

Biogeography of crop progenitors and wild plant resources in the terminal Pleistocene and Early Holocene of West Asia, 14.7–8.3 ka

Joe Roe^{a,b,*}, Amaia Arranz-Otaegui^c

^a University of Bern, Switzerland,

^b University of Copenhagen, Denmark,

^c University of the Basque Country, Spain,

Abstract

This paper presents the first continuous, spatially-explicit reconstructions of the palaeodistributions of 82 plant species found regularly in association with early agricultural archaeological sites in West Asia, including the progenitors of the first crops. We used machine learning to train an ecological niche model of each species based on its present-day distribution in relation to climate and environmental variables. Predictions of the potential ranges of these species at key stages of the Pleistocene–Holocene transition could then be derived from these models using downsampled data from palaeoclimate simulations.

The models predict significant reductions and/or shifts in species ranges in the terminal Pleistocene and Early Holocene compared to present conditions. Many species that are found throughout the region’s ‘hilly flanks’ today are indicated to have much more restricted distributions centered on the Levant, Cyprus and Western Anatolia. In addition, overall ranges shrunk by an average of c. 25% from the terminal Pleistocene to the Early Holocene. The models performance in predicting the occurrence of specific species at archaeological sites is highly variable, but in aggregate the predictions are coherent and align with broad-scale trends in the ubiquity and relative abundance of species in the archaeological record.

1. Introduction

The Pleistocene–Holocene transition in West Asia marked a turning point in global environmental history, as humans brought the first plants under cul-

*Corresponding author
Email address: joeroe@hey.com (Joe Roe)

tivation and began modifying surrounding ecosystems to support their own subsistence. West Asia is part of the native range of a remarkable number of domesticable plant species, including wild relatives of wheat, barley, peas, lentils, and other crops of global importance (Harlan and Zohary, 1966; Diamond, 2002; Zohary et al., 2012). These species supported uniquely dense and complex Late Epipalaeolithic (15–11.7 ka) societies (Bar-Yosef, 1998; Maher et al., 2012) based on foraging (Harris and Hillman, 1989; Colledge, 2001; Weiss et al., 2004) and eventually plant management and pre-domestication cultivation (Colledge, 2001; Weiss et al., 2006; Harris, 2007; Willcox et al., 2008). The first agro-ecosystems emerged as these plants were domesticated in the Pre-Pottery Neolithic period (11.7–8.5 ka) and were shaped by the broader ecosystem in which it was embedded.

Decades of research in archaeobotany and zooarchaeology have reconstructed the subsistence economies of Late Epipalaeolithic and Neolithic sites in great detail. Together with studies of other environmental archaeological records and a variety of palaeoclimate archives (Jones et al., 2019), they also tell us much about the environments surrounding these settlements. However, each of these sources of evidence is subject to the wide variety of taphonomic and recovery biases that are inherent in any direct record of the past. They are also, by definition, records of the (human) environment at particular times and places. Interpolating these snapshots to give a holistic picture of the regional ecologies is not straightforward – to date, it has tended to rely on non-explicit, inductive modelling. The majority are also filtered through human action, producing a mixed signal that makes it difficult to disentangle anthropic effects from the background of environmental change in this period of rapid climatic alteration.

In this paper we present a complementary, deductive approach based on ecological niche modelling. Rather than inferring environmental conditions from preserved physical evidence, we predict the ranges of individual species relevant to human subsistence based on a model of their current environmental niche and simulations of past palaeoclimate. Though hypothetical, this gives us an independent line of evidence on past ecologies that is independent of the environmental archaeological and palaeoclimatic records. This means that the ancient data can be reserved for assessing the model's ability to ‘hindcast’ past conditions. In this sense, discrepancies between the two records are perhaps the most interesting result, as they indicate processes affecting one or both records that are not fully accounted for and therefore generate new questions. Our computational approach is also readily scaled up, allowing us to model spatially-explicit palaeodistributions for a large number of species, for the whole region, under multiple past climatologies.

2. Background

The transition to agriculture represents one of the most fundamental changes in human history. West Asia is one of the regions where this process has been

studied in the most detail: decades of research have traced the gradual development of a Neolithic way of life and the changes that occurred in the plant species and their geographical distribution as a result. Although archaeobotanical assemblages can be biased due to issues of preservation, sampling, recovery techniques, and lab procedures (Dennell, 1976; Hastorf and Popper, 1988)—and although they include not just food remains but plant resources that were used for other purposes or arrived at the site accidentally (Hastorf and Popper, 1988)—large-scale studies have still revealed coherent patterns in the exploitation of plants over time (Colledge et al., 2004; Arranz-Otaegui et al., 2016).

The possibility of an abrupt, geographically-constrained process of plant domestication was proposed in the 1990s (Hillman and Davies, 1990, 1992; Heun et al., 1997; Özkan et al., 2002) and developed as an explanatory model in the 2000s (Lev-Yadun et al., 2000a; Gopher et al., 2001; Abbo et al., 2005). As part of this model, some authors (Lev-Yadun et al., 2000a; Gopher et al., 2001; Abbo et al., 2010, 2012) argued that eight plant species, collectively referred to as ‘founder crops’ or the ‘Neolithic crop package’ (Zohary and Hopf, 1988) were selected and domesticated once, without any phase of pre-domestication cultivation (Abbo et al., 2011, p. 177). This process could have been rapid under strong artificial selection (Hillman and Davies, 1990, 1992) and may have occurred in a single region or ‘core area’ – generally located in southeast Turkey (Ladizinsky and Adler, 1976; Heun et al., 1997; Özkan et al., 2002; Ozkan et al., 2005; Mori, 2003; Luo et al., 2007). From this single point of origin, it was supposed that domesticated or semi-domesticated plants radiated outwards to other regions (Abbo et al., 2006; Kilian et al., 2007; Özkan et al., 2011).

The ‘short gestation’ paradigm was challenged by others (Helbæk, 1969; Harris, 1989; Kislev, 1989; Colledge, 2001; Weiss et al., 2004; Willcox et al., 2008; Fuller et al., 2018). Helbæk (in Kirkbride, 1966) argued that before the appearance of domesticated plants, a phase of cultivation of wild seeds must have taken place. The existence of a phase of cultivation of morphologically wild cereals or ‘pre-domestication cultivation’ was identified in the archaeological record through the study of plant domestication traits such as grain size, shattering v. non-shattering rachises. The archaeobotanical evidence shows that during the Pre-Pottery Neolithic A (PPNA) cereals exhibited sizes similar to those recorded in domestic species, but their dispersal mechanism was still the same as the one present in morphologically-wild species (i.e. shattering, see Kirkbride, 1966; Kislev, 1989; Hillman et al., 2001; Colledge, 2001; Willcox et al., 2008). This evidence suggested that wild cereal stands could have been cultivated for as much as a thousand years before non-shattering domestic forms became prevalent in the archaeological record (Tanno and Willcox, 2006, 2012; Arranz-Otaegui et al., 2016). Additional archaeobotanical (Colledge, 2001; Willcox et al., 2008, 2009; Riehl et al., 2013; Arranz-Otaegui et al., 2016; Weide et al., 2018; Douché and Willcox, 2018; Whitlam et al., 2018) and genetic data (Badr et al., 2000; Molina-Cano et al., 2005; Kilian et al., 2007; Özkan et al., 2011; Iob and Botigué, 2023) in recent years has further challenged the short-gestation model to explain the origins of plant domestication and agriculture in West Asia.

Similarly, the concept of a limited set of eight ‘founder crops’ ([Zohary and Hopf, 1988](#)) that were the first species cultivated, domesticated and then spread as the basis of Neolithic agricultural systems, is not supported by the latest evidence. Our previous analyses of the composition of available archaeobotanical datasets shows that these crops were of marginal importance during the Epipalaeolithic period ([Arranz-Otaegui et al., 2018](#)) and that Neolithic subsistence did not rely either solely or primarily on the exploitation of these species ([Arranz-Otaegui and Roe, 2023](#)). Instead, multiple species of grasses, legumes, fruits, nuts, and other plants were exploited over the Late Pleistocene-Early Holocene transition in southwest Asia.

2.1. Biogeography and agricultural origins

The study of the natural distribution of the progenitors of domesticated crops has been a central part of discussions on the origins of agriculture and plant domestication from the beginning. [von Humboldt \(1807\)](#) acknowledged the importance of the natural distribution of wild species to explain the origin and domestication of crops like spelt and rye. [de Candolle \(1886\)](#) integrated the study of plant ecology and biogeography and influenced [Darwin \(1859\)](#), who later reflected in detail about the geographical distribution of plants and species diversity. At that time, there was intense debate about whether there were single or multiple “centres of creation” of species. Researchers aimed to evaluate whether plant and animal species emerged in the same locations where they were currently distributed.

After Darwin, this early interest in crop origins evolved into more specific discussions about the “centres of plant domestication”. [Vavilov \(1926\)](#) was among the first to seek to determine the number of regions in which plants had been independently domesticated ([Harris, 1990](#)). His main method was ‘differential phytogeography’: he classified the variation within a crop and established the regions of maximum diversity, to locate the geographic regions in which crops originated. Using this method, Vavilov suggested that there were at least eight centres of origin. His work was later criticised by [Harlan \(1971\)](#), who argued that ‘centres of origin’ and ‘centres of diversity’ had to be separated. For Harlan, a ‘centre’ was as an “area in which things originate and out of which things are dispersed” ([Harlan, 1971](#), p. 468), and he suggested that three main centres of origin of domesticated crops existed. He further indicated that Vavilov’s approach to the question was simplistic and that more data proxies had to be considered (e.g. archaeology, history, geology), an approach more in the tradition of [de Candolle \(1886\)](#). Indeed, the inclusion of archaeobotany and genetics in the last decades, together with the study of wild relative distributions has been fundamental in characterising the origins of agriculture ([Fuller and Colledge, 2008](#)). As a result of modern interdisciplinary studies, the number of recognised centres of plant domestication has increased considerably, from the three centres suggested by Harlan in 1971 to the six to eight centres argued for in the 1990s ([Smith, 1995](#)) and up to as much as 24 potential centres reported in 2009 ([Purugganan and Fuller, 2009](#); see also [Fuller, 2010](#)).

Biogeographic research into the centres of origin and/or domestication of crops has also long informed broader understanding of the process of agricultural origins. In *Man Makes Himself*, Childe (1936) correctly located the centre of origin of European agriculture in the ‘Fertile Crescent’ of West Asia (unlike for example Pumpelly (1908) before him). This was not based on the region’s prehistoric archaeological record, which at the time had only been cursorily explored. Instead he was guided to the region by biogeographic work by Vavilov and Peake and Fleure (1927); only later was this prediction validated by archaeological work on the Epipalaeolithic and Neolithic of Palestine (Boyd, 2018). In subsequent decades, the search for more precise origin zones of specific domestic plants relied on the assumption that “the locus of domestication of a wild plant would presumably be within its area of original distribution in the wild state” (Butzer, 1971; paraphrasing Helbæk, 1959) – and that this “natural habitat” has not changed significantly over the last 12,000 years (Butzer, 1971).

Contemporary research on crop origins was pioneered by Harlan and Zohary, who compared the current distribution of the wild progenitors of domesticated plants in southwest Asia (Harlan and Zohary, 1966; Zohary, 1969, 1973; Zohary and Hopf, 1973) to the rapidly-expanding archaeobotanical record (Harlan, 1971; Harlan, 1977; see also Zohary and Hopf, 1988; Harlan and Zohary, 1966). They both were interested in evaluating which were the wild ancestors of domesticated crops and studying their natural distribution to understand their domestication process (Zohary, 1969, 1973; Zohary and Spiegel-Roy, 1975). Indeed, the natural distribution of the wild relatives of domestic plant species was later used as a criterion to infer ‘pre-domestic cultivation’ in the archaeological record. For example, the presence of seeds of chickpea at Jericho led Hopf (1986) to interpret the remains as cultivars, as the natural distribution of the wild form of chickpea was located further to the north. The same rationale was applied to the einkorn remains found at several Pre-Pottery Neolithic sites in the southern Levant (Hopf, 1969; Colledge, 2001), as the wild progenitors of this species was thought to be restricted to the northern Levantine area (Heun et al., 1997; Zohary et al., 2012). The same idea—presence of plants outside their natural range—has been repeated in the literature more recently by several other authors (Tanno and Willcox, 2006; Willcox et al., 2008; Hillman et al., 2001).

Despite the importance of biogeography in the development and validation of hypotheses regarding the origins of agriculture, there have been few studies of the wild range of specific crop progenitors or other relevant plant species (cf. for domestic animals, e.g. Yeomans et al., 2017). Observations regarding translocation or range expansion must therefore rely on a relatively rough and ahistoric notion of a species’ ‘natural distribution’ – that is, one based primarily on contemporary or recent-historic occurrences. Yet we know there has been considerable climatic and environmental change in West Asia since the terminal Pleistocene (Jones et al., 2019), so it is very unlikely that these ranges were in fact static. Reconstructions of broader environments have attempted to trace their fluctuations through time, either for the entire region (e.g. van Zeist and Bottema, 1991; Hillman in Moore et al., 2000) or parts of it (e.g. Cordova, 2007),

but these are at the level of the vegetation zone rather than individual species. They also invariably rely on what might be called ‘expert interpolation’ (where the author composes a map based on his or her own knowledge of the relevant data) rather than an explicit modelling process. This makes it difficult, if not impossible, for users of such reconstructions to understand exactly how they were derived or what could explain, for example, the significant discrepancies between the predictions of different experts.

