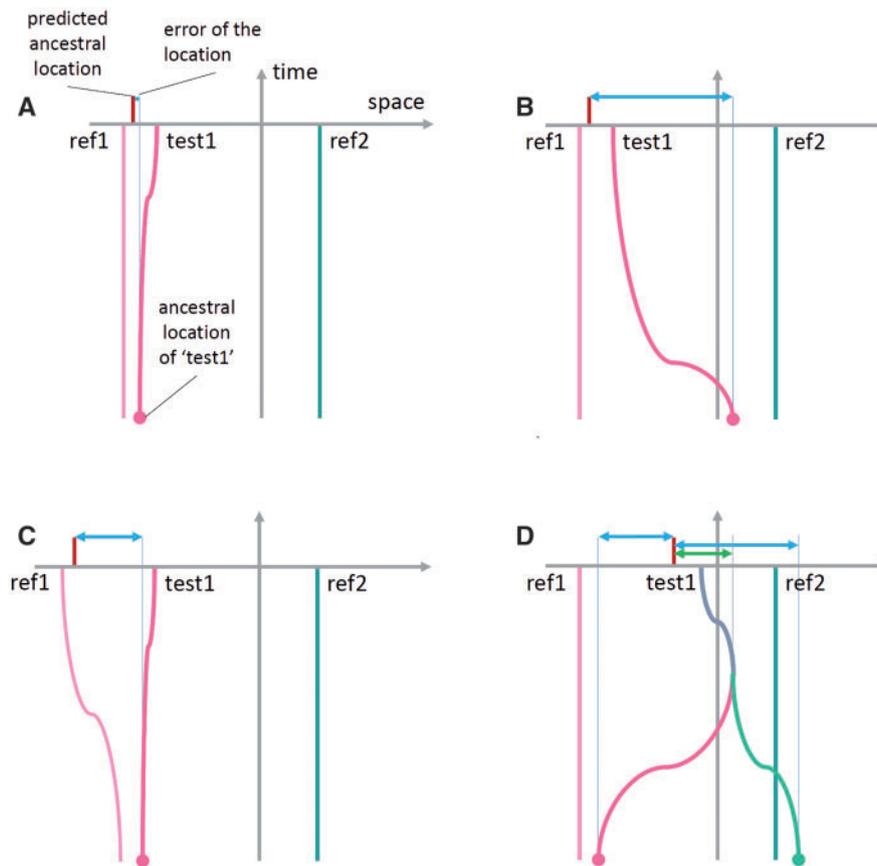


FIG. 2.—Space-time diagrams illustrating behavior of a very simple system of two reference populations (ref1 and ref2) and one test population (test1). Genetically similar populations are depicted with similar colors; ancestral locations that would be predicted by the GPS algorithm are shown with dark-red vertical lines, and their distances from the true ancestral locations are shown with blue arrows. Spatial coordinates for the ancestral population of test1 would be accurately predicted by GPS only in the simplest scenario, when neither test nor references move much in space during a time window appropriate for the method (A). If test1 has moved considerably during its history (B), its ancestral location cannot be predicted without knowing coordinates of any ancient genomes along its “worldline”. In the scenario B, closely related populations ref1 and test1 occupied a large territory in the past, and then their range underwent considerable contraction. The Kets, a Yeniseian-speaking Siberian ethnic group, represent a good example of this scenario: Yeniseian tribes over the last 1000 years have moved a long distance from the south to the north, and their range underwent an extreme contraction under various pressures (Flegontov et al. 2016). Thus, location of the Kets today is very far from their original homeland ~1000 years ago. Similarly, if a reference moves over time, GPS would also make an error in estimating the ancestral location of test1 (C). If test1 is a one-to-one mixture of ref1 and ref2, single ancestral location is meaningless for this population (D), and GPS would place the ancestral location of test1 on the midpoint between ref1 and ref2. The statement “GPS predictions should therefore be interpreted as the last place that admixture has occurred, termed here *geographical origin*” (Das et al. 2016) is invalid. Obviously, location of the mixture partners and of the mixed population can change over time, and GPS has no information to trace these past movements, producing an erroneous location of the last admixture event (the green arrow shows that the inferred last place of admixture is misplaced).

