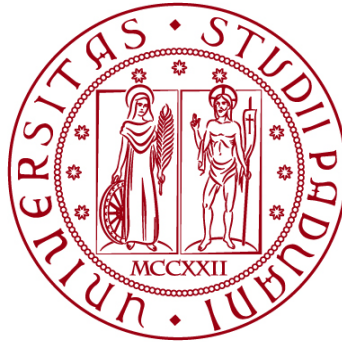


UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI BIOLOGIA

Corso di Laurea magistrale in Quantitative and Computational Biosciences



TESI DI LAUREA

Comparing Genetic and Archaeological Proxies for Reconstructing Past Human Demography in West Asia

Relatore: Prof. Luca Pagani
Dipartimento di Biologia

Correlatore: Hélios Delbrassine
Dipartimento di Biologia

Laureanda: Bryn Philotis

ANNO ACCADEMICO 2025/2026

1. **Introduction**
2. **Radiocarbon and Archaeological Data as Demographic Proxies**
3. **Genetics and Population Demography**
4. **Study Area and Anthropological Context**
5. **Objectives**
6. **Materials and *Methods***
 - 6.1 Initial Filtering and ROH calling (callROH)
 - 6.2 *Estimating Population Sizes (Ne) from Inferred ROH (estimateNe)*
 - 6.3 *Analysis of the Ne Trend and Comparative Analysis of Proxies*
7. **Results**
 - 7.1 *Binning and Confidence Interval Reduction - hapROH Exploration*
 - 7.2 *Regional ROH Distribution and ROH Time Series*
 - 7.3 *Analysis of the Ne Trend and Comparative Analysis of Proxies*
 - 7.31 *Anatolia*
 - 7.32 *Caucasus*
 - 7.33 *Iran*
 - 7.34 *Levant*
8. **Discussion & Conclusion**
 - 8.1 *Binning and Confidence Interval Reduction Exploration*
 - 8.2 *Analysis of the Ne Trend and Comparative Analysis of Proxies*
 - 8.3 *Limitations, Future Directions, and Conclusions*
9. **Acknowledgements**
10. **Supplementary Materials**
11. **References**

1. Introduction

It is commonly theorized that census size is a driver for both social and cultural change, especially in conjunction with environmental factors (Fogarty, 2017; van Schaik et al, 2019). However, modelling demographic changes has proven to be a challenge. Popular proxies of census size, estimates based on abundance of archaeological record or evidence of land use are generally good. However, limitations due to data availability or even their inherent nature, whether it be calibration or dating biases, data quality, or others, leave room for improvement and are thus frequently under scrutiny (Williams, 2012; Conteras and Meadows, 2014; Weitzel and Coddling, 2020). In addition, these proxies often show some degree of disagreement, leaving a conclusive “correct” model yet to be found, especially in the absence of a ground truth, which is a signature characteristic of all historical sciences. A completely different proxy to infer census size, under certain assumptions, may be provided by genetic-based demographic inferences. These were traditional hindered by the scarce availability from past populations. Luckily, with the onset of the Human Genome Project and new technologies such as high-throughput sequencing, we can not only incorporate genetic proxies into demographic analysis, but specifically that of ancient individuals. As such, novel approaches have been curated over the last few decades to infer demographic changes from genetic data, generally through the estimation of a population’s effective population size.

Effective population size is used to explain how a population’s genetic diversity has changed through time, potentially pointing to changes in the overall population and hence census size (Tenesa et al., 2007). More broadly, unlike census size, which is possibly even more complex to estimate for ancestral populations, effective population size can give us insights into complexities of a region’s demography, such as its relative size or inbreeding mechanics.

In the following dissertation, we evaluate estimates of genetic effective population sizes for six regions in Western Asia and compare them with archaeological settlement and radiocarbon data presented by Palmisano and Colleagues (2021), investigating their trend agreement and the historical and climatic events that they may coincide with.

2. Radiocarbon and Archaeological Data as Demographic Proxies

Oftentimes, a single demographic proxy is not enough to recreate a cohesive historical demographic reconstruction. Many publications have attempted to compare different proxies, and while there is often overlap and agreement, they are still imperfect and ultimately limit the certainty of the demographic in question (Palmisano et al, 2017; Palmisano et al, 2021). Two of the most popular of these proxies are archaeological data and radiocarbon data.

The first type of data, represented by archaeological site count and settlement size, is one of the most easily recorded demographic proxies, as it commonly derived from field investigation such as landscape survey or site mapping. However, these types of data are often subject to biases in archaeological research focus, whether it be period or region. In addition, climatic, cultural, and political practices may impact the preservation of these sites, resulting in identification or analysis which may not be entirely accurate due to missing or low-quality data. Moreover, this proxy relies on the assumptions that site count and size are directly proportional – which may not always be true, and that sites are consistently occupied – though this second point has been partly improved upon through simulations and the use of aoristic weights, whether a sample's statistical contribution is distributed proportionally to time series bins (Palmisano et al, 2017).

The second data type, known as summed probability distribution (SPD) of radiocarbon dates, has its own challenges. These radiocarbon dates often come from samples from combined archaeological contexts, including remnants of bone, charcoal, and seeds. An extensive number of independent dates can be gathered from this method, often resulting in a more than substantial dataset for demographic reconstruction. However, despite these high numbers, they are commonly effective only until 3000/2500 cal. yr. BP due to the preferred use of short-lived pottery types to subsequently date archaeological layers. Similarly, radiocarbon data is impacted by the Hallstatt radiocarbon calibration plateau (2750-2350 cal. yr. BP), which occurs roughly at the same time point, making the calibration of radiocarbon dates even more challenging (Palmisano et al, 2021). The use of this proxy as a direct representation of demographic change has also been challenged and has been proposed to rather represent energy consumption (Freeman et al, 2017).

In sum, both archaeological settlement data and radiocarbon have their strengths and weaknesses as proxies for demography, hence it may be useful to use both simultaneously when reconstructing historical demographics. To further strengthen this synchronous use of data, we propose the inclusion of a third demographic proxy: genetic effective population size, or N_e , inferred Runs of Homozygosity (ROHs).

3. Genetics and Demographic Proxies

Thanks to the rapid technological advancements in high throughput DNA sequencing, the human genome can be read with incredible accuracy, even from the smallest amounts of DNA. This has allowed for the development of new genetic methods to investigate demographic changes in populations over time, such as through the inference of effective population size, a proxy for census size in humans (Pankratov et al, 2020). Such methods include tracing the inheritance of Single Nucleotide Polymorphisms (SNPs), sex chromosome haplotyping, and analysing larger patterns of homo and heterozygosity. However, as these methods rely on large quantities of high-quality data, they are mostly suitable for present-day data, or few high-quality ancient samples. As a result, lower-quality ancient samples are often unable to be used with such methods, despite their potential value for reconstructing population demographics through time.

To rectify this problem, a novel method called hapROH was introduced by Ringbauer and colleagues in their 2021 publication “*Parental relatedness through time revealed by runs of homozygosity in ancient DNA*”. With it, one can detect and identify Runs of Homozygosity (ROH), even in low-quality, ancient DNA data. ROH are contiguous lengths of homogenous genotypes inherited from each parent that, once identified, can be used to infer historical inbreeding and effective population size at a given point in time (Kirin et al, 2010).

The frequency and length of ROH in a population are dependent on several demographic factors. ROH often accumulate in smaller populations with less genetic diversity, as they are inherited from individuals of the same ancestral background (Kirin et al, 2010). The length of each ROH, on the other hand, is determined in part by the number of generations since the last common ancestor, as they tend to be cut into smaller pieces through crossover events at each generation (Kirin et al, 2010, Ringbauer et al, 2021).

Thanks to a haplotype reference panel and an underlying Hidden Markov Model (HMM) algorithm, hapROH allows the calling of ROH of shorter lengths (4-8 cM), which can then be used to infer effective population size. These shorter ROH arise from parental ancestry that can date all the way back between and estimated 6.25 and 12.5 generations before an individual’s lifetime when considering the recombination rate of the human genome (Howrigan et al, 2011). Thus, their abundance or lack thereof reflect the effective population size over approximately the last 375 years (for a generation time of 30 years in humans) (Equation 1) (Howrigan et al, 2021).

$$g \approx \frac{50}{ROH \text{ length in cM}}$$

Equation 1: “Generations” from ROH length are approximately equal to 50 over the ROH length in centimorgans (Howrigan et al, 2011). 50 is half the length of a Morgan, which is 100 centimorgans. This is divided as there are two “generations”, being the paternal and maternal lineages. The standard crossover event per meiosis occurs once every Morgan. For a ROH length of 4, this is 6.25 generations. For a ROH length of 8, this is 12.5 generations.