2.2. Ecological niche modelling in archaeology

Ecological niche modelling or species distribution modelling is widely used by ecologists to predict the geographic range of a species based on a set of environmental predictors (Franklin and Miller, 2009). Essentially, it involves combining records of where an organism has been observed with environmental data (climate, topography, etc.) for those locations to model the range of environmental values at which that species – its environmental niche. This model can then be used to predict the range of the organism in question either in the same or a different environment. Townsend Peterson and Soberón (2012) suggests reserving the term ‘species distribution modelling’ for when the method is used to recover the verifiable range of a species in a real and existing environment, and using ‘ecological niche modelling’ as the broader term covering hypothetical or predictive applications – a convention we follow here when referring to predictive or ‘hindcast’ models of past ranges. Within this overarching framework, ecological niche modelling encompasses a wide range of applications and a variety of potential environmental predictors, modelling approaches, and methodologies, which we will not attempt to review here.

Ecological niche modelling has long been of interest to archaeologists as both a means of exploring the biological niche of humans and for reconstructing the past environments they inhabited (David Polly and Eronen, 2011; Franklin et al., 2015). In the first sense, it has been used most extensively to model the range of humans and other hominin species (e.g. Benito et al., 2017; Yousefi et al., 2020; Banks et al., 2021; Yaworsky et al., 2024a,b; Gurin et al., 2024), especially in the Palaeolithic. This overlaps with what archaeologists usually call generically ‘predictive modelling’ (Verhagen and Whitley, 2020)—or more precisely ‘site distribution modelling’—which is essentially the same approach as (and often borrows methodologies from) ecological niche modelling but applied to the occurrence of archaeological sites. Here what is modelled is not strictly a biological niche alone, but also aspects of human geography, taphonomy, and archaeological visibility. These applications can be distinguished from ‘palaeoecological niche modelling’, where the object of model remains, as in ecology, a non-human biological niche.

Franklin et al. (2015) review palaeoecological niche modelling and advocate for its greater adoption in environmental archaeology. In an early application to West Asia, Conolly et al. (2012) used the occurrence of wild and domestic *Bos* remains at prehistoric archaeological sites to map the evolving niche of cattle over the Pleistocene–Holocene transition. It has been used to model

the availability of fauna exploited by humans at wider scales (e.g. [de Andrés-Herrero et al., 2018](#); [Yaworsky et al., 2023](#)) and, in a West Asian context, of foraged plant resources in the landscape around the Neolithic sites on the Konya Plain ([Collins et al., 2018](#)). Modelling the spread of crops has been another significant archaeological application (e.g. [Krzyzanska et al., 2022](#); [Krzyzanska, 2023](#)), though not as yet applied to West Asia.

In the majority of studies to date (palaeo)ecological niche modelling has been applied to archaeological data in an ‘inductive’ fashion, i.e. faunal and botanical remains from ancient sites are used as the occurrence dataset for training a model using either past or present environmental data. However, both the zooarchaeological and archaeobotanical records are sparse and subject to a complex array of depositional, taphonomic and recovery biases factors that , many of which are not fully understood and/or cannot be corrected for. This means that while the archaeological attestation of the presence of a species might generally be relied upon, it is highly unlikely that its absence is representative of true past distributions.

The alternative approach is to train the model using contemporary occurrence and environmental data and then use palaeoenvironmental data to ‘hindcast’ its predictions backwards in time. Like [Franklin et al. \(2015\)](#), we view the hindcasting approach as more promising, because training datasets for both occurrences and environment are far more readily available, complete and reliable for the present than the past. There is some scepticism in the ecological niche modelling literature about the ability of such models to make accurate predictions in unknown environments (like the past, [Franklin et al., 2015](#)), but here the hindcasting approach also presents an opportunity: it reserves archaeological occurrence data as an independent dataset that can be used to assess the retrodictive performance of the model. This possibly was suggested by [Franklin et al. \(2015\)](#) but to our knowledge our study represents the first attempt to actually do so.

The major practical limitation of the hindcasting approach is that it relies on spatially explicit, high resolution palaeoenvironmental surfaces with continuous coverage of the region and periods of interest. Until recently, this has not been widely available for most applications, which is perhaps why only a minority of studies use it (cf. [Krzyzanska et al., 2022](#); [Yaworsky et al., 2023](#)). In this study, we are able to take advantage of the increasing availability of high resolution, global palaeoclimate data derived from simulation experiments with general circulation models of climate ([Brown et al., 2018, 2020](#); [Karger et al., 2023](#)).

3. Data and model

The aim of our study was to model the biogeography of species relevant to human subsistence economies in West Asia (excluding the Southern Arabian peninsula, see Figure 1) during the archaeological Late Epipalaeolithic (15–11.7

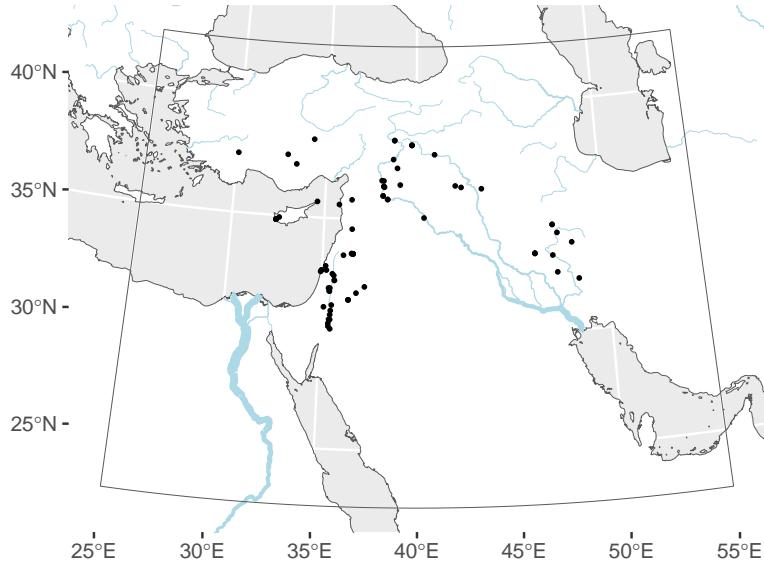


Figure 1: Late Epipalaeolithic & Pre-Pottery Neolithic archaeological sites used to generate modelled flora

ka) and Pre-Pottery Neolithic (11.7–8.3 ka) periods. Based on current understandings, we assume that plant-based subsistence during this period was broad, geographically- and temporally-varied, and reflects a gradual, geographically decentralised, and nonlinear transition to greater reliance on cultivars (i.e. agriculture, see Section 2). Our starting point was a list of 66 taxa (Table 1) comprising the identified species observed at at least 3 Late Epipalaeolithic/Pre-Pottery Neolithic sites, according to our previous study of the regional archaeobotanical data ([Arranz-Otaegui and Roe, 2023](#)). This was based on dataset collated from three previously published regional archaeobotanical databases: ADEMNES ([Riehl and Kümmel, 2005](#)), ORIGINS ([Wallace et al., 2018b](#)), and COMPAG ([Lucas and Fuller, 2018](#); [Fuller et al., 2018](#); based on [Colledge et al., 2004](#); [Shennan and Conolly, 2007](#)). We did not attempt to distinguish between the source of the remains (cf. [Wallace et al., 2018a](#)); archaeobotanical assemblages are subject to a variety of preservational and recovery biases, so by no means were all the species on our list consumed or even deliberately collected by people. However, we assume that their presence at a site of human settlement at least implies that they were part of the wider ecosystem that supported habitation there.

The taxonomic identifications of archaeobotanical material given in our source databases were previously controlled to ensure consistency between sources and to remove taxa that cannot be reliably distinguished (for details see [Arranz-Otaegui and Roe, 2023](#)). Taxonomic names were then matched to the canonical form specified in the GBIF Backbone Taxonomy ([GBIF Secretariat, 2023](#)) so

they could be related to modern occurrences. Archaeologically-attested domestic species meeting our inclusion criteria were substituted for their wild progenitors (where different) when gathering occurrence data, since the domestic forms are now widespread and presumably uninformative of the species original niche.

3.1. Occurrence data

```
Rows: 1 Columns: 11
-- Column specification -----
Delimiter: ","
chr (5): key, doi, license, status, downloadLink
dbl (3): size, totalRecords, numberDatasets
dttm (3): created, modified, eraseAfter

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Georeferenced occurrence data for West Eurasia between 0 and 60° of latitude was obtained from the Global Biodiversity Information Facility (GBIF) using via and the R package ‘rgbif’ (Chamberlain and Boettiger, 2017; Chamberlain et al., 2024). The GBIF dataset (?) excluded fossil occurrences, recorded absences, and records with missing or dubious coordinates. Although niche models have reasonable predictive power even with small training samples (Stockwell and Peterson, 2002; Hernandez et al., 2006; Wisz et al., 2008), we did not attempt to model 3 taxa with less than 40 usable occurrences, following recommendations for niche models generally and Random Forest-based models specifically (Stockwell and Peterson, 2002; Luan et al., 2020). Multiple records of the same taxon at the same coordinate were discarded because they do not impart information to the model. The resulting cleaned dataset used to train our niche models comprises 3.387196×10^6 from 4764 constituent datasets.

GBIF is currently the best available general-purpose occurrence dataset for the West Asia region, its coverage is uneven both geographically and from species to species. Figure 2 shows clearly that the Southern Levant, and Israel specifically, is significantly more densely sampled than other parts of West Asia.

Random Forest is a presence-absence approach to niche modelling and therefore requires not just data on where a species is present, but where it is definitely not present. However ‘absence’ data is rarely available because it requires exhaustive survey. In practice, most applications of niche modelling are ‘presence-only’ and, where absence data is required (as for Random Forest), it is supplied as a random background sample of points.¹ The purpose of this sample is to inform the model about the nature of the underlying environment. The stochastic generation process means that some of these points will overlap or fall close to

¹We use a background sample that is representative of the entire study area, as opposed to attempting to construct a ‘pseudo-absence’ sample. See ? for the distinction.

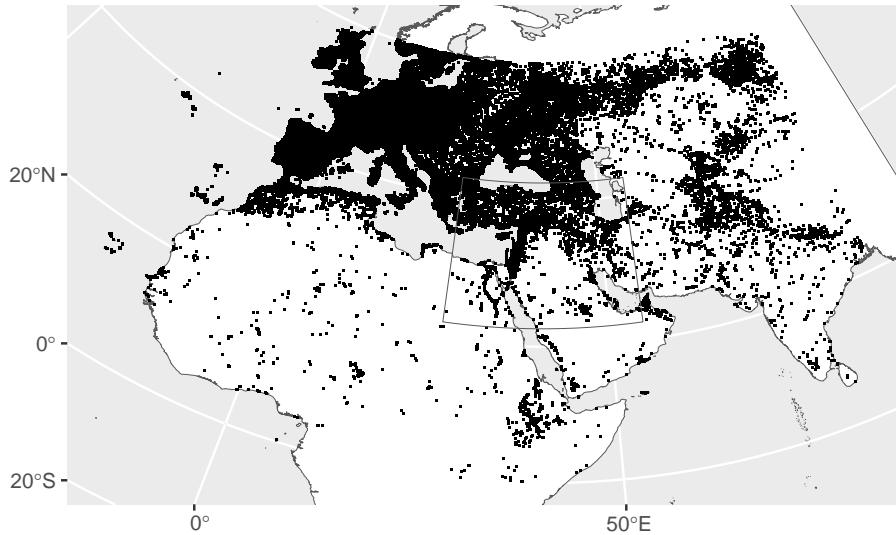


Figure 2: Georeferenced occurrence records from West Asia used to train models (N=1409939)

presences, so ensuring the model is not overly influenced by background samples is critical to its predictive importance ([Valavi et al., 2022](#)). Here we follow the advice of [Barbet-Massin et al. \(2012\)](#) for regression-based species distribution models and use a large (10000) uniform sample of points from across the land area of the study region. These points are then weighted equally against the presences in the regression to produce a ‘balanced Random Forest’ ([Valavi et al., 2022](#)).

3.2. Predictor data

We modelled the occurrence of species as a function of 24 geospatial predictor variables. These included:

```
bio_1,
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : GDAL Message 1: dimension #0 (period) is not a Time or
Vertical dimension.