reduction problem. GPS has placed 25% of 249 Sardinians into their home villages, and 50% within 15 km from their villages (Elhaik et al. 2014). Perhaps this result prompted Das et al. to claim that Ashkenazi Jews have been located to their primeval villages. However, underlying data on nine admixture components among Sardinians (supplementary fig. 2, Supplementary Material online in Elhaik et al. 2014) lack structure: average proportions of components differ among villages by 2% at most (in absolute terms), and the pair of

villages most distant geographically (Villagrande and Sant’Antioco) differs by 2% of the Mediterranean, 2% of the Middle Eastern, and 1% of the North European component, with the other components identical (supplementary fig. S1 and table S1, Supplementary Material online). Five components that reach >1% in Sardinia are Mediterranean (58%), North European (19%), Middle Eastern (16%), South–East Asian (3%), and Sub-Saharan African (1%); their distributions are summarized by box plots in supplementary figure S2,

Supplementary Material online. Among 45 possible pairs of villages, the ANOVA test shows that only 2 pairs have significantly different fractions of the Mediterranean component (adjusted p -value <0.05), 10 pairs—of the North European; 13 pairs—of the Middle Eastern, 3 pairs—of the Sub-Saharan African, and no pairs—of the Southeast Asian component (**supplementary fig. S2, Supplementary Material** online). Apparently, placement of a quarter of Sardinians into their home villages was possible due to these differences in admixture profiles. However, variability of 1–2% in absolute values is generally considered as noise in admixture analyses, and depends much on dataset composition and on the number of algorithm iterations, among which the best one is selected. Thus, GPS results on the Sardinian population are, in our view, unreliable due to extreme data reduction.

Above we have considered conceptual problems that arise even with idealized densely spaced and stationary references. Coordinates inferred by GPS and reported as “ancestral” are influenced by positions of reference populations, which are rather sparse in reality. In the original GPS implementation (Elhaik et al. 2014), reference populations were almost lacking in some regions: USA, Canada, most of South America, Siberia, most of North Africa, Australia, and Southeast Asia were not covered. In published GPS results, quite a few individuals were mapped along straight lines connecting best proxies of their mixture partners (Elhaik et al. 2014): 1) many Italians and some Spanish were placed in Greece, on a line connecting the Italian and Lebanese reference populations; 2) all Tunisians and some Kuwaiti were placed in the Mediterranean Sea; 3) most Bermudians were placed on a line in the Atlantic Ocean, connecting the Bermudian and Yoruba reference populations; 4) most Puerto Ricans were placed on another line in the Atlantic Ocean, connecting the Puerto Rican and Spanish reference populations; 5) all Peruvians and Mexicans were placed on a line connecting the two countries, and crossing the Pacific Ocean. These cases are sufficient to illustrate that mapping of test individuals has little or nothing to do with ancestral locations (see fig. 2), but is determined by their collapsed mixture proportions and by coordinates of sparsely positioned references. In parts of the map more densely populated with references, positioning along straight lines is less common due to differential pull of ten genetically closest references (Das et al. 2016).