4. Study Area and Anthropological Context

Genetic Samples of West Asia

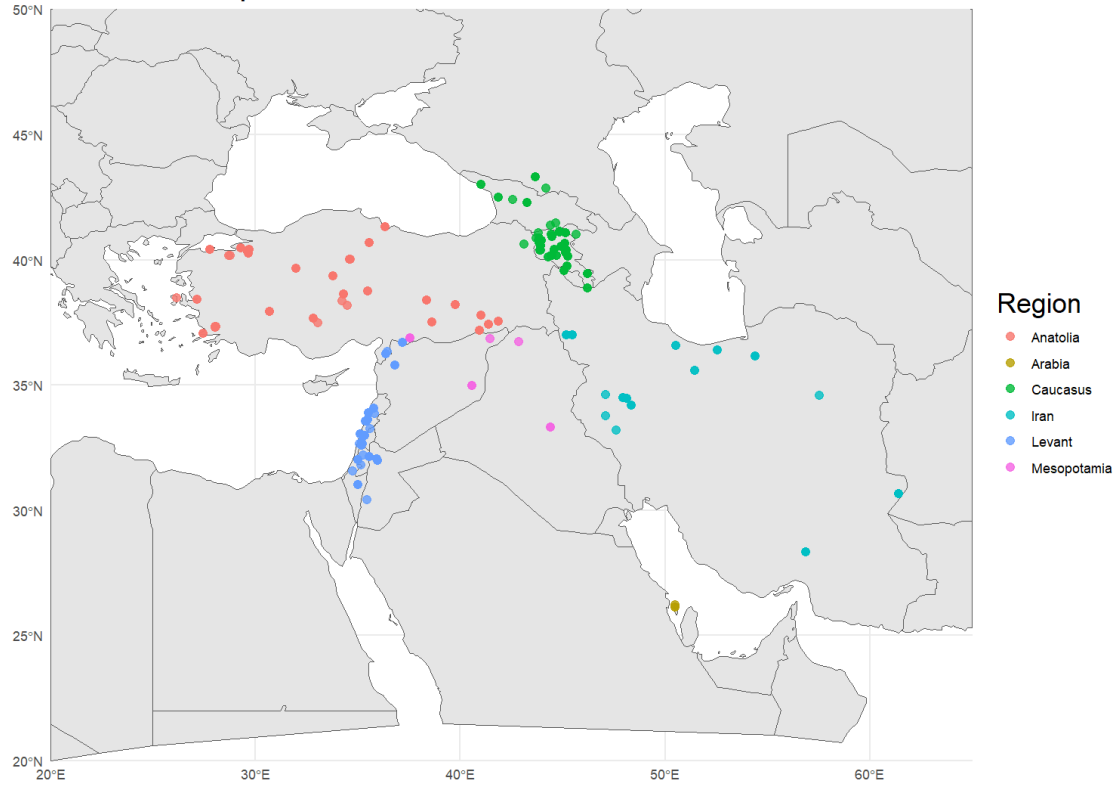


Figure 1: Map of the West Asia – Near East region analysed in this study. Each point is an eligible, individual sample taken from the AADR after initial SNP thresholding. Red: Anatolia, Yellow: Arabia, Green: Caucasus, Light Blue: Iran, Dark Blue: Levant, Pink: Mesopotamia.

In this study, we will investigate six regions of West Asia (Figure 1). These six regions are the same as those by Palmisano and Colleagues (2021) in their publication “Holocene regional population dynamics and climatic trends in the Near East: A first comparison using archaeo-demographic proxies”, that is: Anatolia, Arabia, Caucasus, Iran, the Levant, and Mesopotamia. Eventually, we only compare proxies of Anatolia, the Caucasus, Iran, and the Levant, while Arabia and Mesopotamia are excluded due to genetic data limitations. In the original publication, demographic reconstructions based on radiocarbon and

archaeological settlement proxies are compared to one another, as well as climatic variations or specific historical events. One of, if not the most, important of these events is the intensification of agriculture during the Bronze Age (5000 – 3200 cal. yr. BP), after the initial onset of agriculture at the beginning of the Neolithic (12000 cal. yr. BP). This event is often followed by a decrease in population size after the initial boom, which is also thought to be a manifestation of population stability (Palmisano et al, 2021). In addition, the five major climatic events are recorded by Palmisano and Colleagues (2021) at 10200 (10.2k), 9200 (9.2k), 8200 (8.2k), 4200 (4.2k), and 3200 (3.2k) cal. yr. BP. Each event is characterized by unique climatic dry-wet episodes that are thought to be at least in part responsible for the demographic shifts observed in West Asia. While these will not be elaborated in explicit detail in this report, these events caused climatic shifts that trended both towards more arid environments as well as more frequent or intensive dry and wet episodes.

5. Objectives

Both radiocarbon and settlement proxies are common proxies for understanding population size changes throughout time. While both have their advantages and often correlate, their limitations are evident when it comes to corroborating a larger picture. Here, we attempt to introduce and explore an additional proxy, effective population size (N_e) estimated from Runs of Homozygosity on ancient genomic data. This approach will be explored in its advantages and limitations before defining factors that make N_e estimations more confident. Then, it will be compared to see how well it corroborates to the other proxies previously explored by Palmisano and colleagues (2021).

6. Materials and Methods

6.1 Initial Filtering and ROH calling (callROH)

Genetic data was obtained from the Allen Ancient DNA Resource (AADR; Mallick and Reich, 2026). We downloaded a total of 855 individuals from 6 populations, Anatolia (N=237), Arabia (N=5), Caucasus (N=201), Iran (N=104), Levant (N=297), and Mesopotamia (N=11). Firstly, we filtered individuals based on their overall quality assessment (we selected the individuals labelled as “Passed” by the data curators), and we used as an inclusion threshold a SNP count equal to or exceeding 400,000 (Ringbauer et al, 2021). Any duplicates were subsequently removed, with the single copy being retained based on the highest SNP count. Lastly, both types of files were run under hapROH’s `e_model` “haploid” setting (following the advice of the program’s author, Harald Ringbauer, through personal communication with Hélios Delbrassine).

We performed initial ROH acquisition through hapROH using the resources and support provided by the BioData Hub at the Department of Biology, University of Padova. The primary Python script used was taken from the quick start vignettes provided by Ringbauer and colleagues (2021), under their "Documentation of hapROH and hapCon". The global 1000 Genome data (The 1000 Genomes Project Consortium, 2015) was used as the reference data panel, filtered down to bi-allelic 1240k SNPs. Once processed, two types of data were exported and used for further analysis: 1. the specific ROH data per individual ("roh_full"), and 2. individual summary data ("merged_full"). The "roh_full" datasets feature a list of all ROH per individual with corresponding length and chromosomal locations. The "merged_full" summary data summarize each individual's ROH characteristics, counting the number and total length of ROH within predetermined length bins, being ROH over 4, 8, 12, or 20 cM. The summarized data was further filtered to remove inbred individuals whose ROH > 20 cM exceeded a length of 50. Indeed, such individuals are indicative of extremely recent familial inbreeding. As they are outliers that reflect a uniquely inbred circumstance in an otherwise outbred population, including them in the computation of N_e might skew its estimates. Making the population appear smaller than it actually is (Ringbauer et al, 2021). This cutoff is based on analytical calculations and simulations by Ringbauer and colleagues, where this threshold detects a large fraction of close kin offspring while still taking into consideration background relatedness of small population sizes (Ringbauer et al, 2021). After filtering, the ROH data was then combined with all other individual feature data included in the 2026 AADR "anno" file (v66.pl_1240K.aadr.PUB.anno), such as radiocarbon dating (cal. yr. BP) and geographic coordinates.

Sum of ROH blocks was investigated using the supplementary material provided by Ringbauer and colleagues 2021 in their plotROH vignette. Two investigations were done. The first of such was the distribution of the sum of ROH among individuals for each region (Figure 2), which were compared to the expected distributions of background relatedness and size of ancestral backgrounds (see section 7.1 for more information)

The second of these was by plotting the sum of ROH 4-8 along a time series and the expected diploid population size ($2N_e$) for the density of ROH found (see section 7.1, Equation 2 for more information). Trends in the change of ROH density over time were investigated by plotting best-fit lines modelled by Lowess smoothing and Gaussian Process (GP) estimates. Both lines involved smoothed fits, but Lowess fits local regressions to nearby observations, while the GP makes more assumptions, particularly about observations are closer together in time (Cleveland, 1979; Rasmussen and Williams, 2006, Chapter 4).