Warning in parse_ncdf_meta(meta_data, get_names(x)): NAs introduced by
coercion

bio_10,
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : GDAL Message 1: dimension #0 (period) is not a Time or
Vertical dimension.
```

```
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : NAs introduced by coercion
bio_11,
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : GDAL Message 1: dimension #0 (period) is not a Time or
Vertical dimension.
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : NAs introduced by coercion
bio_12,
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : GDAL Message 1: dimension #0 (period) is not a Time or
Vertical dimension.
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : NAs introduced by coercion
bio_13,
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : GDAL Message 1: dimension #0 (period) is not a Time or
Vertical dimension.
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : NAs introduced by coercion
bio_14,
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : GDAL Message 1: dimension #0 (period) is not a Time or
Vertical dimension.
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : NAs introduced by coercion
bio_15,
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : GDAL Message 1: dimension #0 (period) is not a Time or
Vertical dimension.
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : NAs introduced by coercion
bio_16,
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : GDAL Message 1: dimension #0 (period) is not a Time or
Vertical dimension.
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : NAs introduced by coercion
bio_17,
```

```
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : GDAL Message 1: dimension #0 (period) is not a Time or
Vertical dimension.
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : NAs introduced by coercion
bio_18,
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : GDAL Message 1: dimension #0 (period) is not a Time or
Vertical dimension.
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : NAs introduced by coercion
bio_19,
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : GDAL Message 1: dimension #0 (period) is not a Time or
Vertical dimension.
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : NAs introduced by coercion
bio_2,
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : GDAL Message 1: dimension #0 (period) is not a Time or
Vertical dimension.
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : NAs introduced by coercion
bio_3,
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : GDAL Message 1: dimension #0 (period) is not a Time or
Vertical dimension.
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : NAs introduced by coercion
bio_4,
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : GDAL Message 1: dimension #0 (period) is not a Time or
Vertical dimension.
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : NAs introduced by coercion
bio_5,
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : GDAL Message 1: dimension #0 (period) is not a Time or
Vertical dimension.
Warning in CPL_read_gdal(as.character(x), as.character(options),
```

```

as.character(driver), : NAs introduced by coercion
bio_6,
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : GDAL Message 1: dimension #0 (period) is not a Time or
Vertical dimension.
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : NAs introduced by coercion
bio_7,
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : GDAL Message 1: dimension #0 (period) is not a Time or
Vertical dimension.
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : NAs introduced by coercion
bio_8,
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : GDAL Message 1: dimension #0 (period) is not a Time or
Vertical dimension.
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : NAs introduced by coercion
bio_9,
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : GDAL Message 1: dimension #0 (period) is not a Time or
Vertical dimension.
Warning in CPL_read_gdal(as.character(x), as.character(options),
as.character(driver), : NAs introduced by coercion

```

- Sixteen ‘bioclimatic’ variables derived from monthly temperature and precipitation values, following standard practice for species distribution models ([Hijmans et al., 2005](#)). Contemporary bioclimatic predictor data for West Asia was extracted from the global CHELSA dataset ([Karger et al., 2017](#)), which predicts temperature and precipitation from downscaled general circulation model output at 1 km resolution.
- Terrain aspect and slope, which at high resolution perform well as proxies for solar radiation when modelling plant occurrence ([Austin and Van Niel, 2011; Leempoel et al., 2015](#)); and the topographic wetness index (TWI), which serves as a proxy for soil moisture and is particularly important in modelling arid environments ([Kopecký and Čížková, 2010; Campos et al., 2016; Di Virgilio et al., 2018](#)). All three were derived from the SRTM30+ digital elevation model using algorithms from WhiteboxTools ([Lindsay, 2016](#)).

- Edaphic data from SoilGrids (Hengl et al., 2014, 2017), which improves model performance for plants (Dubuis et al., 2013; Mod et al., 2016; Velasco et al., 2017). Based on a recent assessment of the reliability of SoilGrids data for species distribution modelling (Miller et al., 2024), we used a subset of four variables relating to soil texture (clay, silt, sand) and pH at the surface (0–5 cm depth).

For hindcasting, we used reconstructed bioclimatic data for three key climatologies generated from downscaled paleoclimate simulations from the HadCM3 general circulation model (Fordham et al., 2017; Brown et al., 2018): the Bølling–Allerød (c. 14.7–12.9 ka), the Younger Dryas (c. 12.9–11.7 ka), and the Early Holocene (11.7–8.3 ka). Terrain and soil predictors were held constant, since reconstructions of these variables in the past are not available at sufficient scale. It is unlikely that either macroscale topography or soil characteristics have altered significantly over the period of time considered here, so we assume that this does not degrade model performance, and may in fact benefit it by providing ‘anchoring’ predictors that are independent of climate change.

For training, test predictions, and archaeological predictions predictor data was left in its native projection and resolution. For hindcast palaeodistributions, it was transformed to common equal-area projection and resolution of 5 km.

Warning: There were 57 warnings in `mutate()`.

The first warning was:

```
i In argument: `occ = map(occ, bind_extract, st_as_stars(hack(climate,
  "period", "cur")))`.
Caused by warning in `CPL_get_metadata()`:
! GDAL Message 1: dimension #0 (period) is not a Time or Vertical dimension.
i Run `dplyr::last_dplyr_warnings()` to see the 56 remaining warnings.
```

3.3. Random Forest

Ecological niche modelling is a classification problem that can be approached with a wide range of statistical methods. A substantial literature exists on the relatively performance of these approaches and their respective parameterisations (reviewed in Valavi et al., 2022). Random Forest, a widely-used machine learning algorithm, is amongst the best performing methods for presence-only species distribution models, providing it is appropriately parameterised to account for the class imbalance between presence and background samples (Valavi et al., 2021, 2022). For our application, it also has the advantage of requiring little to no manual parameter tuning to achieve good predictive results, which makes it easier to model a larger numbers of taxa.

For each taxon we trained a classification model to predict occurrence (presence or absence/background) based on up to 24 predictor variables (Section 3.2). Highly correlated (Pearson’s $r > 0.7$) predictors were removed on a taxon-by-taxon basis, to mitigate issues of overfitting due to collinearity (?), as were redundant predictors with zero variance. We used the Random Forest algorithm implemented in the R package ‘ranger’ (Wright and Ziegler, 2017) and

the ‘tidymodels’ ([Kuhn and Silge, 2022](#)) framework for data preprocessing and model selection. To avoid overfitting, we follow [Valavi et al. \(2021\)](#) in their recommended hyperparameters and use of down-sampling to balance presence and background samples. Models for each taxon were fit independently, with redundant zero-variance predictors excluded, and assessed based on balanced training (¾) and test (¼) partitions.

4. Model assessment

We trained Random Forest models for 63 taxa using contemporary occurrence data from GBIF, a random sample of background points, and the predictor variables described in Section 3.2. Substituting the “current” climate predictors for those derived from palaeoclimatic simulations ([Brown et al., 2018](#)), we could then generate hindcast predictions for reconstructed past environments in 4 key climate periods – a total of 252 modelled palaeodistributions. Predicted distributions for individual taxa are presented in the appendix and accompanying material.

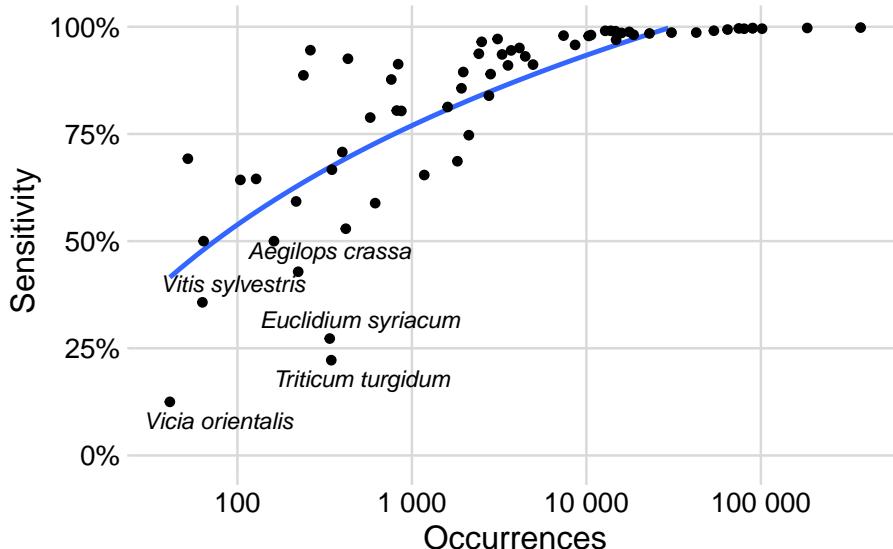


Figure 3: Model sensitivity by number of training occurrences

We assessed the predictive performance of the fitted niche models in the contemporary environment based on the reserved test partition Table 1. Model accuracy (proportion of correctly classified presence and background samples) ranged between 92% and 100%, with an average of 97%. Sensitivity (proportion of correctly classified presence samples) ranged between 12% and 100%, with an average of 82%. The area under the models’ receiver operating characteristic curves (ROC-AUC) was on average 0.987 ± 0.018 . Model sensitivity is loosely

correlated with the number of occurrences available for training (Figure 3), with the worst-performing models all having less than 300 recorded occurrences: *Lathyrus sativus*, *Lepidium perfoliatum*, *Triticum aestivum*, *Bolboschoenus glaucus*, *Gypsophila vaccaria*, *Euclidium syriacum*, *Gypsophila elegans*, *Citrullus colocynthis*, *Triticum monococcum*, *Triticum turgidum*, *Aegilops crassa*, *Secale strictum*, *Cicer reticulatum*, *Medicago astroites*, *Triticum durum*, *Gypsophila pilosa*, *Arnebia decumbens*, *Salvia absconditiflora*, *Suaeda fruticosa*, *Buglossoides tenuiflora*, *Lathyrus oleraceus*, *Vitis sylvestris*, and *Vicia orientalis*. Test metrics and ROC curves for the individual models are included in the appendix.

The ability of the hindcast models to predict the occurrence of specific species at archaeological sites is worse, with 10% of presences in archaeobotanical assemblages successfully predicted. Model sensitivity (proportion of corrected predicted presenses) in relation to the archaeological data is on average 0.07 ± 0.15 Table 1. A full assessment of the hindcasting performance of the individual models can be found in the appendix.

Table 1: Summary of species modelled

| Taxon | Occurrences | | Model | | | |
|---|-------------|----------------|-------|---------|-------|---------------|
| | Arch. | Cur. | Acc. | ROC-AUC | Sens. | Sens. (arch.) |
| <i>Aegilops crassa</i> | 4 | 223 | 0.98 | 0.98 | 0.43 | 0.00 |
| <i>Aizoanthemopsis hispanica</i> ¹ | 4 | 762 | 0.99 | 1.00 | 0.88 | 0.50 |
| <i>Ammi majus</i> | 4 | 10311 | 0.97 | 0.99 | 0.98 | 0.00 |
| <i>Androsace maxima</i> | 17 | 4946 | 0.94 | 0.98 | 0.91 | 0.06 |
| <i>Arenaria serpyllifolia</i> | 3 | 101612 | 0.98 | 0.98 | 1.00 | 0.00 |
| <i>Arnebia decumbens</i> | 21 | 616 | 0.97 | 0.97 | 0.59 | 0.05 |
| <i>Arnebia linearifolia</i> | 13 | 239 | 1.00 | 0.99 | 0.89 | 0.00 |
| <i>Atriplex prostrata</i> | 4 | 80181 | 0.98 | 0.99 | 1.00 | 0.00 |
| <i>Avena sterilis</i> | 5 | 15906 | 0.96 | 0.99 | 0.99 | 0.40 |
| <i>Bassia arabica</i> | 4 | 262 | 1.00 | 1.00 | 0.95 | 0.00 |
| <i>Bolboschoenus glaucus</i> | 5 | 418 | 0.98 | 0.97 | 0.53 | 0.00 |
| <i>Bolboschoenus maritimus</i> ² | 31 | 53809 | 0.98 | 1.00 | 0.99 | 0.00 |
| <i>Brachypodium distachyon</i> | 5 | 13800 | 0.98 | 1.00 | 0.99 | 0.20 |
| <i>Bromus sterilis</i> | 3 | 89764 | 0.99 | 0.99 | 1.00 | 0.00 |
| <i>Buglossoides arvensis</i> | 23 | 23004 | 0.95 | 0.98 | 0.98 | 0.09 |
| <i>Buglossoides tenuiflora</i> | 26 | 399 | 0.99 | 0.99 | 0.71 | 0.04 |
| <i>Capparis spinosa</i> | 4 | 8617 | 0.97 | 0.99 | 0.96 | 0.00 |
| <i>Carex divisa</i> | 9 | 7400 | 0.97 | 0.99 | 0.98 | 0.00 |
| <i>Chenopodium album</i> | 5 | 184403 | 0.99 | 0.98 | 1.00 | 0.00 |
| <i>Cicer reticulatum</i> ³ | 3 | 52 | 1.00 | 1.00 | 0.69 | 0.00 |
| <i>Citrullus colocynthis</i> | 4 | 1823 | 0.94 | 0.95 | 0.69 | 0.00 |
| <i>Crithopsis delileana</i> | 3 | 430 | 1.00 | 1.00 | 0.93 | 0.00 |
| <i>Euclidium syriacum</i> | 3 | 338 | 0.98 | 0.97 | 0.27 | 0.00 |
| <i>Ficus carica</i> | 8 | 64424 | 0.98 | 0.99 | 0.99 | 0.38 |
| <i>Fumaria densiflora</i> | 4 | 3704 | 0.96 | 0.99 | 0.94 | 0.00 |
| <i>Gypsophila elegans</i> | 3 | 1603 | 0.95 | 0.98 | 0.81 | 0.00 |
| <i>Gypsophila pilosa</i> | 5 | 348 | 0.99 | 0.99 | 0.67 | 0.00 |
| <i>Gypsophila vaccaria</i> ⁴ | 8 | 1177 | 0.95 | 0.98 | 0.65 | 0.00 |
| <i>Henrardia pubescens</i> | 3 | 1 ⁵ | — | — | — | — |
| <i>Hordeum bulbosum</i> | 4 | 4134 | 0.97 | 1.00 | 0.95 | 0.00 |

| | | | | | | |
|---|----|-----------------|------|------|------|------|
| <i>Hordeum murinum</i> | 5 | 74770 | 0.98 | 0.99 | 1.00 | 0.00 |
| <i>Hordeum spontaneum</i> ⁶ | 76 | 3098 | 0.99 | 1.00 | 0.97 | 0.20 |
| <i>Lathyrus aphaca</i> | 4 | 18652 | 0.96 | 0.99 | 0.98 | 0.00 |
| <i>Lathyrus oleraceus</i> ⁷ | 8 | 819 | 0.98 | 0.99 | 0.80 | 0.00 |
| <i>Lathyrus sativus</i> | 4 | 2120 | 0.92 | 0.96 | 0.75 | 0.00 |
| <i>Lepidium perfoliatum</i> | 3 | 2764 | 0.94 | 0.98 | 0.84 | 0.00 |
| <i>Linum bienne</i> ⁸ | 14 | 12824 | 0.98 | 0.99 | 0.99 | 0.07 |
| <i>Lolium rigidum</i> | 5 | 10620 | 0.97 | 0.99 | 0.98 | 0.00 |
| <i>Lolium temulentum</i> | 3 | 4463 | 0.94 | 0.98 | 0.93 | 0.00 |
| <i>Medicago astroites</i> ⁹ | 15 | 104 | 1.00 | 0.99 | 0.64 | 0.00 |
| <i>Medicago radiata</i> | 20 | 834 | 0.99 | 0.99 | 0.91 | 0.30 |
| <i>Phalaris paradoxa</i> | 3 | 3287 | 0.97 | 0.99 | 0.94 | 0.00 |
| <i>Phragmites australis</i> | 4 | 373041 | 0.99 | 0.99 | 1.00 | 0.00 |
| <i>Pistacia atlantica</i> | 6 | 2419 | 0.99 | 1.00 | 0.94 | 0.33 |
| <i>Poa bulbosa</i> | 5 | 30769 | 0.96 | 0.98 | 0.99 | 0.20 |
| <i>Polygonum arenarium</i> ¹⁰ | 7 | 25 ⁵ | — | — | — | — |
| <i>Polygonum corrigioloides</i> | 6 | 12 ⁵ | — | — | — | — |
| <i>Prosopis farcta</i> | 5 | 2512 | 0.99 | 1.00 | 0.96 | 0.60 |
| <i>Rumex pulcher</i> | 6 | 17606 | 0.97 | 0.99 | 0.99 | 0.00 |
| <i>Salsola kali</i> | 6 | 14699 | 0.98 | 1.00 | 0.99 | 0.00 |
| <i>Salvia absconditiflora</i> ¹¹ | 3 | 128 | 0.99 | 1.00 | 0.65 | 0.00 |
| <i>Secale cereale</i> | 4 | 14803 | 0.95 | 0.98 | 0.97 | 0.00 |
| <i>Secale strictum</i> ¹² | 3 | 162 | 0.99 | 0.99 | 0.50 | 0.00 |
| <i>Suaeda fruticosa</i> | 3 | 577 | 0.99 | 1.00 | 0.79 | 0.00 |
| <i>Taeniatherum caput-medusae</i> | 4 | 1969 | 0.96 | 0.99 | 0.89 | 0.00 |
| <i>Triticum aestivum</i> ¹³ | 4 | 217 | 0.99 | 0.98 | 0.59 | 0.00 |
| <i>Triticum durum</i> | 3 | 64 | 1.00 | 0.99 | 0.50 | 0.00 |
| <i>Triticum monococcum</i> ¹⁴ | 47 | 870 | 0.98 | 0.99 | 0.80 | 0.15 |
| <i>Triticum turgidum</i> ¹⁵ | 53 | 345 | 0.97 | 0.93 | 0.22 | 0.00 |
| <i>Verbena officinalis</i> | 3 | 89654 | 0.99 | 0.99 | 1.00 | 0.00 |
| <i>Vicia ervilia</i> | 26 | 1924 | 0.95 | 0.98 | 0.86 | 0.35 |
| <i>Vicia faba</i> | 7 | 42636 | 0.97 | 0.99 | 0.99 | 0.43 |
| <i>Vicia narbonensis</i> ¹⁶ | 3 | 2826 | 0.95 | 0.99 | 0.89 | 0.33 |
| <i>Vicia orientalis</i> ¹⁷ | 16 | 41 | 1.00 | 0.95 | 0.12 | 0.00 |
| <i>Vitis sylvestris</i> | 3 | 63 | 1.00 | 0.99 | 0.36 | 0.00 |
| <i>Zygophyllum fabago</i> | 3 | 3554 | 0.97 | 0.99 | 0.91 | 0.00 |

¹Including *Aizoon hispanicum*

²Including *Scirpus maritimus*

³Including *Cicer arietinum*

⁴Including *Vaccaria pyramidata*

⁵Excluded from modelling due to sample size

⁶Including *Hordeum vulgare*

⁷Including *Pisum sativum* and *Pisum elatius*

⁸Including *Linum usitatissimum*

⁹Including *Trigonella astroites*

¹⁰Including *Polygonum venantianum*

¹¹Including *Salvia cryptantha*

¹²Including *Secale montanum*

¹³Including *Triticum spelta* and *Triticum aestivocompactum*

¹⁴Including *Triticum boeoticum*

¹⁵Including *Triticum aestivum*, *Triticum dicoccum*, and *Triticum dicoccoides*

¹⁶Including *Vicia narbonense*

¹⁷Including *Lens culinaris* and *Lens orientalis*

5. Discussion

5.1. Reduction in range sizes over the Pleistocene/Holocene boundary

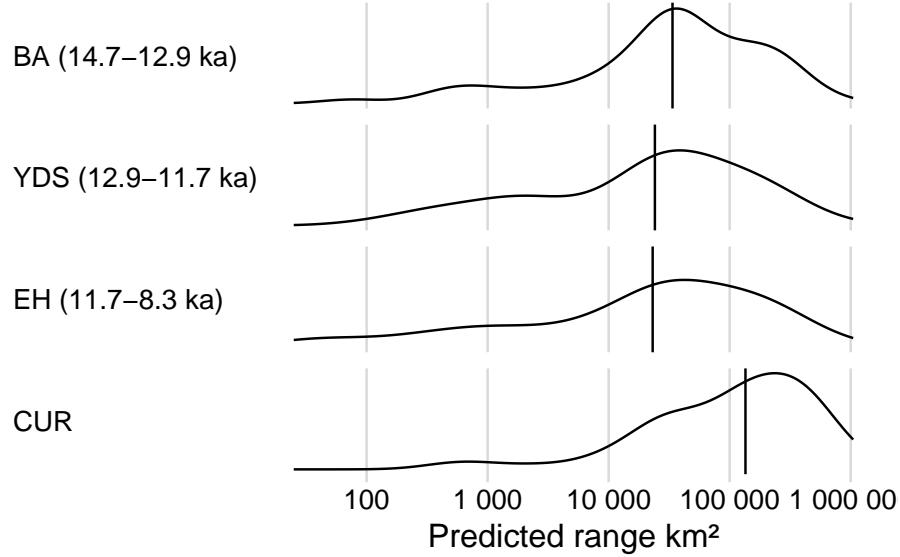


Figure 4: Distribution of predicted species ranges by period. Dashed lines indicate the median range.

Our reconstructed palaeodistributions (shown in full in the appendix) indicate that the majority of species had significantly different geographic ranges in the Late Pleistocene/Early Holocene compared today. 55 of 63 species are predicted reduced ranges in the past; 54 of more than 10% or more. Though the magnitude of the change in range size from prehistory to the present likely reflects a degree of overfitting in the model (discussed further in Section 5.3), fluctuations in modelled range size between the Bølling-Allerød (14.7–12.9 ka), Younger Dryas (12.9–11.7 ka), and Early Holocene (11.7–8.3 ka) are more directly comparable (Figure 4). The average range of modelled species was 31% in the Early Holocene compared to the Bølling-Allerød, and 28% during the Younger Dryas (i.e. ranges recovered slightly between the Younger Dryas and Early Holocene). This perhaps indicates that although this period is considered one of climatatic amelioration globally (Jones et al., 2019), the colder conditions of the Pleistocene may have supported more extensive plant-based economies in West Asia specifically.

Many taxa that occur (or are predicted to occur) across the ‘hilly flanks’ today—including most crop progenitors—are reconstructed to have had a significantly more restricted distribution in the terminal Pleistocene/Early Holocene (Figure 5). These include *Ficus carica* (fig); *Hordeum* spp. (wild barleys); *Lathyrus aphaca* and *L. sativus* (both marginally edible legumes); *Triticum aestivum compactum* (in the N. Levant), *T. monococcum aegilopoides*, *T. durum*,

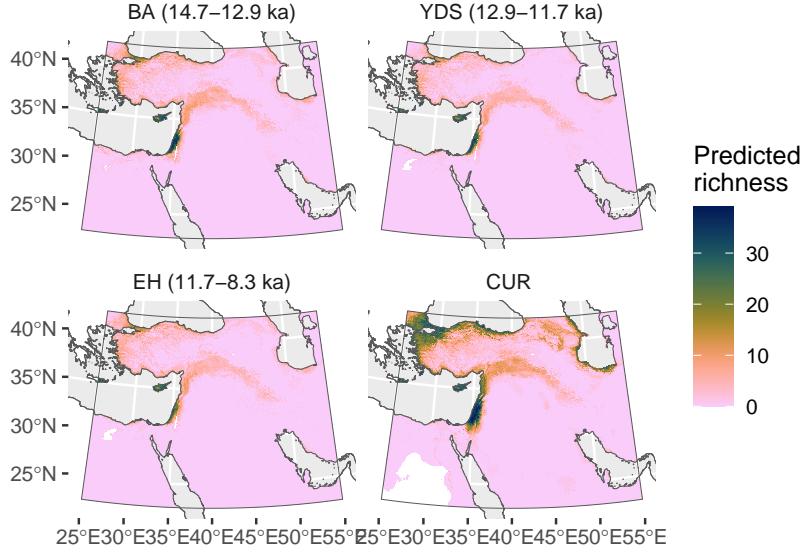


Figure 5: Predicted species richness (sum of predicted ranges) by period