According to Das et al. (2016), the GPS approach introduced in the original publication was used without modification in the work on Ashkenazi Jews. Although the original reference dataset (Elhaik et al. 2014) was updated by Das et al. (2016), the sampling remained sparse, with just 26 reference populations. Not surprisingly, similar positioning artefacts are seen in the paper discussed here: 1) most Italians and apparently all Greeks were positioned in Bulgaria and in the Black Sea; 2) all Lebanese were scattered along a line connecting Egypt and the Caucasus; 3) all Nogais from the Caucasus and Pamiri Tajiks were placed in Turkmenistan or in the

Caspian Sea, to give just few examples. Therefore, one is left to wonder why so much weight is put on the inferred locations of Ashkenazi Jews in Bulgaria, amid the Black Sea, and in Turkey: “The Geographic Population Structure (GPS) analysis localized most Ashkenazi Jews along major primeval trade routes in northeastern Turkey adjacent to primeval villages with names that may be derived from Ashkenaz” (Das et al. 2016). Because Ashkenazi Jews represent a population with a clear admixture signature (Atzmon et al. 2010; Behar et al. 2010, 2013; Costa et al. 2013), their locations on the map produced by Das et al. (2016) form a gradient between mixture partners located in Europe and in the Middle East, or rather between their proxies among the modern reference populations. Moreover, genotypes of Ashkenazi Jews were obtained from a genetic testing company, ancestry of their grandparents was not controlled, and 86% individuals originated from the USA (Das et al. 2016). Therefore, recent European admixture in these Jewish samples is rather likely.

Major Problems of the Yiddish Relexification Theory

Based on overwhelming empirical evidence, modern linguistics generally defines primary evidence for genetic relationship of languages as 1) a significant number of etymological matches between their basic vocabularies, and 2) a significant number of etymological matches between their main grammatical exponents (such as number, case, person, etc.), see, for example, Campbell & Poser (2008).

The Germanic (or, more precisely, High German) affiliation of Yiddish is thus firmly based on two observations: 1) the Yiddish basic vocabulary is predominantly Germanic, and 2) the majority of grammatical exponents, including the main ones, are Germanic. This may be easily demonstrated by consulting such standardized basic wordlists as the 200-item wordlist of Morris Swadesh (where only a small handful of items are of Hebrew or Slavic origin), or the 700-item T. Kaufman’s basic concept list, only approximately 10% of which is of Slavic, and approximately 5% of Hebrew origin. Likewise, the majority of Yiddish grammatical exponents are also transparently Germanic (Jacobs et al. 1994), regardless of whether they are applied to indigenous Germanic or borrowed Slavic words (specific lexical and grammatical examples are listed in the linguistic supplement, **supplementary text S1, Supplementary Material** online). Consequently, there is a natural consensus in modern linguistics on the German affiliation of Yiddish (Rothstein 2006; Harbert 2007; Roberge 2010, *inter alia*), and it would take much more evidence than has been presented by Wexler to support the contrary assertion that Yiddish is a “fifteenth Slavic language” (Wexler’s original proposal, Wexler 1991) or even a “relexified Slavic language.” Although the Slavic component in Yiddish lexicon is indeed significant (ca., 5–10% overall), it predominantly represents cultural vocabulary. Likewise, despite Wexler’s claim that

“Yiddish grammar and phonology are Slavic (with some Irano–Turkic input)” (Das et al. 2016, similarly in Wexler 1991, 2010), he has managed to offer only a few grammatical/phonological matches between Yiddish and Slavic, generally confined to secondary grammatical features (such as semantic shifting of some German aspectual/spatial verbal prefixes and some nominal derivational suffixes towards the functions of their Slavic counterparts). Typological studies on language contact (Thomason & Kaufman 1988; van Coetsem 2000; Thomason 2001; Winford 2003) clearly suggest that all these phenomena may be optimally explained as later Slavic influence on Yiddish. In other words, Slavic languages functioned as adstrate and superstrate for Yiddish, rather than an underlying substrate (see [supplementary text S1, Supplementary Material](#) online for details).

A key point for Das et al. (2016) is that there are allegedly Iranian and Turkic loanwords in Yiddish which should indicate ancient contact between Yiddish and Anatolian communities (as Das et al. 2016 state: “a Slavic origin [of Yiddish] with strong Iranian and weak Turkic substrata”). But in reality, however, Wexler offers only a few Yiddish cultural words which eventually go back to Persian or Turkic forms, and all reliable cases represent areally diffused words which also happen to be spread across Slavic languages (see [supplementary text S1, Supplementary Material](#) online for some individual cases). Thus, such Yiddish terms are explainable as Slavic cultural loans, and there is no firm linguistic evidence for positing early Yiddish–Persian or Yiddish–Turkic contacts.