6.2 Estimating Population Sizes (N_e) from Inferred ROH (*estimateNe*)

The ROH data obtained by hapROH was further processed in two manners, the first being N_e estimation based on single individuals. An N_e estimation was performed for each individual using their corresponding “roh_full” dataset, which were compiled into a single dataset per region (roh_vec). This was done using a Maximum Likelihood procedure to fit N_e within a confidence interval of 95 percent. The output resulted in a single N_e estimation per individual, based on each individual’s quantity and length of ROH between 4-8 cM (hereby referred to as ROH 4-8).

The second manner of N_e estimation was by binning the individuals into generational time bins before N_e estimation. An average generational time span of 30 years was decided, which was applied to four bin categories – 60 years for two generations, 90 for three, 120 for four, and 150 for five. Bins began at 0 cal. yr. BP and increase backwards in time in their respective increments (for example, from 0 to 150, 300, 450 years for 5 generations). Every individual was then assigned to a bin in each of these binning strategies according to their age. N_e estimation followed using the same Maximum Likelihood procedure, but considering the ROH information (quantity, length) of all individuals within a given bin,

6.3 Analysis of the N_e Trend and Comparative Analysis of Proxies

Exploratory analysis in the reduction of the confidence interval of N_e estimations was done for several features, being 1. Number of ROH within bin, 2. Generational bin size, and 3. Number of individuals within bins. This was done through visual representation using the ggplot package (scatter and bar plots), as well as statistically through generalized linear models.

After initial interpretation of N_e results, only the five-generational bin series (150 year time bins) was considered as the genetic proxy for further comparative analysis. This decision was based predominantly on the reduction of the confidence interval with the increase in the number of ROH 4-8 (see below (section 7.2, Figure 5) for further explanations).

Radiocarbon and settlement data were obtained from the supplementary materials provided by Palmisano et al 2021, “*Holocene regional population dynamics and climatic trends in the Near East: A first comparison using archaeo-demographic proxies*”. As elaborated upon in the original publication, the radiocarbon dates come from an archaeological context, where most of them are taken from samples of bone, charcoal, and seeds (Palmisano et al, 2021). In addition, they have been heavily filtered to exclude organic material with low confidence and/or poorly understood temporal offsets. While this data succeeds the minimum sample size

requirement to produce reliable SPD dates, it should also be noted that the dates are only effective until 3000/2500 cal. yr. BP due to the gradual replacement of radiocarbon data with archaeological findings as the predominant method for dating of archaeological strata (Palmisano et al, 2021). Archaeological data was combined into a single proxy using raw counts of archaeological sites and sums of estimated settlement data.

Radiocarbon data was directly binned into the same bin series as genetic data of 150 years, from 10,000 cal. yr. BP to 0 cal. yr. BP. Aoristic sum settlement data provided by Palmisano et al. 2021 was used as a single settlement proxy, a decision based on its consideration of the temporal uncertainty of occupation periods. Aoristic weights were recalibrated for Anatolia for 150-year bins, starting at 12,000 cal. yr. BP and ending at 0 cal. yr. BP. This time span is slightly shifted from the original (11,600 to 400 cal. yr. BP) to ensure the same 150-year time bin division as genetic and radiocarbon data. Levant settlement data, due to the unavailability of raw site data which is required for deliberate bin recalibration of aoristic weights, was instead recalibrated indirectly from the original 200-year time bins (10,000 to 2,600 cal. yr. BP) into 150-year time bins using a second layer of aoristic weights. This was done by generating aoristic weights for the new time series times based on the old ones and applying them to the appropriate new 150-year time bins - for example, by taking the weight of 0.75 of the 200 cal. yr. BP time bin and 0.25 weight of the 0 cal. yr. BP time bin to get an aoristic weight for the new 150 cal. yr. BP time bin. Both recalibrations resulted in aoristic sums for the given time bins.

The data from the areas of Anatolia, the Caucasus, Iran, and the Levant were used in comparative analysis. While Arabia and Mesopotamia were included in the original publication, as well as the subregional areas of the Southern Levant, and Upper Mesopotamia, they were excluded in this thesis after initial Ne time series evaluation analysis due to the lack of genetic data available for proxy comparison.

Final filtering was done to standardize the time series to be between 10,000 and 1,500 cal. yr. BP as each proxy was calibrated differently and this temporal range is what is covered by the archaeological dataset. In addition, the three proxies were normalized from 0 to 1 (min-max normalization) to be compared on the same scale and were smoothed using the Lowess method to account for noise and fluctuations and to create a smooth trend line with continuous values. Comparison of Ne against the two proxies, as well as the radiocarbon against settlement data when available, was done by plotting an “agreement graph” illustrating where the two trends agree in trend direction along the time series. This was done using the Lowess-smoothed model as opposed to the raw data to minimize noise and provide continuous values over time bins. All dates are recorded in “Calendaer Years Before Present (Cal. Yr. BP)”, based on radiocarbon date calibration.

7. Results

7.1. Binning and Confidence Interval Reduction - hapROH Exploration

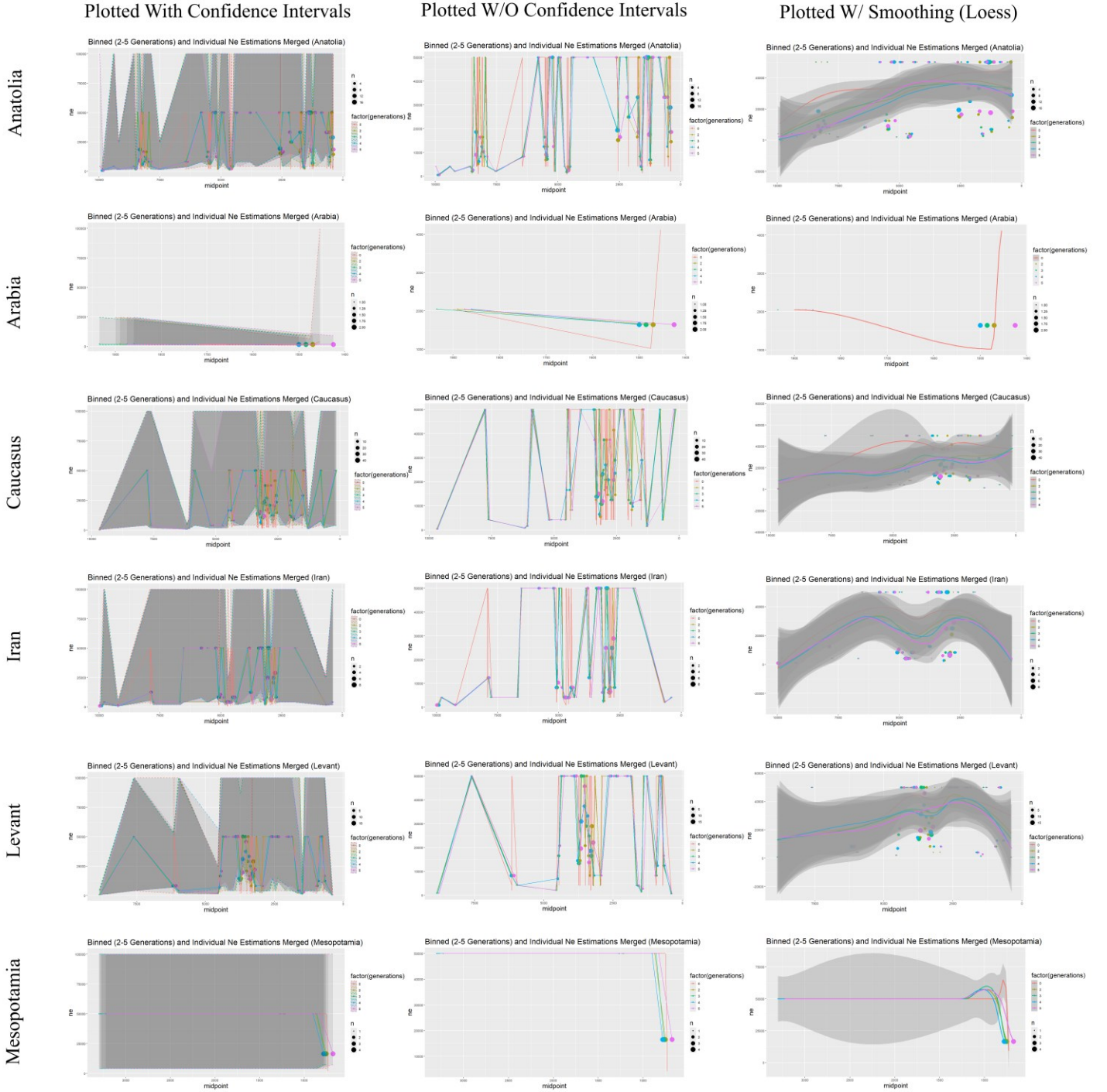


Figure 2: Ne time series trends when not binned (red) and binned generationally (2: yellow, 3: green, 4: blue, 5: pink). From left to right: Plotted with 95% confidence intervals, plotted without confidence intervals, and Lowess-smoothed trend. From top to bottom: Anatolia, Arabia, Caucasus, Iran, Levant, Mesopotamia. See Supplementary (Section 10) for larger images.