and *Triticum urartu* (but not the other wheat progenitor, *T. turgidum dicoccum* – see Section 5.2); *Aegilops speltoides*, but not *Aegilops tauschii* (goatgrasses); and *Vicia* spp. (vetches), including *Vicia faba* (broad beans). Most of Anatolia, Northern Mesopotamia, and the Zagros Mountains in particular disappear from the predicted ranges of these species, leaving the Levant and to a lesser extent the Aegean and Cyprus as refugia.

Our results for the Levant are consistent with the current understanding of this region as developing early intensive foraging economies (the Natufian culture, Bar-Yosef, 1998) and as a centre of origin of agriculture (Zeder, 2011). Within the Levant, many species show moderate reductions in ranges over the Pleistocene/Holocene boundary, retreating from the Badia/transjordan region (e.g. barley, fig. ?@fig-palaeodist-barely-levant).

Loss of the Northern Mesopotamia–Anatolia region from the predicted ranges of crop progenitors is interesting in light of the ‘golden triangle’ hypothesis (Lev-Yadun et al., 2000b; Kozłowski and Aurenche, 2005; Abbo et al., 2010), which puts this region at the centre of the development of agriculture and plant domestication. Multiple lines of archaeological evidence have emerged that point away from this hypothesis and towards a more geographically diverse origin (Asouti, 2006; Fuller et al., 2011; Arranz-Otaegui et al., 2016), and our reconstructions are also consistent with the late arrival of intensive plant-based foraging economies in this region (cf. the Natufian of the Levant).

The near-absence of the Zagros in any predicted ranges is also surprising, given

mounting evidence that animal domestication took place just as early in the eastern *Mashriq* as it did in the west (Zeder, 2024). We consider that the most likely explanation for this is that our flora does not include the species that were most important to plant subsistence in the east. Archaeobotanical data on Neolithic sites in the Zagros is limited (compared to the Levant in particular) due to a hiatus in field research there from the 1980s to early 2010s (Zeder, 2024). Recent research (Riehl et al., 2013; Weide et al., 2017, 2018; Whitlam et al., 2018; González Carretero et al., 2023) indicates that plant subsistence in this region was based on a distinct set of species, compared to the Levant and Anatolia.

Cyprus and the Aegean are not conventionally considered part of the primary zone of domestication but rather amongst the first regions that acquired agriculture from West Asia. Our analysis complicates this picture, as it indicates that the wild ranges of many crop progenitors included these regions. Early examples of several domesticates are recorded at sites on Cyprus, Western Anatolia and Greece (Arranz-Otaegui and Roe, 2023), and the Aegean region was probably connected to West Asia by a land bridge via Anatolia until the Early Holocene (Aksu and Hiscott, 2022). Were these areas part of the same broader ‘interaction sphere’ that produce Neolithic agriculture in West Asia?

Exceptions to the dominant trend of range reduction include *Cicer reticulatum* (wild chickpea), which has a relatively stable range centered on Northern Mesopotamia; and *Triticum turgidum dicoccum* (wild emmer wheat), which is predicted to have two limited ranges centered around the Black Sea Coast of Anatolia and the Palmyra basin. In the latter case, neither of these areas are part of the predicted (by our model) modern distribution of wild emmer, which is centered around the Caucasus and Northern Mesopotamia. But it would be consistent with archaeological evidence for early cultivation at sites in the Upper Euphrates (Willcox, 2024).

5.2. Biogeography of crop progenitors

Almost all the cereal and legume crop progenitors we modelled are predicted to have only been found in the Levant during the terminal Pleistocene and Early Holocene (see appendix). Part of this may be due to the fact that both our initial flora and training occurrence dataset have a strong bias towards the southern Levant, but strikingly the modelled *current* ranges of these plants do tend to include Anatolia and the Zagros, so this cannot be the only factor. One notable exception is the wild ancestor of chickpea (*Cicer reticulatum*), which is predicted to have a distribution centered on Northern Mesopotamia but encompassing much of the ‘hilly flanks’ (except the southern Levant, Figure 6). Another is rye (*Secale cereale*), which is inferred to be primarily Anatolian (Figure 6). This is perhaps relevant to rye’s unusual domestication history, as a crop of West Asian origin that was intensively exploited (Hillman et al., 2001; Douché and Willcox, 2018) but apparently not first cultivated until much later than the ‘founder crops’, in Europe (Schreiber et al., 2021).

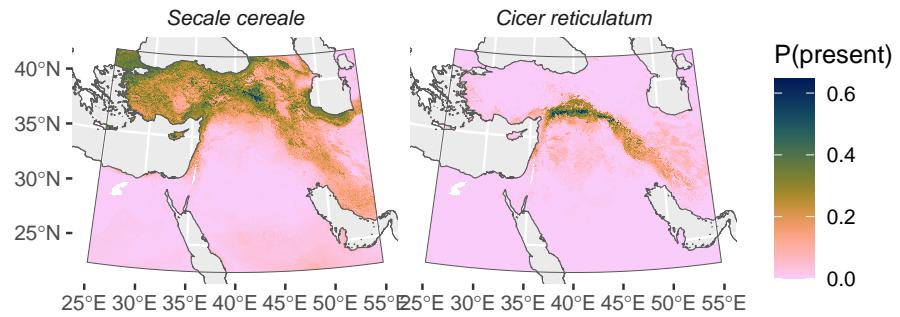


Figure 6: Predicted palaeodistribution of wild chickpea and rye in the Early Holocene (11.7–8.2 ka)

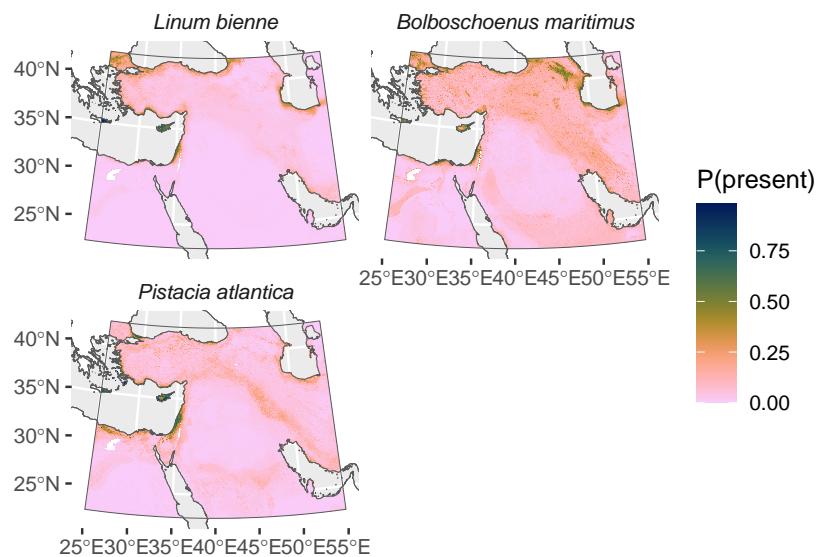


Figure 7: Predicted palaeodistribution of flax, pistachio and clubrush in the Early Holocene (11.7–8.2 ka)

Flax (*Linum bienne*) is predicted to have had a highly concentrated distribution in Cyprus and along the Mediterranean coast of the southern Levant. This is consistent with its low ubiquity in archaeobotanical assemblages (Arranz-Otaegui and Roe, 2023), despite conventionally being considered a ‘founder crop’, and presumably implies that its domestication was similarly geographically constrained. It is the only unambiguous crop progenitor with such a restricted range, though pistachio (*Pistacia atlantica*) and clubrush (*Bolboschoenus maritimus*) are similarly constrained to the Mediterranean coast (and North Africa, in the case of pistachio). This is despite the fact that they are well-attested in the archaeotanical record from across West Asia.

Wild barley (*Hordeum spontaneum*) and its relatives show a contraction of their predicted ranges from the Pleistocene to the Holocene, concurrent with it being brought into cultivation (?@fig-palaeodist-barely-levant). It also sees a marked decline in the archaeobotanical record from the Early PPNA/Early PPNB (where it was amongst the most common taxa) to the Late PPNB and Late Neolithic (Arranz-Otaegui and Roe, 2023). Pistachio (*Pistacia atlantica*) shows similar trends, but it is less certain that this species was managed in the Neolithic. Conversely, the various wild wheat species native to West Asia show almost no response to Pleistocene/Holocene climate change, even within the Levant, and in the archaeobotanical record wheat displays the opposite trend to barley and pistachio – becoming gradually more abundant through the course of the Neolithic and dominant by its end (Arranz-Otaegui and Roe, 2023).

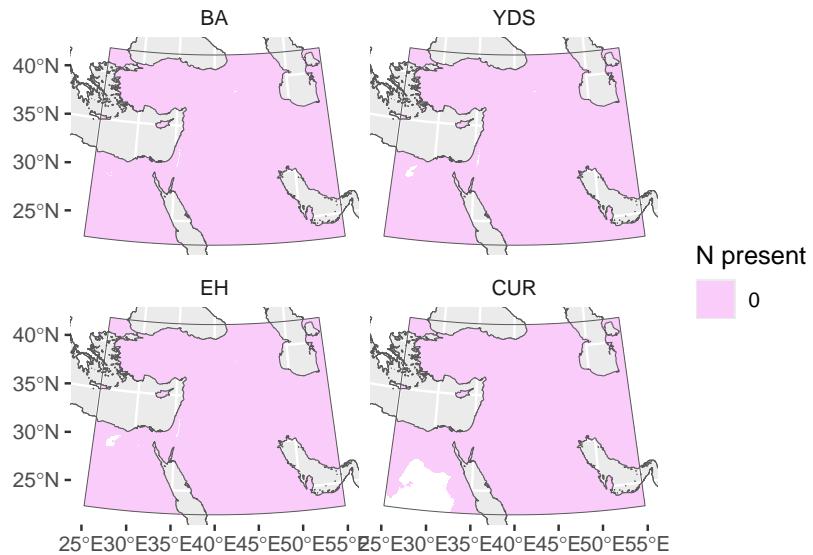


Figure 8: Combined predicted palaeodistributions of bread wheat progenitors

Bread wheat (*Triticum aestivum*), the most common wheat cultivar today,

has a complex ancestry that involves two recent hybridisation events (Levy and Feldman, 2022): most recently between domestic emmer (*Triticum turgidum dicoccum*) and a goatgrass (*Aegilops tauschii*) c. 9 ka, and before that, in emmer, between wild red einkorn (*T. urartu*) and another goatgrass (*Aegilops speltoides*). Although there is fairly wide overlap between these species today, according to our modelling the only place all four (or even just three) are predicted to have coincided or been found in close proximity to each other in the time frame of domestication is a narrow area around the Orontes river in the northern Levant (Figure 8). This area is c. 200 km from the only two sites in our archaeobotanical database where bread wheat (*T. aestivum*) co-occurs with both of its recent progenitors (*A. tauschii* and *T. turgidum dicoccum*) or both species of wheat and the goatgrasses are found together: Tell Abu Hureyra and El Kowm II (Arranz-Otaegui and Roe, 2023).² Taken together this suggests an origin of domesticated bread wheat, otherwise only loosely geographically constrained to the Levant–Upper Euphrates corridor (Levy and Feldman, 2022), is possibly occurred in this vicinity.

5.3. Hindcast models do not predict archaeobotanical composition

The failure of our hindcast models to predict the occurrence of species in archaeobotanical assemblages has several possible explanations. Since they do accurately predict the test dataset, a likely culprit is overfitting of the models to the present environment. This implies that the modelled palaeodistributions should be seen as conservative estimates or a minimal range. Another obvious flaw in our methodology is that the time slices used for palaeoclimatic reconstruction are very broad—each covering around two millennia—and therefore potentially unrepresentative of the environment around sites at the specific time at which they were occupied. The variable quality of the archaeological test dataset, especially in terms of chronology, is also a plausible factor.

For a variety of reasons, our models almost certainly underestimate the *fundamental* niche or potential distribution of the target taxon. These include:

- The uneven coverage of the GBIF occurrence dataset (see Section 3.1)

At the same time, we cannot rule out more substantive reasons for the discrepancy between predicted and observed archaeological occurrences. The niches of the modelled species could have changed since the Early Holocene, which would not be captured in a model trained purely on modern specimens. Human economic choices—mobility, foraging strategies, cultivation, etc.—could also produce archaeobotanical assemblages whose composition depart significantly from that of the surrounding local flora. Further refinement of the methodology for hindcast palaeoecological niche models, for example using more finely resolved palaeoclimate sequences (e.g. Karger et al., 2023),

²We are grateful to Benjamin Nowak for pointing this out to us.

hyperparameter tuning to avoid overfitting, and improved archaeological datasets, would help disentangle these potential explanations.

6. Conclusion

We present the first continuous, spatially explicit models of the palaeodistributions of 63 plant species found regularly in association with Late Epipalaeolithic and Early Neolithic sites in West Asia. This deductive approach—modelling the niche of a species based on its occurrence in relation to environmental factors today, and using this together with palaeoclimate simulations to infer its past distribution—represents a new line of evidence on the archaeoecology of the world’s first agricultural societies. It provides a complementary picture to that gleaned from environmental archaeology and climatological archives because it is independent of the taphonomic, anthropic, and recovery-related processes that affect these records.

“All models are wrong” (Box, 1976), but the performance of our models on independent test datasets give confidence in its predictions, which are plausible and consistent with broad-scale patterns in the archaeological record. Results with regard to specific archaeologically-attested species are not as promising, but this also reflects the incomplete and coarsely temporally-resolved nature of the archaeobotanical verification dataset. Discrepancies between the modelled distribution, whether on the broad scale (e.g. the restricted geographic range of most species compared to their attestation in the archaeological record), or relating to specific species (i.e. false positives and false negatives), suggest several avenues for future investigation.