Conclusions

In our view, Das et al. have attempted to fit together a marginal and unsupported interpretation of the linguistic data with a genetic provenancing approach, GPS, that is at best only suited to inferring the most likely geographic location of modern and relatively unadmixed genomes, and tells little or nothing of population history and origin. Using explication of the GPS workflow and examples from the original GPS publication (Elhaik et al. 2014), as well as from the paper discussed here (Das et al. 2016) we find that this inference methodology provides no more information on an individual’s population origin than a few generations of family history. As opposed to GPS and similar tools (Kozlov et al. 2015) operating on highly reduced data, we advocate the use of more data-intensive and sophisticated approaches for the study of population history within the last 5,000 years: Rarecoal (Schiffels et al. 2016), ChromoPainter, fineSTRUCTURE (Lawson et al. 2012), and GLOBETROTTER (Hellenthal et al. 2014), among others.

Das et al. support the Khazar hypothesis of Ashkenazi ancestry, placing their alleged “Irano-Turko-Slavo Jewish merchants” within the Khazar Empire. Note that Das et al. designate this empire as the “Slavo-Iranian confederation”—a historically meaningless term invented by the authors under review. Having been popular in the mid-20th

century, the idea that the Khazars directly contributed to Ashkenazi ancestry is currently abandoned by practically all historians and linguists. To say more, according to a recent analysis of historical sources (Stampfer 2013), the conversion of Khazars to Judaism might have never happened, being a medieval legend. The Khazar hypothesis has previously been advocated in a genetic study (Elhaik 2013) reanalyzing autosomal SNP data from Behar et al. (2010). As no ancient DNA of Khazars was available, modern Armenians and Georgians were chosen by the author as genetic proxies for the ancient Khazar population (Elhaik 2013). This questionable choice of modern proxies may have biased the conclusions of the study, and a further analysis of a significantly extended dataset, spanning Europe, the Middle East, and the region historically associated with the Khazar Khaganate, has found no particular similarity of Ashkenazi Jews with populations from the Caucasus, including populations in the Khazar region (Behar et al. 2013). A large-scale study based on mitochondrial DNA (mtDNA) has also found no evidence for the Khazar hypothesis, estimating that >80% of Ashkenazi mtDNAs were probably assimilated within Europe, and virtually no mtDNAs were traced to the North Caucasus (Costa et al. 2013). A study focused on the Y chromosome has found strong support for the Near Eastern origin of a significant portion of Ashkenazi Y chromosomes (Rootsi et al. 2013). In summary, genetic studies support the traditional view on the history of the European Jewish diaspora: its Levantine origin, migration to the North Mediterranean followed by substantial local admixture, especially on the maternal side, and subsequent limited East European admixture in the Ashkenazi community (Atzmon et al. 2010; Behar et al. 2010, 2013; Costa et al. 2013; Rootsi et al. 2013).

Supplementary Material

Supplementary figures S1 and S2, text S1, and table S1 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

Acknowledgments

The authors are grateful to Asya Pereltsvaig, Alexandra Polyan, and especially to Alexis Manaster Ramer for long and productive discussions. P.F. was supported by the Institution Development Program of the University of Ostrava.

Literature Cited

- Alexander DH, Lange K. 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* 12:246.
- Allentoft ME, et al. 2015. Population genomics of Bronze Age Eurasia. *Nature* 522:167–172.
- Atzmon G, et al. 2010. Abraham’s children in the genome era: major Jewish diaspora populations comprise distinct genetic clusters with shared middle eastern ancestry. *Am J Hum Genet.* 86:850–859.
- Behar DM, et al. 2010. The genome-wide structure of the Jewish people. *Nature* 466:238–242.