As a prerequisite to comparing demographic proxies, we investigate hapROH itself and the resulting Ne trends it produces for each region from the quantity and length of ROH 4-8 in samples. The resulting Ne estimation time series can be seen in Figure 2, where unbinned samples (0 generations, red line), and binned samples (generations: 2 - yellow; 3 - green; 4 - blue; 5 - pink) are plotted with their respective 95% confidence intervals. During this investigation, we sought to reduce the width of the confidence intervals, which were taken as a potential indicator of the robustness and accuracy of Ne inferences. Confidence intervals are considerably large (~100,000) for samples that have high Ne estimations (~50,000) and remain relatively constant regardless of bin size. For samples or groupings of samples that do not change with the increase in bin size, the confidence intervals remain the same, although often shifted slightly in time due to the increase in the bin width. For groupings that increase in size with bin size, the confidence interval, in some instances, appears to decrease (for example, Caucasus 5 generations at ~5200 cal. yr. BP). The inclusion of new individuals is highly dependent on the start and end date of the bin, however, and confidence intervals may appear to shift (e.g. reduction in bins 2 and 4, but not 3 and 5) because of the inclusion or omission of an individual at a threshold of two bins). For samples that were not binned where Ne estimations are calculated individually, confidence intervals are, on average, even higher. There are more samples that hit the maximum Ne estimation for hapROH, and thus also have the maximum confidence interval. Below we discuss three different variables of the dataset (number of ROH in groupings, bin size, and group size) to see if they impact the width of the confidence interval and, if so, how.

When considering the number of ROH >4 cM in a single individual (n), the confidence interval rapidly decreases. The lack of any ROH >4 cM, and thus any cue on deeper ancestry, estimates the maximum population size for hapROH (50000) with the maximum confidence interval (100000). With the inclusion of even one ROH >4, the width of the confidence decreases by over two-thirds, and with three, it decreases by over 80% (Figure 3), resulting in a more supported Ne estimation. This pattern is universal among all populations according to a standard GLM, using the CI as the dependent variable (All populations: $t = -36.72$, $P < 2e-16$, Anatolia: $t = -14.99$, $P < 2e-16$, Arabia: $t = -2.952$, $P = 0.032$, Caucasus: $t = 25.19$, $P < 2e-16$, Iran: $t = 13.86$, $P < 2e-16$, Levant: $t = -22.60$, $P < 2e-16$, Mesopotamia: $t = 5.749$, $P < 2.4e-05$).

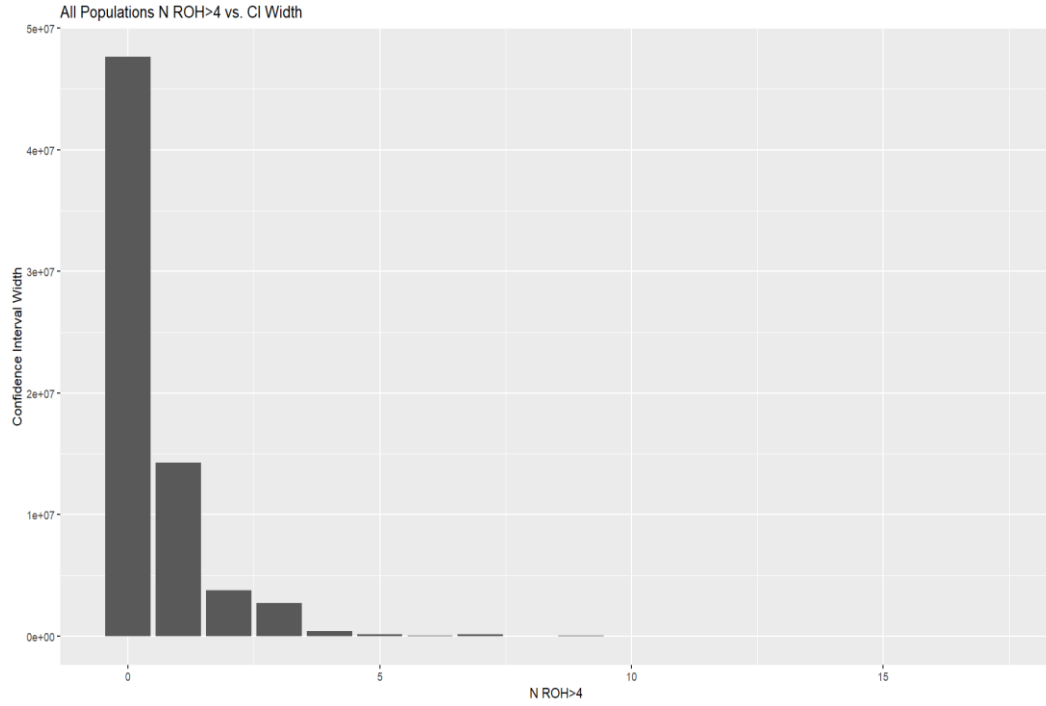


Figure 3: Number of ROH > 4 in a single grouping versus the width of the confidence interval. Results of all regions have been compiled together into a single cumulative graph as they all followed the same trend and were significant in the same manner.

As a second factor, the binning method of samples for Ne estimation was evaluated with a GLM with respect to confidence interval width, where they were compared between no binning (“0” generations, with independent, individual based Ne estimations) and binning of 2 to 5 generations (Figure 4). When generational binning included independent Ne estimations, the relationship was significant for nearly every population (all populations: $t = -7.703$, $P = 2.9e-14$, Anatolia: $t = -4.318$, $P = 2.05e-05$, Arabia: $t = -1.342$, $P = 0.212$, Caucasus: $t = -5.039$, $P = 8.1e-07$, Iran: $t = -1.475$, $P = 0.142$, Levant: $t = -4.18$, $P = 4e-05$, Mesopotamia: $t = -1.608$, $P = 0.123$), and a decrease in the confidence interval can be seen with the increase in bin size. It should also be noted that due to the lack of available data, little change occurred with the incorporation of binning for Arabia and Mesopotamia, likely resulting in the high p-values.

With the removal of the non-binned Ne category (0 generations) (Figure 5), this significance disappears entirely. There is no significant statistical difference between any generational bin over all populations ($t = -1.770$, $P = 0.0772$), nor for Anatolia ($t = -0.628$, $P = 0.531$), Arabia ($t = 0$, $P = 1$), Caucasus ($t = -1.109$, $P = 0.269$), Iran ($t = -0.327$, $P = 0.744$), Levant ($t = -1.589$, $P = 0.114$), or Mesopotamia ($t = 0$, $P = 1$). For the comparative analysis of proxies in section 7.3, a bin size of 5 generations was used. This is due to the potentiality to include more ROH within a bin due to the inclusion of more samples, which is the one feature that significantly

reduces the confidence interval, even if bin size does not directly do this.