Modelling a large number of species using machine learning, the substantial occurrence datasets available for the present day, and a hindcasting approach to past distributions also represents a significant advance in the methodology of palaeoecological niche modelling. This approach is enabled by the availability of high quality, global open datasets in ecology (GBIF, 2025; GBIF Secretariat, 2023), earth science (Farr et al., 2007), and climatology (Karger et al., 2017; Brown et al., 2018). Unfortunately in relation to these fields open data in archaeology lags conspicuously behind. Though we have benefited from the relatively long tradition of compiling archaeobotanical data in our region of study (Colledge et al., 2004; Shennan and Conolly, 2007; Riehl and Kümmel, 2005; Lucas and Fuller, 2018; Fuller et al., 2018; Wallace et al., 2018a,b), further development of open, comprehensive and up-to-date ‘backbone’ datasets on site locations and chronologies is needed to advance archaeoecological modelling to the same level.

7. Data availability

The data and R code used to produced is archived with Zenodo at <https://doi.org/10.5281/zenodo.14629984>

8. Acknowledgements

We are grateful to Alex Weide and the second, anonymous reviewer of our manuscript, whose suggestions have greatly improved this study.

This work was partially funded by the European Research Council (ERC-2021-STG project 101039060, *PalaeOrigins: Tracing the Epipalaeolithic origins of plant management in southwest Asia*).

References

- Abbo, S., Gopher, A., Peleg, Z., Saranga, Y., Fahima, T., Salamini, F., Lev-Yadun, S., 2006. The ripples of "The Big (agricultural) Bang": The spread of early wheat cultivation. *Genome* 49, 861–863. doi:[10.1139/g06-049](https://doi.org/10.1139/g06-049).
- Abbo, S., Gopher, A., Rubin, B., Lev-Yadun, S., 2005. On the Origin of Near Eastern Founder Crops and the 'Dump-heap Hypothesis'. *Genetic Resources and Crop Evolution* 52, 491–495. doi:[10.1007/s10722-004-7069-x](https://doi.org/10.1007/s10722-004-7069-x).
- Abbo, S., Lev-Yadun, S., Gopher, A., 2010. Agricultural Origins: Centers and Noncenters; A Near Eastern Reappraisal. *Critical Reviews in Plant Sciences* 29, 317–328. doi:[10.1080/07352689.2010.502823](https://doi.org/10.1080/07352689.2010.502823).
- Abbo, S., Lev-Yadun, S., Gopher, A., 2011. Origin of Near Eastern plant domestication: Homage to Claude Levi-Strauss and "La Pensée Sauvage". *Genetic Resources and Crop Evolution* 58, 175–179. doi:[10.1007/s10722-010-9630-0](https://doi.org/10.1007/s10722-010-9630-0).
- Abbo, S., Lev-Yadun, S., Gopher, A., 2012. Plant Domestication and Crop Evolution in the Near East: On Events and Processes. *Critical Reviews in Plant Sciences* 31, 241–257. doi:[10.1080/07352689.2011.645428](https://doi.org/10.1080/07352689.2011.645428).
- Aksu, A.E., Hiscott, R.N., 2022. Persistent Holocene outflow from the Black Sea to the eastern Mediterranean Sea still contradicts the Noah's Flood Hypothesis: A review of 1997–2021 evidence and a regional paleoceanographic synthesis for the latest Pleistocene–Holocene. *Earth-Science Reviews* 227, 103960. doi:[10.1016/j.earscirev.2022.103960](https://doi.org/10.1016/j.earscirev.2022.103960).
- Arranz-Otaegui, A., Colledge, S., Zapata, L., Teira-Mayolini, L.C., Ibáñez, J.J., 2016. Regional diversity on the timing for the initial appearance of cereal cultivation and domestication in southwest Asia. *Proceedings of the National Academy of Sciences* 113, 14001–14006. doi:[10.1073/pnas.1612797113](https://doi.org/10.1073/pnas.1612797113).
- Arranz-Otaegui, A., González Carretero, L., Roe, J., Richter, T., 2018. "Founder crops" v. wild plants: Assessing the plant-based diet of the last hunter-gatherers in southwest Asia. *Quat. Sci. Rev.* 186, 263–283. doi:[10.1016/j.quascirev.2018.02.011](https://doi.org/10.1016/j.quascirev.2018.02.011).

- Arranz-Otaegui, A., Roe, J., 2023. Revisiting the concept of the ‘Neolithic Founder Crops’ in southwest Asia. *Vegetation History and Archaeobotany* doi:[10.1007/s00334-023-00917-1](https://doi.org/10.1007/s00334-023-00917-1).
- Asouti, E., 2006. Beyond the Pre-Pottery Neolithic B interaction sphere. *Journal of World Prehistory* 20, 87–126. doi:[10.1007/s10963-007-9008-1](https://doi.org/10.1007/s10963-007-9008-1).
- Austin, M.P., Van Niel, K.P., 2011. Improving species distribution models for climate change studies: Variable selection and scale. *Journal of Biogeography* 38, 1–8. doi:[10.1111/j.1365-2699.2010.02416.x](https://doi.org/10.1111/j.1365-2699.2010.02416.x).
- Badr, A., M, K., Sch, R., Rabey, H.E., Effgen, S., Ibrahim, H.H., Pozzi, C., Rohde, W., Salamini, F., 2000. On the Origin and Domestication History of Barley (*Hordeum vulgare*). *Molecular Biology and Evolution* 17, 499–510. doi:[10.1093/oxfordjournals.molbev.a026330](https://doi.org/10.1093/oxfordjournals.molbev.a026330).
- Banks, W.E., Moncel, M.H., Raynal, J.P., Cobos, M.E., Romero-Alvarez, D., Woillez, M.N., Faivre, J.P., Gravina, B., d'Errico, F., Locht, J.L., Santos, F., 2021. An ecological niche shift for Neanderthal populations in Western Europe 70,000 years ago. *Scientific Reports* 11, 5346. doi:[10.1038/s41598-021-84805-6](https://doi.org/10.1038/s41598-021-84805-6).
- Bar-Yosef, O., 1998. The Natufian culture in the Levant, threshold to the origins of agriculture. *Evolutionary Anthropology* 6, 159–177. doi:[10.1002/\(SICI\)1520-6505\(1998\)6:5<159::AID-EVAN4>3.0.CO;2-7](https://doi.org/10.1002/(SICI)1520-6505(1998)6:5<159::AID-EVAN4>3.0.CO;2-7).
- Barbet-Massin, M., Jiguet, F., Albert, C.H., Thuiller, W., 2012. Selecting pseudo-absences for species distribution models: How, where and how many?: How to use pseudo-absences in niche modelling? *Methods Ecol. Evol.* 3, 327–338. doi:[10.1111/j.2041-210X.2011.00172.x](https://doi.org/10.1111/j.2041-210X.2011.00172.x).
- Benito, B.M., Svenning, J.C., Kellberg-Nielsen, T., Riede, F., Gil-Romera, G., Mailund, T., Kjaergaard, P.C., Sandel, B.S., 2017. The ecological niche and distribution of Neanderthals during the Last Interglacial. *J. Biogeogr.* 44, 51–61. doi:[10.1111/jbi.12845](https://doi.org/10.1111/jbi.12845).
- Box, G.E.P., 1976. Science and Statistics. *J. Am. Stat. Assoc.* 71, 791–799. doi:[10.1080/01621459.1976.10480949](https://doi.org/10.1080/01621459.1976.10480949).
- Boyd, B., 2018. Settled? Recent debates in the archaeology of the Epipalaeolithic and Pre-Pottery Neolithic of Southwest Asia. *Asian Archaeology* 1, 63–73. doi:[10.1007/s41826-018-0006-3](https://doi.org/10.1007/s41826-018-0006-3).
- Brown, J.L., Hill, D.J., Dolan, A.M., Carnaval, A.C., Haywood, A.M., 2018. PaleoClim, high spatial resolution paleoclimate surfaces for global land areas. *Sci Data* 5, 180254. doi:[10.1038/sdata.2018.254](https://doi.org/10.1038/sdata.2018.254).
- Brown, S.C., Wigley, T.M.L., Otto-Bliesner, B.L., Fordham, D.A., 2020. StableClim, continuous projections of climate stability from 21000 BP to 2100 CE at multiple spatial scales. *Scientific Data* 7, 335. doi:[10.1038/s41597-020-00663-3](https://doi.org/10.1038/s41597-020-00663-3).

- Butzer, K.W., 1971. Agricultural origins in the Near East as a geographical problem, in: Streuver, S. (Ed.), Prehistoric Agriculture. American Museum Sourcebooks in Anthropology.
- Campos, V.E., Cappa, F.M., Viviana, F.M., Giannomi, S.M., 2016. Using remotely sensed data to model suitable habitats for tree species in a desert environment. *Journal of Vegetation Science* 27, 200–210. doi:[10.1111/jvs.12328](https://doi.org/10.1111/jvs.12328).
- de Candolle, A., 1886. *The Origin of Cultivated Plants*. Cambridge University Press, Cambridge. doi:[10.1017/CBO9781139107365](https://doi.org/10.1017/CBO9781139107365).
- Chamberlain, S., Barve, V., Mcglinn, D., Oldoni, D., Desmet, P., Geffert, L., Ram, K., 2024. rgbif: Interface to the Global Biodiversity Information Facility API. URL: <https://CRAN.R-project.org/package=rgbif>. r package version 3.8.1.
- Chamberlain, S., Boettiger, C., 2017. R python, and ruby clients for gbif species occurrence data. PeerJ PrePrints URL: <https://doi.org/10.7287/peerj.preprints.3304v1>.
- Childe, V.G., 1936. *Man Makes Himself*. Watts & Co, London.
- Colledge, S., 2001. Plant Exploitation on Epipalaeolithic and Early Neolithic Sites in the Levant. BAR International Series, Oxbow, Oxford.
- Colledge, S., Conolly, J., Shennan, S., 2004. Archaeobotanical Evidence for the Spread of Farming in the Eastern Mediterranean. *Current Anthropology* 45, S35–S58. doi:[10.1086/422086](https://doi.org/10.1086/422086).
- Collins, C., Asouti, E., Grove, M., Kabukcu, C., Bradley, L., Chiverrell, R., 2018. Understanding resource choice at the transition from foraging to farming: An application of palaeodistribution modelling to the Neolithic of the Konya Plain, south-central Anatolia, Turkey. *J. Archaeol. Sci.* 96, 57–72. doi:[10.1016/j.jas.2018.02.003](https://doi.org/10.1016/j.jas.2018.02.003).
- Conolly, J., Manning, K., Colledge, S., Dobney, K., Shennan, S., 2012. Species distribution modelling of ancient cattle from early Neolithic sites in SW Asia and Europe. *Holocene* 22, 997–1010. doi:[10.1177/0959683612437871](https://doi.org/10.1177/0959683612437871).
- Cordova, C.E., 2007. *Millennial Landscape Change in Jordan: Geoarchaeology and Cultural Ecology*. University of Arizona Press, Tucson, AZ.
- Darwin, C., 1859. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. John Murray, London.
- David Polly, P., Eronen, J.T., 2011. Mammal Associations in the Pleistocene of Britain: Implications of Ecological Niche Modelling and a Method for Reconstructing Palaeoclimate, in: Ashton, N., Lewis, S.G., Stringer, C. (Eds.), *Developments in Quaternary Sciences*. Elsevier. volume 14, pp. 279–304. doi:[10.1016/B978-0-444-53597-9.00015-7](https://doi.org/10.1016/B978-0-444-53597-9.00015-7).

- de Andrés-Herrero, M., Becker, D., Weniger, G.C., 2018. Reconstruction of LGM faunal patterns using Species Distribution Modelling. The archaeological record of the Solutrean in Iberia. *Quat. Int.* 485, 199–208. doi:[10.1016/j.quaint.2017.10.042](https://doi.org/10.1016/j.quaint.2017.10.042).
- Dennell, R.W., 1976. The economic importance of plant resources represented on archaeological sites. *Journal of Archaeological Science* 3, 229–247. doi:[10.1016/0305-4403\(76\)90057-1](https://doi.org/10.1016/0305-4403(76)90057-1).
- Di Virgilio, G., Wardell-Johnson, G.W., Robinson, T.P., Temple-Smith, D., Hesford, J., 2018. Characterising fine-scale variation in plant species richness and endemism across topographically complex, semi-arid landscapes. *Journal of Arid Environments* 156, 59–68. doi:[10.1016/j.jaridenv.2018.04.005](https://doi.org/10.1016/j.jaridenv.2018.04.005).
- Diamond, J., 2002. Evolution, consequences and future of plant and animal domestication. *Nature* 418, 700–707. doi:[10.1038/nature01019](https://doi.org/10.1038/nature01019).
- Douché, C., Willcox, G., 2018. New archaeobotanical data from the Early Neolithic sites of Dja'de el-Mughara and Tell Aswad (Syria): A comparison between the Northern and the Southern Levant. *Paléorient* 44, 45–58. [arXiv:26595374](https://arxiv.org/abs/26595374).
- Dubuis, A., Giovanettina, S., Pellissier, L., Pottier, J., Vittoz, P., Guisan, A., 2013. Improving the prediction of plant species distribution and community composition by adding edaphic to topo-climatic variables. *Journal of Vegetation Science* 24, 593–606. doi:[10.1111/jvs.12002](https://doi.org/10.1111/jvs.12002).
- Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D., 2007. The Shuttle Radar Topography Mission. *Rev. Geophys.* 45, RG2004. doi:[10.1029/2005RG000183](https://doi.org/10.1029/2005RG000183).
- Fordham, D.A., Saltré, F., Haythorne, S., Wigley, T.M.L., Otto-Bliesner, B.L., Chan, K.C., Brook, B.W., 2017. PaleoView: A tool for generating continuous climate projections spanning the last 21 000 years at regional and global scales. *Ecography* 40, 1348–1358. doi:[10.1111/ecog.03031](https://doi.org/10.1111/ecog.03031).
- Franklin, J., Miller, J.A., 2009. Mapping Species Distributions: Spatial Inference and Prediction. Cambridge University Press.