- Behar DM, et al. 2013. No evidence from genome-wide data of a Khazar origin for the Ashkenazi Jews. *Hum Biol.* 85:859–900.
- Campbell L, Poser WJ. 2008. *Language classification: history and method*. New York: Cambridge University Press.
- Costa MD, et al. 2013. A substantial prehistoric European ancestry amongst Ashkenazi maternal lineages. *Nat Commun.* 4:2543.
- Das R, et al. 2016. Localizing Ashkenazic Jews to primeval villages in the ancient Iranian lands of Ashkenaz. *Genome Biol Evol.* 8:1132–1149.
- Elhaik E. 2013. The missing link of Jewish European ancestry: contrasting the Rhineland and the Khazarian hypotheses. *Genome Biol Evol.* 5:61–74.
- Elhaik E, et al. 2014. Geographic population structure analysis of worldwide human populations infers their biogeographical origins. *Nat Commun.* 5:3513.
- Flegontov P, et al. 2016. Genomic study of the Ket: a Paleo-Eskimo-related ethnic group with significant ancient North Eurasian ancestry. *Sci Rep.* 6:20768.
- Haak W, et al. 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522:207–211.
- Harbert W. 2007. *The Germanic languages*. Cambridge: Cambridge University Press.
- Hellenthal G, et al. 2014. A genetic atlas of human admixture history. *Science* 343:747–751.
- Jacobs NG, et al. 1994. Yiddish. In König E, Auwera J van der, editors. *The Germanic languages*. London: Routledge. p. 388–419.
- Kozlov K, et al. 2015. Differential Evolution approach to detect recent admixture. *BMC Genomics* 16(Suppl 8):S9.
- Lawson DJ, Hellenthal G, Myers S, Falush D. 2012. Inference of population structure using dense haplotype data. *PLoS Genet.* 8:11–17.
- Leslie S, et al. 2015. The fine-scale genetic structure of the British population. *Nature* 519:309–314.
- Roberge P. 2010. Contact and the history of Germanic languages. In Hickey R, editor. *The handbook of language contact*. Chichester: Wiley-Blackwell. p. 406–431.
- Roots S, et al. 2013. Phylogenetic applications of whole Y-chromosome sequences and the Near Eastern origin of Ashkenazi Levites. *Nat Commun.* 4:2928.
- Rothstein RA. 2006. Yiddish. In Brown K, editor. *Encyclopedia of language and linguistics*. 2nd ed. Vol. 13. Oxford: Elsevier. p. 730–733.
- Schiffels S, et al. 2016. Iron age and Anglo-Saxon genomes from East England reveal British migration history. *Nat Commun.* 7:10408.
- Stampfer S. 2013. Did the Khazars convert to Judaism? *Jewish Social Studies.* 19:1–72.
- Thomason SG, Kaufman T. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Thomason SG. 2001. *Language contact*. Edinburgh: Edinburgh University Press.
- van Coetsem F. 2000. *A general and unified theory of the transmission process in language contact*. Heidelberg: Winter.
- Wexler P. 1991. Yiddish — the fifteenth Slavic language. A study of partial language shift from Judeo-Sorbian to German. *Int J Soc Lang.* 1991:9–150.
- Wexler P. 2002. Two-tiered relexification in Yiddish: Jews, Sorbs, Khazars, and the Kiev-Polessian dialect. Berlin: Mouton de Gruyter.
- Wexler P. 2010. Do Jewish Ashkenazim (i.e., “Scythians”) originate in Iran and the Caucasus and is Yiddish Slavic? In Stadnik-Holzer E, Holzer G, editors. *Sprache und Leben der frühmittelalterlichen Slaven: Festschrift für Radoslav Katičić zum 80. Geburtstag*. Frankfurt: Peter Lang. p. 189–216.
- Winford D. 2003. *An introduction to contact linguistics*. Oxford: Blackwell.

Associate editor: Bill Martin