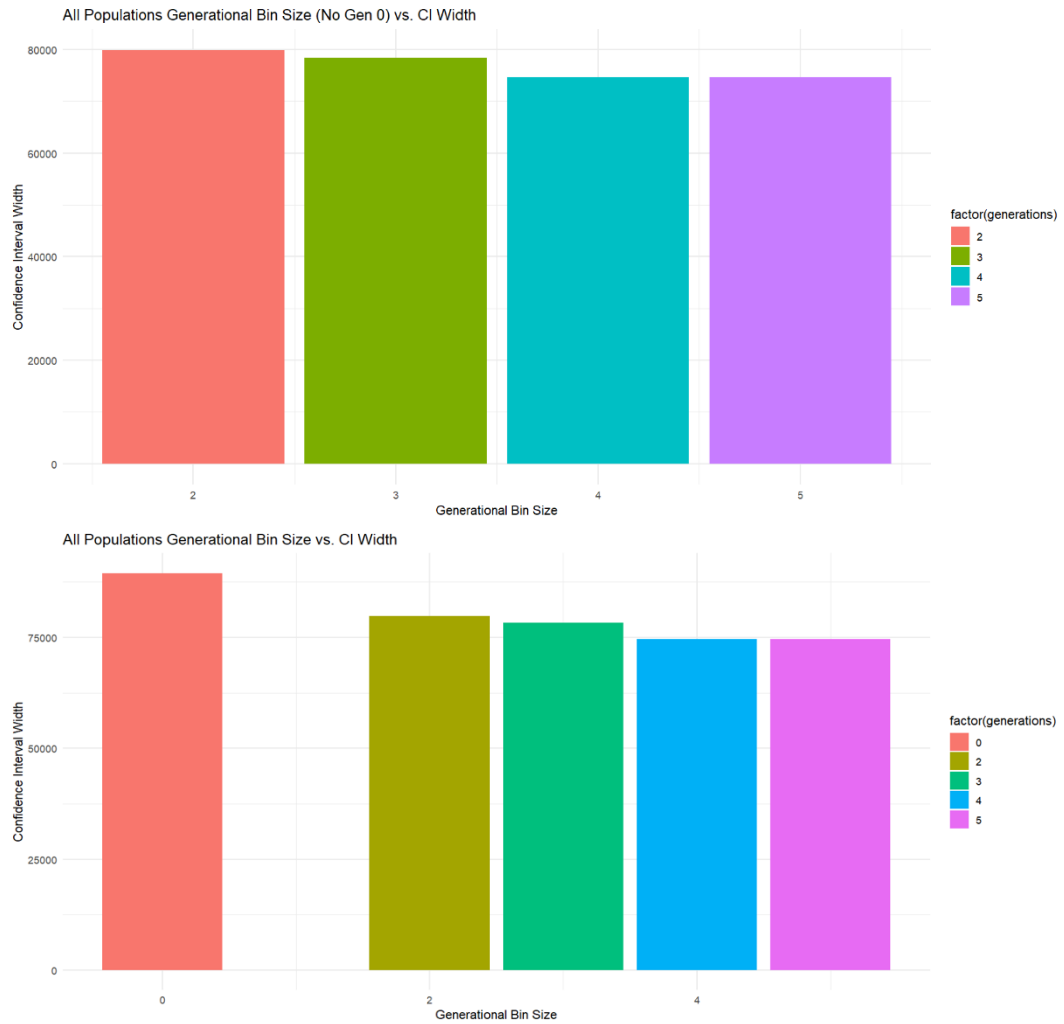


Figure 4, 5: Number of generations in a generational bin versus the width of the confidence interval. Top (4): no generations/individual Ne estimation (0) and generations 2-5. Bottom (5): generational binning 2-5 generations. Results of all regions have been compiled together into a single cumulative graph.

When evaluated with a GLM, the significance of the number of individuals in a grouping depends on the region. When all regions are considered simultaneously, the increase in individuals within a bin reduces the confidence interval ($t = -3.108$, $P = 0.002$), which can be seen visually in Figure 6. This is also the case for several independent regions, being Arabia ($t = -1.8e+15$, $P < 2e-16$), Caucasus ($t = -2.698$, $P = 0.008$), Iran ($t = -2.638$, $P = 0.009$), and Mesopotamia ($t = -1.7e+14$, $P < 2e-16$). Anatolia ($t = -1.514$, $P = 0.13$) and the Levant ($t = -0.097$, $P = 0.923$) remain insignificant - both the largest available datasets. At the larger range of sample sizes available (>27), there appears to be a general decrease in confidence size interval. In addition, for bins where the confidence interval starts high and there are few individuals, there appears to be a decreasing trend when more individuals are added

(Figure 7).

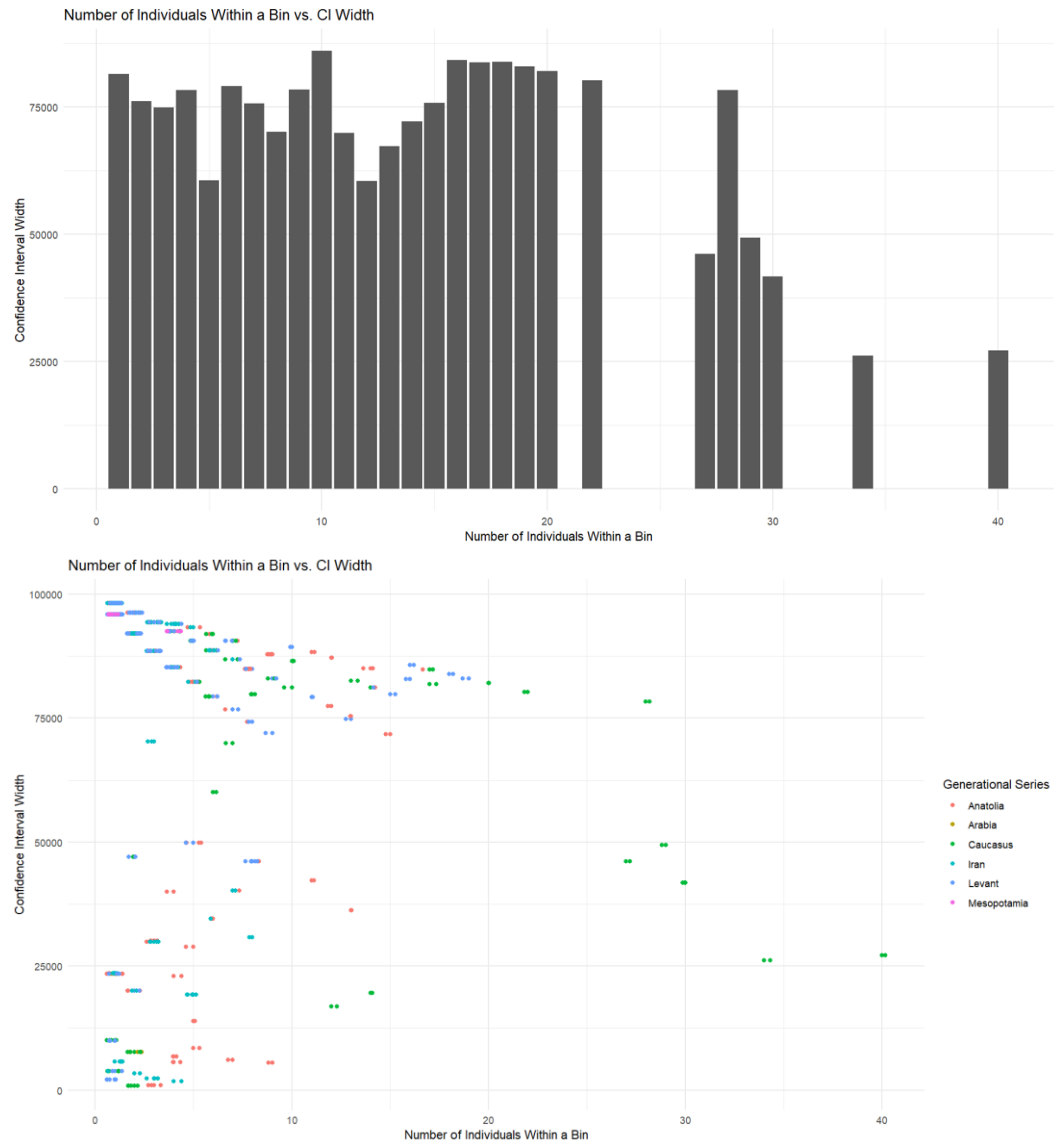


Figure 6, 7: Bar graph of the number of individuals within a bin vs. confidence interval width around the N_e estimate for said bin (top panel). Scatterplot of the number of individuals within a bin vs confidence interval width around the N_e estimate for said bin. Dots are colored according to the geographic region under study (bottom panel).