- Franklin, J., Potts, A.J., Fisher, E.C., Cowling, R.M., Marean, C.W., 2015. Paleodistribution modeling in archaeology and paleoanthropology. *Quat. Sci. Rev.* 110, 1–14. doi:[10.1016/j.quascirev.2014.12.015](https://doi.org/10.1016/j.quascirev.2014.12.015).
- Fuller, D.Q., 2010. An emerging paradigm shift in the origins of agriculture. *General Anthropology* 17, 1–12.

- Fuller, D.Q., Colledge, S., 2008. Recent lessons from Near Eastern archaeobotany: Wild cereal use, pre-domestication cultivation and tracing multiple origins and dispersals. *Pragdhara* 18, 105–134.
- Fuller, D.Q., Lucas, L., Carretero, L.G., Stevens, C., 2018. From intermediate economies to agriculture: Trends in wild food use, domestication and cultivation among early villages in Southwest Asia. *Paléorient* 44, 59–74. [arXiv:26595375](https://arxiv.org/abs/26595375).
- Fuller, D.Q., Willcox, G., Allaby, R.G., 2011. Cultivation and domestication had multiple origins: Arguments against the core area hypothesis for the origins of agriculture in the Near East. *World Archaeol.* 43, 628–652. doi:[10.1080/00438243.2011.624747](https://doi.org/10.1080/00438243.2011.624747).
- GBIF, 2025. What is GBIF? <https://www.gbif.org/what-is-gbif>.
- GBIF Secretariat, 2023. GBIF Backbone Taxonomy. doi:[10.15468/390MEI](https://doi.org/10.15468/390MEI).
- González Carretero, L., Lucas, L., Stevens, C., Fuller, D.Q., 2023. Investigating early agriculture, plant use and culinary practices at Neolithic Jarmo (Iraqi Kurdistan). *Journal of Archaeological Science: Reports* 52, 104264. doi:[10.1016/j.jasrep.2023.104264](https://doi.org/10.1016/j.jasrep.2023.104264).
- Gopher, A., Abbo, S., Yadun, S.L., 2001. The “when”, the “where” and the “why” of the Neolithic revolution in the Levant. *Documenta Praehistorica* 28, 49–62. doi:[10.4312/dp.28.3](https://doi.org/10.4312/dp.28.3).
- Guran, S.H., Yousefi, M., Kafash, A., Ghasidian, E., 2024. Reconstructing contact and a potential interbreeding geographical zone between Neanderthals and anatomically modern humans. *Scientific Reports* 14, 20475. doi:[10.1038/s41598-024-70206-y](https://doi.org/10.1038/s41598-024-70206-y).
- Harlan, J.R., 1971. Agricultural Origins: Centers and Noncenters. *Science* 174, 468–474. doi:[10.1126/science.174.4008.468](https://doi.org/10.1126/science.174.4008.468).
- Harlan, J.R., 1977. The Origins of Cereal Agriculture in the Old World, in: Reed, C.A. (Ed.), *Origins of Agriculture*. De Gruyter Mouton, pp. 357–384. doi:[10.1515/9783110813487.357](https://doi.org/10.1515/9783110813487.357).
- Harlan, J.R., Zohary, D., 1966. Distribution of Wild Wheats and Barley. *Science* 153, 1074–1080. URL: <https://www.science.org/doi/abs/10.1126/science.153.3740.1074>. doi:[10.1126/science.153.3740.1074](https://doi.org/10.1126/science.153.3740.1074). publisher: American Association for the Advancement of Science.
- Harris, D.R., 1989. An evolutionary continuum of people–plant interaction, in: *The Emergence of Agriculture*. Routledge.
- Harris, D.R., 1990. Vavilov’s concept of centres of origin of cultivated plants: Its genesis and its influence on the study of agricultural origins. *Biological Journal of the Linnean Society* 39, 7–16. doi:[10.1111/j.1095-8312.1990.tb01608.x](https://doi.org/10.1111/j.1095-8312.1990.tb01608.x).

- Harris, D.R., 2007. Agriculture, Cultivation and Domestication: Exploring the Conceptual Framework of Early Food Production, in: Rethinking Agriculture. Routledge.
- Harris, D.R., Hillman, G.C., 1989. Introduction, in: Foraging and Farming: The Evolution of Plant Exploitation. Routledge.
- Hastorf, C.A., Popper, V.S., 1988. Current Paleoethnobotany: Analytical Methods and Cultural Interpretations of Archaeological Plant Remains. University of Chicago Press.
- Helbæk, H., 1969. Plant Collecting, Dry-Farming and Irrigation Agriculture in Prehistoric Deh Luran. volume 1. University of Michigan Ann Arbor.
- Helbæk, H., 1959. How farming began in the Old World. Archaeology 12, 183–189.
- Hengl, T., de Jesus, J.M., MacMillan, R.A., Batjes, N.H., Heuvelink, G.B.M., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Gonzalez, M.R., 2014. SoilGrids1km—global soil information based on automated mapping. PLoS One 9, e105992. doi:[10.1371/journal.pone.0105992](https://doi.org/10.1371/journal.pone.0105992).
- Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: Global gridded soil information based on machine learning. PLoS One 12, e0169748. doi:[10.1371/journal.pone.0169748](https://doi.org/10.1371/journal.pone.0169748).
- Hernandez, P.A., Graham, C.H., Master, L.L., Albert, D.L., 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29, 773–785. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0906-7590.2006.04700.x>. doi:[10.1111/j.0906-7590.2006.04700.x](https://doi.org/10.1111/j.0906-7590.2006.04700.x).
- Heun, M., Schäfer-Pregl, R., Klawan, D., Castagna, R., Accerbi, M., Borghi, B., Salamini, F., 1997. Site of Einkorn Wheat Domestication Identified by DNA Fingerprinting. Science 278, 1312–1314. doi:[10.1126/science.278.5341.1312](https://doi.org/10.1126/science.278.5341.1312).
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25, 1965–1978. doi:[10.1002/joc.1276](https://doi.org/10.1002/joc.1276).
- Hillman, G., Hedges, R., Moore, A., Colledge, S., Pettitt, P., 2001. New evidence of Lateglacial cereal cultivation at Abu Hureyra on the Euphrates. The Holocene 11, 383–393. doi:[10.1191/095968301678302823](https://doi.org/10.1191/095968301678302823).

- Hillman, G.C., Davies, M.S., 1990. Measured domestication rates in wild wheats and barley under primitive cultivation, and their archaeological implications. *Journal of World Prehistory* 4, 157–222. doi:[10.1007/BF00974763](https://doi.org/10.1007/BF00974763).
- Hillman, G.C., Davies, M.S., 1992. Domestication Rate in Wild Wheats and Barley under Primitive Cultivation: Preliminary Results and Archaeological Implications of Field Measurements of Selection Coefficient. *Monographie du CRA*, 113–158.
- Hopf, M., 1969. Plant remains and early farming in Jericho 1, in: The Domestication and Exploitation of Plants and Animals. Aldine, Chicago.
- Hopf, M., 1986. Archaeological Evidence of the Spread and Use of Some Members of the Leguminosae Family, in: Barigozzi, C. (Ed.), Developments in Agricultural and Managed Forest Ecology. Elsevier. volume 16 of *The Origin and Domestication of Cultivated Plants*, pp. 35–60. doi:[10.1016/B978-0-444-42703-8.50008-7](https://doi.org/10.1016/B978-0-444-42703-8.50008-7).
- Iob, A., Botigué, L., 2023. Genomic analysis of emmer wheat shows a complex history with two distinct domestic groups and evidence of differential hybridization with wild emmer from the western Fertile Crescent. *Vegetation History and Archaeobotany* 32, 545–558. doi:[10.1007/s00334-022-00898-7](https://doi.org/10.1007/s00334-022-00898-7).
- Jones, M.D., Abu-Jaber, N., AlShdaifat, A., Baird, D., Cook, B.I., Cuthbert, M.O., Dean, J.R., Djamali, M., Eastwood, W., Fleitmann, D., Haywood, A., Kwiecien, O., Larsen, J., Maher, L.A., Metcalfe, S.E., Parker, A., Petrie, C.A., Primmer, N., Richter, T., Roberts, N., Roe, J., Tindall, J.C., Ünal-İmer, E., Weeks, L., 2019. 20,000 years of societal vulnerability and adaptation to climate change in southwest Asia. *WIREs Water* 6, e1330. doi:[10.1002/wat2.1330](https://doi.org/10.1002/wat2.1330).
- Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, H.P., Kessler, M., 2017. Climatologies at high resolution for the earth's land surface areas. *Sci Data* 4, 170122. doi:[10.1038/sdata.2017.122](https://doi.org/10.1038/sdata.2017.122).
- Karger, D.N., Nobis, M.P., Normand, S., Graham, C.H., Zimmermann, N.E., 2023. CHELSA-TraCE21k – high-resolution (1 km) downscaled transient temperature and precipitation data since the Last Glacial Maximum. *Climate of the Past* 19, 439–456. doi:[10.5194/cp-19-439-2023](https://doi.org/10.5194/cp-19-439-2023).
- Kilian, B., Özkan, H., Walther, A., Kohl, J., Dagan, T., Salamini, F., Martin, W., 2007. Molecular Diversity at 18 Loci in 321 Wild and 92 Domestic Lines Reveal No Reduction of Nucleotide Diversity during *Triticum monococcum* (Einkorn) Domestication: Implications for the Origin of Agriculture. *Molecular Biology and Evolution* 24, 2657–2668. doi:[10.1093/molbev/msm192](https://doi.org/10.1093/molbev/msm192).

- Kirkbride, D., 1966. Five Seasons at the Pre-Pottery Neolithic Village of Beidha In Jordan. *Palestine Exploration Quarterly* doi:[10.1179/peq.1966.98.1.8](https://doi.org/10.1179/peq.1966.98.1.8).
- Kislev, M.E., 1989. Origins of the cultivation of *lathyrus sativus* and *L. cicera* (fabaceae). *Economic Botany* 43, 262–270. doi:[10.1007/BF02859868](https://doi.org/10.1007/BF02859868).
- Kopecký, M., Čížková, V.S.e., 2010. Using topographic wetness index in vegetation ecology: Does the algorithm matter? *Appl. Veg. Sci.* 13, 450–459. doi:[10.1111/j.1654-109X.2010.01083.x](https://doi.org/10.1111/j.1654-109X.2010.01083.x).
- Kozłowski, S.K., Aurenche, O., 2005. Territories, Boundaries and Cultures in the Neolithic Near East. Archaeopress.
- Krzyzanska, M., 2023. Modelling spatio-temporal changes in the ecological niches of major domesticated crops in China: Application of Species Distribution Modelling. Ph.D. thesis. University of Cambridge.
- Krzyzanska, M., Hunt, H.V., Crema, E.R., Jones, M.K., 2022. Modelling the potential ecological niche of domesticated buckwheat in China: Archaeological evidence, environmental constraints and climate change. *Vegetation History and Archaeobotany* 31, 331–345. doi:[10.1007/s00334-021-00856-9](https://doi.org/10.1007/s00334-021-00856-9).
- Kuhn, M., Silge, J., 2022. Tidy Modeling with R: A Framework for Modeling in the Tidyverse. O'Reilly.
- Ladizinsky, G., Adler, A., 1976. The origin of chickpea *Cicer arietinum* L. *Euphytica* 25, 211–217. doi:[10.1007/BF00041547](https://doi.org/10.1007/BF00041547).
- Leempoel, K., Parisod, C., Geiser, C., Daprà, L., Vittoz, P., Joost, S., 2015. Very high-resolution digital elevation models: Are multi-scale derived variables ecologically relevant? *Methods in Ecology and Evolution* 6, 1373–1383. doi:[10.1111/2041-210X.12427](https://doi.org/10.1111/2041-210X.12427).
- Lev-Yadun, S., Gopher, A., Abbo, S., 2000a. The Cradle of Agriculture. *Science* 288, 1602–1603. doi:[10.1126/science.288.5471.1602](https://doi.org/10.1126/science.288.5471.1602).
- Lev-Yadun, S., Gopher, A., Abbo, S., 2000b. The Cradle of Agriculture. *Science* 288, 1602–1603. doi:[10.1126/science.288.5471.1602](https://doi.org/10.1126/science.288.5471.1602).
- Levy, A.A., Feldman, M., 2022. Evolution and origin of bread wheat. *The Plant Cell* 34, 2549–2567. doi:[10.1093/plcell/koac130](https://doi.org/10.1093/plcell/koac130).
- Lindsay, J.B., 2016. Whitebox GAT: A case study in geomorphometric analysis. *Comput. Geosci.* 95, 75–84. doi:[10.1016/j.cageo.2016.07.003](https://doi.org/10.1016/j.cageo.2016.07.003).
- Luan, J., Zhang, C., Xu, B., Xue, Y., Ren, Y., 2020. The predictive performances of random forest models with limited sample size and different species traits. *Fisheries Research* 227, 105534. doi:[10.1016/j.fishres.2020.105534](https://doi.org/10.1016/j.fishres.2020.105534).

- Lucas, L., Fuller, D., 2018. Dataset: From intermediate economies to agriculture: Trends in wild food use, domestication and cultivation among early villages in southwest Asia.
- Luo, M.C., Yang, Z.L., You, F.M., Kawahara, T., Waines, J.G., Dvorak, J., 2007. The structure of wild and domesticated emmer wheat populations, gene flow between them, and the site of emmer domestication. *Theoretical and Applied Genetics* 114, 947–959. doi:[10.1007/s00122-006-0474-0](https://doi.org/10.1007/s00122-006-0474-0).
- Maher, L.A., Richter, T., Stock, J.T., 2012. The Pre-Natufian Epipaleolithic: Long-term Behavioral Trends in the Levant. *Evolutionary Anthropology: Issues, News, and Reviews* 21, 69–81. doi:[10.1002/evan.21307](https://doi.org/10.1002/evan.21307).
- Miller, T., Blackwood, C.B., Case, A.L., 2024. Assessing the utility of Soil-Grids250 for biogeographic inference of plant populations. *Ecology and Evolution* 14, e10986. doi:[10.1002/ece3.10986](https://doi.org/10.1002/ece3.10986).
- Mod, H.K., Scherrer, D., Luoto, M., Guisan, A., 2016. What we use is not what we know: Environmental predictors in plant distribution models. *Journal of Vegetation Science* 27, 1308–1322. doi:[10.1111/jvs.12444](https://doi.org/10.1111/jvs.12444).
- Molina-Cano, J.L., Russell, J.R., Moralejo, M.A., Escacena, J.L., Arias, G., Powell, W., 2005. Chloroplast DNA microsatellite analysis supports a polyphyletic origin for barley. *Theoretical and Applied Genetics* 110, 613–619. doi:[10.1007/s00122-004-1878-3](https://doi.org/10.1007/s00122-004-1878-3).
- Moore, A.M.T., Hillman, G.C., Legge, A.J., 2000. Village on the Euphrates: From Foraging to Farming at Abu Hureyra. Oxford University Press, Oxford.
- Mori, N., 2003. Origins of domesticated emmer and common wheat inferred from chloroplast DNA fingerprinting, in: Proc. 10th. Int. Wheat Genetics Symp., Paestum, Italy, 2003, pp. 25–28.
- Ozkan, H., Brandolini, A., Pozzi, C., Effgen, S., Wunder, J., Salamini, F., 2005. A reconsideration of the domestication geography of tetraploid wheats. *Theoretical and Applied Genetics* 110, 1052–1060. doi:[10.1007/s00122-005-1925-8](https://doi.org/10.1007/s00122-005-1925-8).
- Özkan, H., Brandolini, A., Schäfer-Pregl, R., Salamini, F., 2002. AFLP Analysis of a Collection of Tetraploid Wheats Indicates the Origin of Emmer and Hard Wheat Domestication in Southeast Turkey. *Molecular Biology and Evolution* 19, 1797–1801. doi:[10.1093/oxfordjournals.molbev.a004002](https://doi.org/10.1093/oxfordjournals.molbev.a004002).
- Özkan, H., Willcox, G., Graner, A., Salamini, F., Kilian, B., 2011. Geographic distribution and domestication of wild emmer wheat (*Triticum dicoccoides*). *Genetic Resources and Crop Evolution* 58, 11–53. doi:[10.1007/s10722-010-9581-5](https://doi.org/10.1007/s10722-010-9581-5).
- Peake, H., Fleure, H.J., 1927. Corridors of Time III: Peasants and Potters. Clarendon Press, Oxford.

- Pumpelly, R., 1908. Explorations in Turkestan. Number 73 in Carnegie Institution Publication, Carnegie Institution, Washington, DC.
- Purugganan, M.D., Fuller, D.Q., 2009. The nature of selection during plant domestication. *Nature* 457, 843–848. doi:[10.1038/nature07895](https://doi.org/10.1038/nature07895).
- Riehl, S., Kümmel, C., 2005. Archaeobotanical Database Of Eastern Mediterranean And Near Eastern Sites (ADEMNES).
- Riehl, S., Zeidi, M., Conard, N.J., 2013. Emergence of Agriculture in the Foothills of the Zagros Mountains of Iran. *Science* 341, 65–67. doi:[10.1126/science.1236743](https://doi.org/10.1126/science.1236743).
- Schreiber, M., Özkan, H., Komatsuda, T., Mascher, M., 2021. Evolution and Domestication of Rye, in: Rabanus-Wallace, M.T., Stein, N. (Eds.), The Rye Genome. Springer International Publishing, Cham, pp. 85–100. doi:[10.1007/978-3-030-83383-1_6](https://doi.org/10.1007/978-3-030-83383-1_6).
- Shennan, S.J., Conolly, J., 2007. Dataset: The Origin and Spread of Neolithic Plant Economies in the Near East and Europe.
- Smith, B.D., 1995. The Emergence of Agriculture. Scientific American Library, New York, NY.
- Stockwell, D.R.B., Peterson, A.T., 2002. Effects of sample size on accuracy of species distribution models. *Ecological Modelling* 148, 1–13. URL: <https://www.sciencedirect.com/science/article/pii/S030438000100388X>, doi:[10.1016/S0304-3800\(01\)00388-X](https://doi.org/10.1016/S0304-3800(01)00388-X).
- Tanno, K.i., Willcox, G., 2006. How Fast Was Wild Wheat Domesticated? *Science* 311, 1886–1886. doi:[10.1126/science.1124635](https://doi.org/10.1126/science.1124635).
- Tanno, K.i., Willcox, G., 2012. Distinguishing wild and domestic wheat and barley spikelets from early Holocene sites in the Near East. *Vegetation History and Archaeobotany* 21, 107–115. doi:[10.1007/s00334-011-0316-0](https://doi.org/10.1007/s00334-011-0316-0).
- Townsend Peterson, A., Soberón, J., 2012. Species Distribution Modeling and Ecological Niche Modeling: Getting the Concepts Right. *NatCon* 10, 102–107. doi:[10.4322/natcon.2012.019](https://doi.org/10.4322/natcon.2012.019).
- Valavi, R., Elith, J., Lahoz-Monfort, J.J., Guillera-Arroita, G., 2021. Modelling species presence-only data with random forests. *Ecography* 44, 1731–1742. doi:[10.1111/ecog.05615](https://doi.org/10.1111/ecog.05615).
- Valavi, R., Guillera-Arroita, G., Lahoz-Monfort, J.J., Elith, J., 2022. Predictive performance of presence-only species distribution models: A benchmark study with reproducible code. *Ecological Monographs* 92, e01486. doi:[10.1002/ecm.1486](https://doi.org/10.1002/ecm.1486).

- van Zeist, W., Bottema, S., 1991. Late Quaternary Vegetation of the Near East. Beihefte Zum Tübinger Atlas Des Vorderen Orients: Naturwissenschaften, Reichert.
- Vavilov, N., 1926. Studies on the Origin of Cultivated Plants. Institute of Applied Botany and Plant Breeding, Leningrad.
- Velazco, S.J.E., Galvão, F., Villalobos, F., De Marco Júnior, P., 2017. Using worldwide edaphic data to model plant species niches: An assessment at a continental extent. *PloS One* 12, e0186025. doi:[10.1371/journal.pone.0186025](https://doi.org/10.1371/journal.pone.0186025), arXiv:[29049298](https://arxiv.org/abs/29049298).
- Verhagen, P., Whitley, T.G., 2020. Predictive Spatial Modelling, in: Archaeological Spatial Analysis. Routledge.
- von Humboldt, A., 1807. Essai Sur La Géographie Des Plantes. Paris.
- Wallace, M., Jones, G., Charles, M., Forster, E., Stillman, E., Bonhomme, V., Livarda, A., Osborne, C.P., Rees, M., Frenck, G., Preece, C., 2018a. Re-analysis of archaeobotanical remains from pre- and early agricultural sites provides no evidence for a narrowing of the wild plant food spectrum during the origins of agriculture in southwest Asia. *Veg. Hist. Archaeobot.* doi:[10.1007/s00334-018-0702-y](https://doi.org/10.1007/s00334-018-0702-y).
- Wallace, M., Livarda, A., Charles, M., Jones, G., 2018b. Origins of agriculture: Archaeobotanical database. doi:[10.5284/1046750](https://doi.org/10.5284/1046750).
- Weide, A., Riehl, S., Zeidi, M., Conard, N.J., 2017. Reconstructing subsistence practices: Taphonomic constraints and the interpretation of wild plant remains at aceramic Neolithic Chogha Golan, Iran. *Vegetation History and Archaeobotany* 26, 487–504. doi:[10.1007/s00334-017-0607-1](https://doi.org/10.1007/s00334-017-0607-1).
- Weide, A., Riehl, S., Zeidi, M., Conard, N.J., 2018. A systematic review of wild grass exploitation in relation to emerging cereal cultivation throughout the Epipalaeolithic and aceramic Neolithic of the Fertile Crescent. *PLOS ONE* 13, e0189811. doi:[10.1371/journal.pone.0189811](https://doi.org/10.1371/journal.pone.0189811).
- Weiss, E., Kislev, M.E., Hartmann, A., 2006. Autonomous Cultivation Before Domestication. *Science* 312, 1608–1610. doi:[10.1126/science.1127235](https://doi.org/10.1126/science.1127235).
- Weiss, E., Wetterstrom, W., Nadel, D., Bar-Yosef, O., 2004. The broad spectrum revisited: Evidence from plant remains. *Proceedings of the National Academy of Sciences* 101, 9551–9555. doi:[10.1073/pnas.0402362101](https://doi.org/10.1073/pnas.0402362101).
- Whitlam, J., Bogaard, A., Matthews, R., Matthews, W., Mohammadifar, Y., Ilkhani, H., Charles, M., 2018. Pre-agricultural plant management in the uplands of the central Zagros: The archaeobotanical evidence from Sheikh-e Abad. *Vegetation History and Archaeobotany* 27, 817–831. doi:[10.1007/s00334-018-0675-x](https://doi.org/10.1007/s00334-018-0675-x).

- Willcox, G., 2024. Sowing, harvesting and tilling at the end of the Pleistocene/beginning of the Holocene in northern Syria: A reassessment of cereal and pulse exploitation. *Vegetation History and Archaeobotany* doi:[10.1007/s00334-023-00984-4](https://doi.org/10.1007/s00334-023-00984-4).
- Willcox, G., Buxo, R., Herveux, L., 2009. Late Pleistocene and early Holocene climate and the beginnings of cultivation in northern Syria. *The Holocene* 19, 151–158. doi:[10.1177/0959683608098961](https://doi.org/10.1177/0959683608098961).
- Willcox, G., Fornite, S., Herveux, L., 2008. Early Holocene cultivation before domestication in northern Syria. *Vegetation History and Archaeobotany* 17, 313–325. doi:[10.1007/s00334-007-0121-y](https://doi.org/10.1007/s00334-007-0121-y).
- Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A., Group, N.P.S.D.W., 2008. Effects of sample size on the performance of species distribution models. *Diversity and Distributions* 14, 763–773. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1472-4642.2008.00482.x>, doi:[10.1111/j.1472-4642.2008.00482.x](https://doi.org/10.1111/j.1472-4642.2008.00482.x).
- Wright, M.N., Ziegler, A., 2017. Ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software* 77, 1–17. doi:[10.18637/jss.v077.i01](https://doi.org/10.18637/jss.v077.i01).
- Yaworsky, P.M., Hussain, S.T., Riede, F., 2023. Climate-driven habitat shifts of high-ranked prey species structure Late Upper Paleolithic hunting. *Scientific Reports* 13, 4238. doi:[10.1038/s41598-023-31085-x](https://doi.org/10.1038/s41598-023-31085-x).
- Yaworsky, P.M., Hussain, S.T., Riede, F., 2024a. The effects of climate and population on human land use patterns in Europe from 22ka to 9ka ago. *Quaternary Science Reviews* 344, 108956. doi:[10.1016/j.quascirev.2024.108956](https://doi.org/10.1016/j.quascirev.2024.108956).
- Yaworsky, P.M., Nielsen, E.S., Nielsen, T.K., 2024b. The Neanderthal niche space of Western Eurasia 145 ka to 30 ka ago. *Scientific Reports* 14, 7788. doi:[10.1038/s41598-024-57490-4](https://doi.org/10.1038/s41598-024-57490-4).
- Yeomans, L., Martin, L., Richter, T., 2017. Expansion of the known distribution of Asiatic mouflon (*Ovis orientalis*) in the Late Pleistocene of the Southern Levant. *Royal Society Open Science* 4, 170409. doi:[10.1098/rsos.170409](https://doi.org/10.1098/rsos.170409).
- Yousefi, M., Heydari-Guran, S., Kafash, A., Ghasidian, E., 2020. Species distribution models advance our knowledge of the Neanderthals' paleoecology on the Iranian Plateau. *Scientific Reports* 10, 14248. doi:[10.1038/s41598-020-71166-9](https://doi.org/10.1038/s41598-020-71166-9).
- Zeder, M.A., 2011. The Origins of Agriculture in the Near East. *Curr. Anthropol.* 52, S221–S235. doi:[10.1086/659307](https://doi.org/10.1086/659307), arXiv:[10.1086/659307](https://arxiv.org/abs/10.1086/659307).

- Zeder, M.A., 2024. Out of the Shadows: Reestablishing the Eastern Fertile Crescent as a Center of Agricultural Origins: Part 1. *Journal of Archaeological Research* doi:[10.1007/s10814-024-09195-5](https://doi.org/10.1007/s10814-024-09195-5).
- Zohary, D., 1969. The progenitors of wheat and barley in relation to domestication and agricultural dispersal in the Old World, in: *The Domestication and Exploitation of Plants and Animals*. Routledge.
- Zohary, D., Hopf, M., 1973. Domestication of Pulses in the Old World. *Science* 182, 887–894. doi:[10.1126/science.182.4115.887](https://doi.org/10.1126/science.182.4115.887).
- Zohary, D., Hopf, M., 1988. Domestication of Plants in the Old World: The Origin and Spread of Domesticated Plants in Southwest Asia, Europe, and the Mediterranean Basin. 1st ed ed., Oxford University Press, Oxford.
- Zohary, D., Spiegel-Roy, P., 1975. Beginnings of Fruit Growing in the Old World. *Science* 187, 319–327. doi:[10.1126/science.187.4174.319](https://doi.org/10.1126/science.187.4174.319).
- Zohary, D., Weiss, E., Hopf, M., 2012. Domestication of Plants in the Old World: The Origin and Spread of Domesticated Plants in Southwest Asia, Europe, and the Mediterranean Basin. 4th ed ed., Oxford University Press, Oxford.
- Zohary, M., 1973. Geobotanical Foundations of the Middle East. Gustav Eischer, Stuttgart.