7.2 Regional ROH Distribution and ROH Time Series

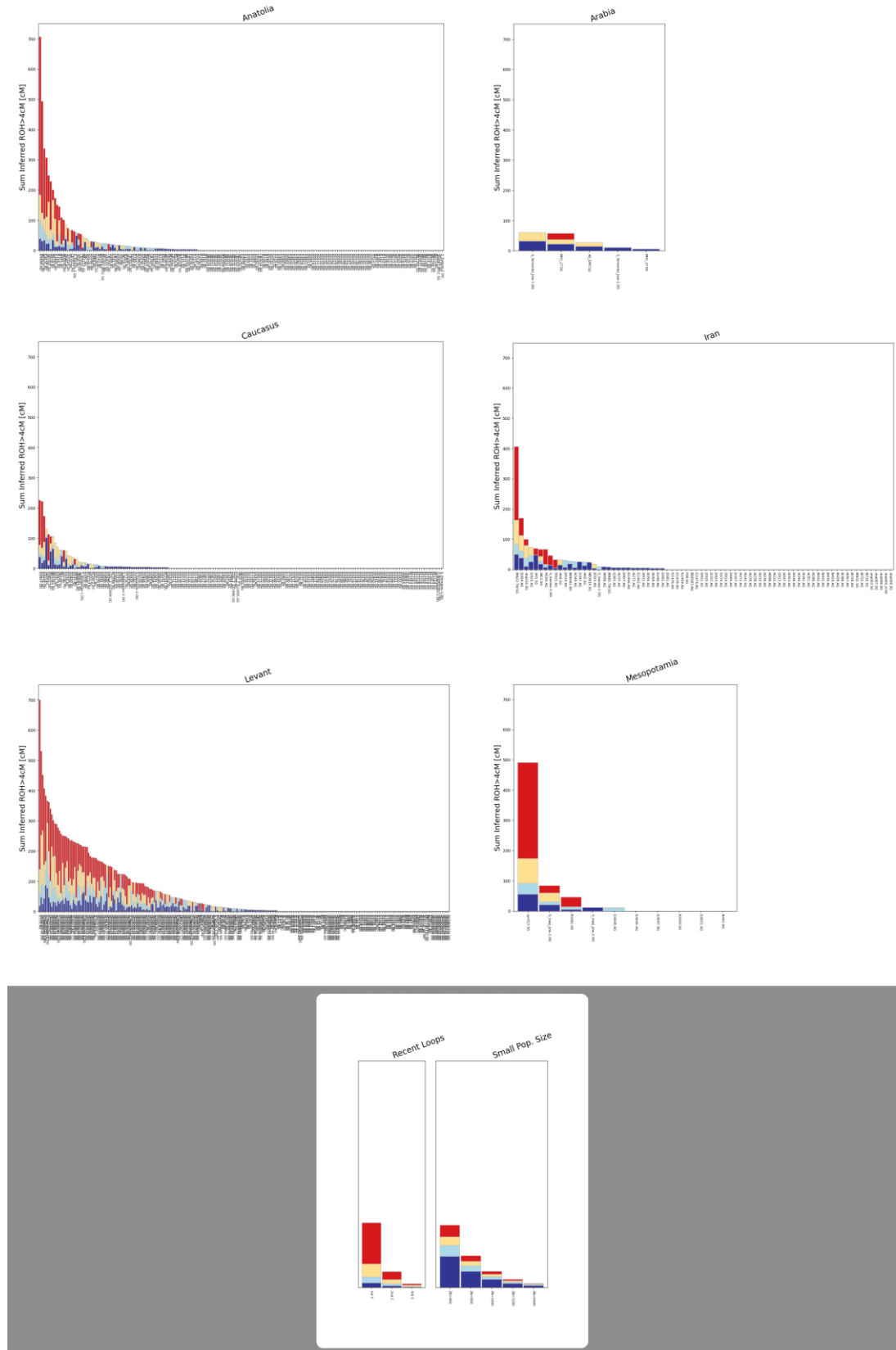


Figure 8: Distribution of the sum of ROH per individual (x-axis ticks) per population. From top left to bottom right: Anatolia, Arabia, Caucasus, Iran, Levant,

Mesopotamia. Red: ROH >20 cM, yellow: ROH >12-20 cM, light blue: ROH >8-12 cM, dark blue: ROH >4-8 cM. We also provide the expected ROH distributions for offspring of first, second, and third-degree cousins, as well as for populations of small sizes ($2N = 400, 800, 1600, 3200, 6400$) (right label).

Different regions demonstrate different compositions of ROH within their populations, as seen in Figure 8.

Anatolia's population makeup involves extreme heterogeneity. Less than half of its samples contain any ROH at all, suggesting a larger effective population through time than could possibly be estimated by this method. For those that do have ROH, most have ROHs of short lengths and of similar proportions. The sum of ROH lengths for ROH 4-8 cM and ROH 8-12 cM varies little among the individuals that have ROH, summing between 20 and 50 cM for each category. Few individuals have ROH 12-20 cM, of which the sum range is slightly larger, typically between 20 and 100 cM in length. Even fewer individuals appear to have stretches of extremely long ROH (ROH >20), of which the sum varies widely between individuals between 40 and 500 cM. This may point to recent inbreeding between more related individuals, but only for a select subset of the population. All regions but Arabia seem to show a generally deep ancestry, as only half of the samples appear to have ROH at all. On the contrary, all samples in Arabia have at least some ROH, although not particularly long ones, which follow the demographic trend of a small population size.

The Levant in particular appears to show a more gradual and homogenous distribution of ROH among all ROH lengths. Like with Anatolia, individuals show on average a higher sum of ROH lengths than in other regions. While a large proportion of individuals within this population show no ROH at all, those who do show similar ROH proportions. Opposed to few individuals showing long ROH lengths, a majority of those that have ROH have ROH of every length. The distribution of ROH is far more gradual than any other region, suggesting not only a more recent inbreeding event, but a more regional than localized one - potentially due to a small population size.

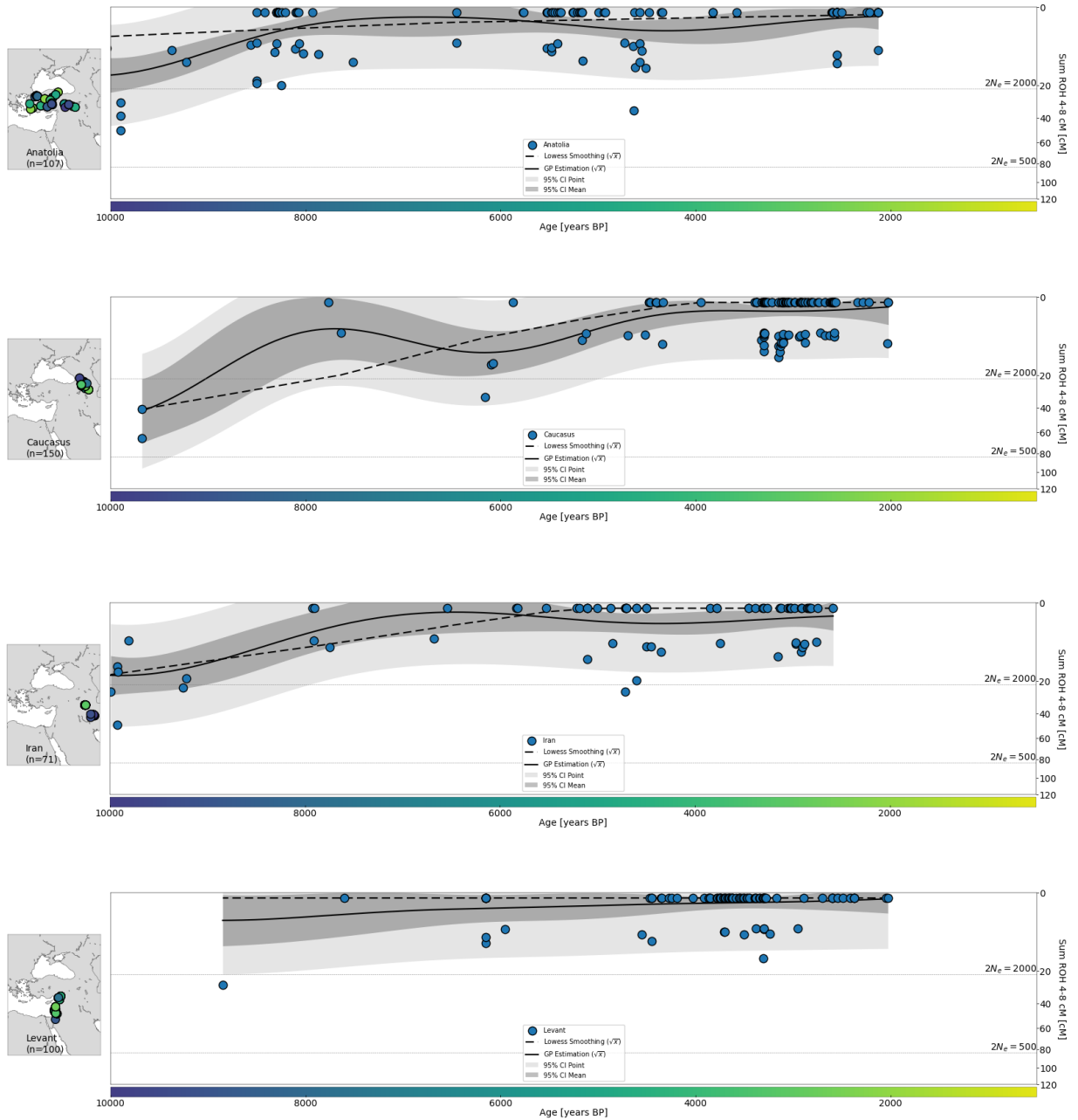


Figure 9: Sum of ROH 4-8 cM time series for Anatolia (N = 107), Caucasus (N = 150), Iran (N = 71), and Levant (N = 100). Maps of the distributions of individuals can be seen on the left side, where color corresponds to their age. The trajectory of ROH 4-8 over time can be seen, where the time series is plotted on the bottom x axis in descending order from 10,000 cal. yr. BP on the left to the present on the right. On the y axis, the sum of ROH 4-8 cM increases from top to bottom, while

the diploid effective population size ($2N_e$) decreases from top to bottom, as the two values are inversely proportional. The theoretical expectations for the sum of ROH 4-8 in the given population sizes are calculated using analytical formulas developed by Ringbauer and colleagues 2021 (Equation 2) (Ringbauer et al 2021, Supplementary Note 4). Single dots represent a single individual, plotted by their age and their sum of ROH 4-8. The dotted (Lowess Smoothing) and solid (GP Estimation) smoothed lines plot the mean estimates calculated by Lowess smoothing and a Gaussian Process model, respectively, that use root-transformed ROH values. Lines are illustrated with 95% empirical confidence intervals for individuals (light gray) and for the estimated mean (dark grey).

It is possible to observe changes in population size depending on the geographic area considered in Figure 9, where the sum of ROH are mapped to the expected population size (Equation 2). When observing the patterns given by the Lowess-smoothed trend (dotted line), the effective population size of Anatolia, Caucasus, and Iran increases from 10000 cal. yr. BP to the present gradually, while the effective population size of the Levant always remains high through this time span. When observing the GP trend (solid line), the trend shows more dynamic decreases and increases. Anatolia, Caucasus, and Iran show similar demographic patterns, where the population begins low and increases to a peak at around 8000 cal. yr. BP. The population then decreases gradually between 8000 and 4000 cal. yr. BP before increasing once more towards the present as the sum of ROH 4-8 becomes smaller. On the contrary, the population of the Levant throughout the entire time series appears to be high and relatively unwavering. The population begins quite high, where $2N_e = 2000$ ($N_e = 14000$), and increases gradually towards the present where the sum of ROH 4-8 becomes smaller.

$$f_N(x)dx = \frac{4N(1 + 2NG)}{(1 + 2Nx)^3} dx$$

Equation 2: Expected density of ROH blocks of length x in a constant panmictic population. $N = 2N_e$, which is divided by 2 for the number of haploid genomes (twice the effective number of diploid individuals). X is ROH length and G is the genetic length of the chromosome, both measured in Morgans.

7.3 Analysis of the N_e Trend and Comparative Analysis of Proxies

We further compare the demographic trends for Anatolia, Caucasus, Iran, and Levant using the genetic effective population size, settlement survey data, and radiocarbon data proxies to assess their relative agreement along the same time series. All three proxies were binned into the same 150-year time bins (five generations) from 10000 to 1500 cal. yr. BP (for the reasons illustrated in section 7.1) and were normalized from 0 to 1 to make them comparable on the same scale

In addition, seen in the comparative plots in Figure 10, the actual period of ancestral demographic estimation for ROH 4-8, which correspond between 6 and 12.5 generations (~180-360 years), was accounted for by shifting the Ne trend line back in time by two time bins (300 years) (Howringer et al 2011). Trend lines were smoothed using the Lowess method to account for noise and fluctuations and to create a smooth trend line with continuous values, which were then compared (Figure 11). The trends of demographic proxies were evaluated by calculating the difference between the current bin and the previous one to understand the trajectory of the curve. At each time step, the direction of each proxy was then compared. The relationship was assigned as 1 when in complete agreement (both positive or negative), 0 when neutral (no change), -1 when in a weak disagreement (one positive or negative, one neutral), and -2 when in complete disagreement (one positive, one negative). Fluctuations smaller than 0.005 (0.5%) in either direction were not interpreted as directional demographic change, creating a 1% tolerance band around 0. In this case, bins were considered as static at the given time point.

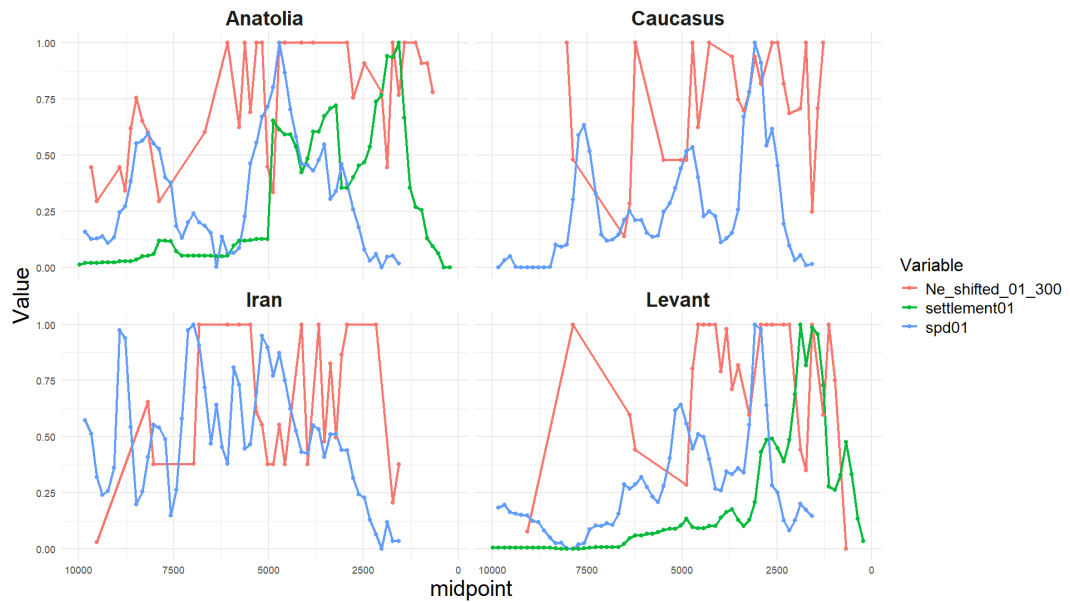


Figure 10: Anatolia, Caucasus, Iran and Levant time series trends with their corresponding scaled Ne (Effective Population Size (Ne)), Settlement (Settlement) and radiocarbon (Radiocarbon (SPD)) demographic proxies.

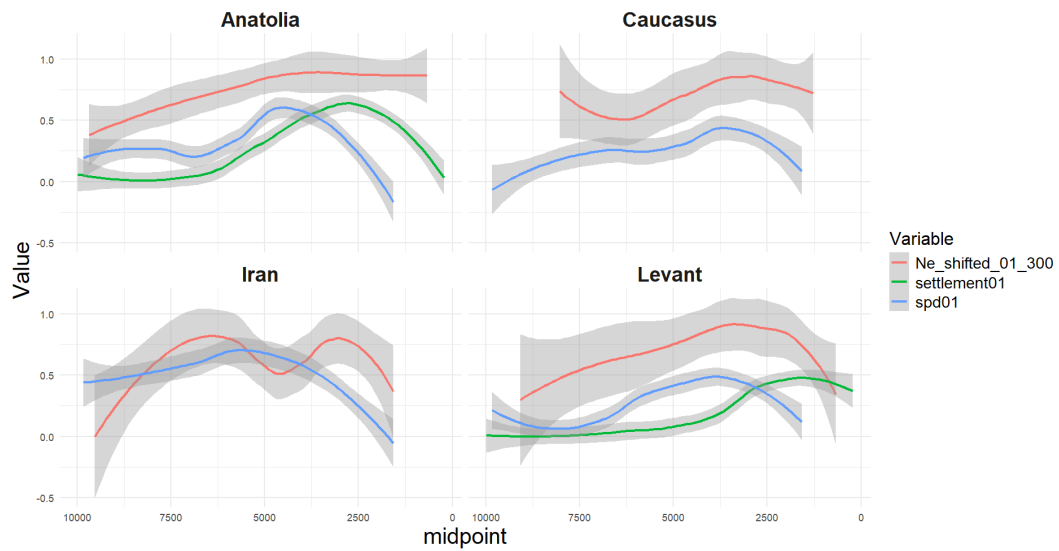


Figure 11: Anatolia, Caucasus, Iran and Levant Loess smoothed time series trends with their corresponding scaled Ne (red; Effective population size (Ne)), radiocarbon (green; Radiocarbon (SPD)), and settlement (blue; Settlement) demographic proxies. Grey areas represent the confidence interval at 95%.

Figure 10 shows the raw plotted data, while Figure 11 shows the data with smoothing applied. Across regions, a similar trend in demographic change can be seen for nearly all proxies from 10000 towards 0 cal. yr. BP. Initially, population size appears to increase in the first half of the time series, from about 10000 to 6000 cal. yr. BP. In general, we see an increase in population size in the first half of the series, followed by a decline in the second half, with trends being similar between geographic areas and proxies despite slight chronological shifting. The long-term agreements of proxies are further evaluated on a population basis in the following sections.

7.31 Anatolia

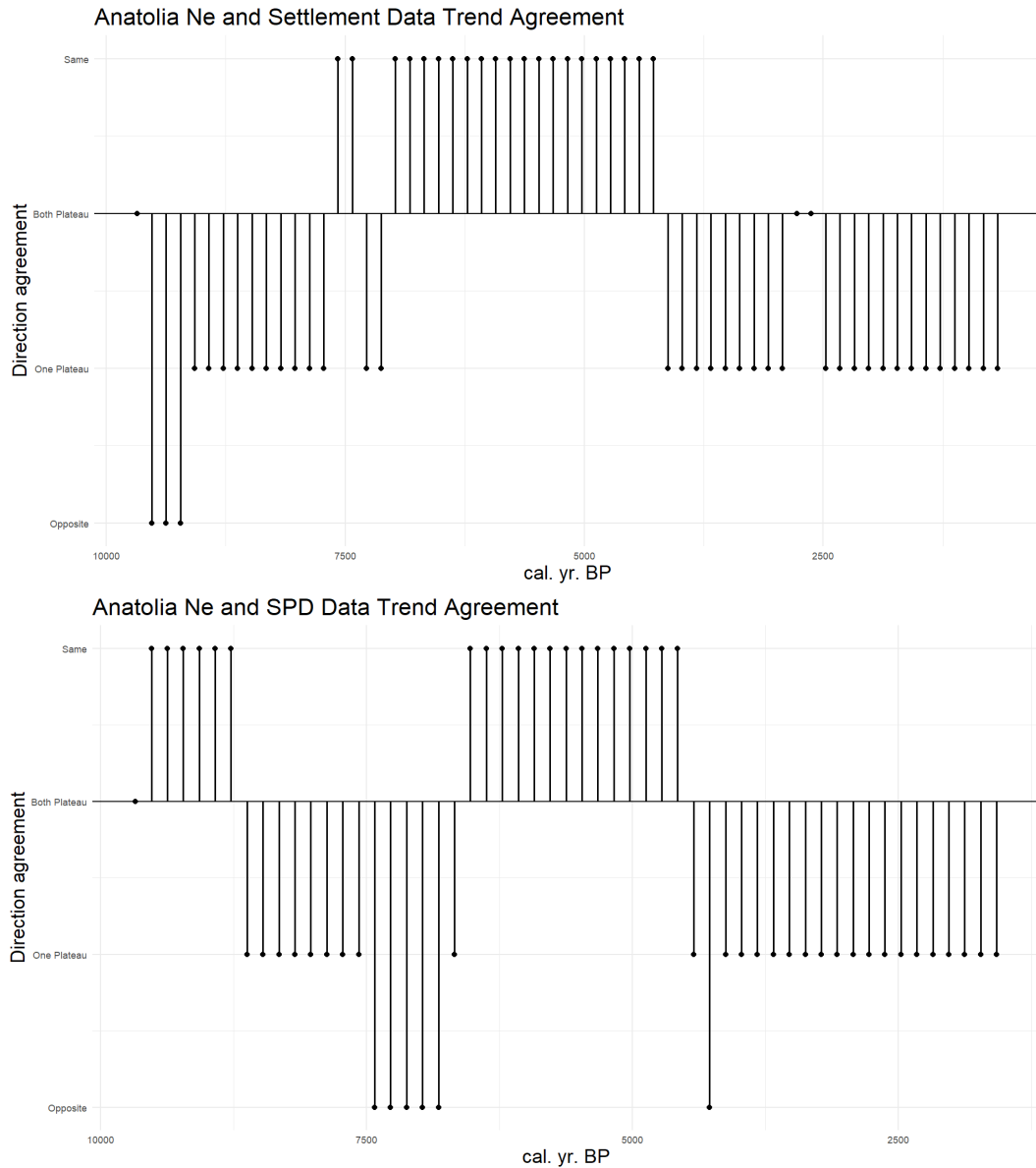


Figure 12 and 13: Anatolia direction agreement graphs between demographic proxies using genetic data (Ne) and radiocarbon data (SPD) (top) and settlement data (bottom).

Anatolia Ne estimations illustrate an increase in population size throughout nearly the entire time series from 10000 to 3250 cal. yr. BP (0.45 to 0.85), before a minor population decrease and plateau from 3250 cal. yr. BP towards the present. When this trend is compared to radiocarbon and settlement data proxies, it agrees broadly in the same way to radiocarbon data as it does to the settlement data, as do SPD and settlement data to one another (Figure 14). All three methods agree to a positive increase in population size from slightly after 7000 cal. yr. BP to slightly after 5000 cal. yr. BP. However, while Ne estimations continue to suggest a growing population after this initial increase, both radiocarbon and archaeological data

appear to show a decrease in population size from 4000 cal. yr. BP towards the present.

Ne and SPD trends appear to agree regarding an initial population increase from 10000 cal. yr. BP towards 8250 cal. yr BP (Figure 12). The agreement between these proxies then becomes slightly less strong as Ne continues to increase while SPD maintains a plateau from 8625 to 6675 cal. yr. BP, with a notable population decrease between 7425 and 6825 cal. yr BP. Settlement data appears to show this same plateau or slight decrease in population from 10000 to around 7000 cal. yr. BP, while also not suggesting the same initial increase in population size as radiocarbon data does (Figure 13).

Despite being independent proxies, both SPD and settlement data show a remarkably similar agreement to Ne. While a weak disagreement between Ne and the respective proxy is seen about half of the time (51% for SPD and 55% for settlement data), this predominantly comes from one proxy trending in a decisive direction (whether positive or negative) while the other plateaus. The proxies are only in strong disagreement and trend in different directions between 5 (SPD) and 11 (settlement) percent of the time, while they are in complete agreement over the direction of population change 36 (SPD) and 34 (settlement) percent of the time.

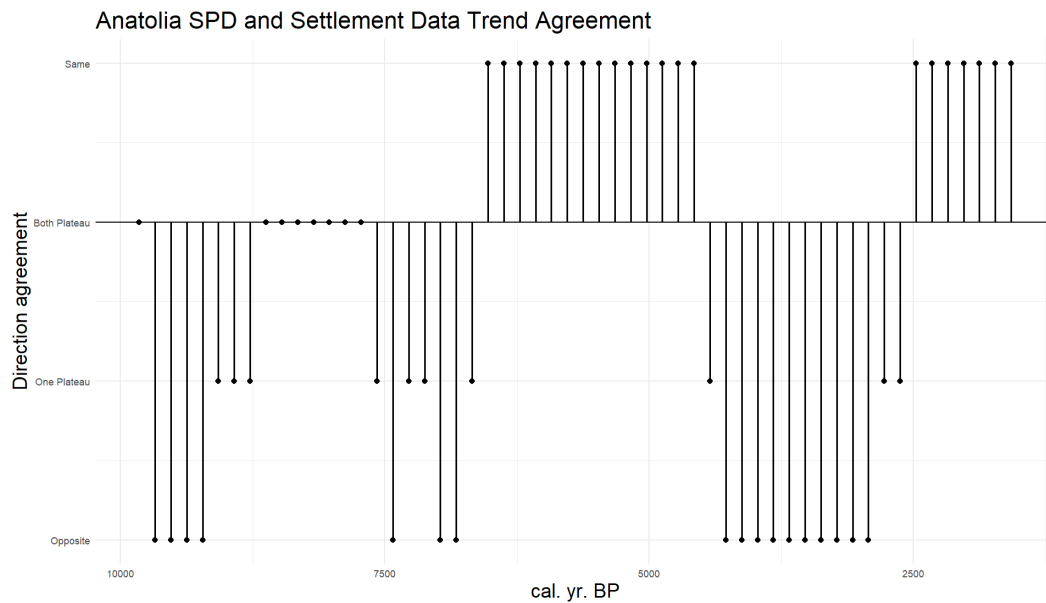


Figure 14: Anatolia direction agreement graphs between demographic proxies using radiocarbon data (SPD) and settlement data.

5.32 Caucasus

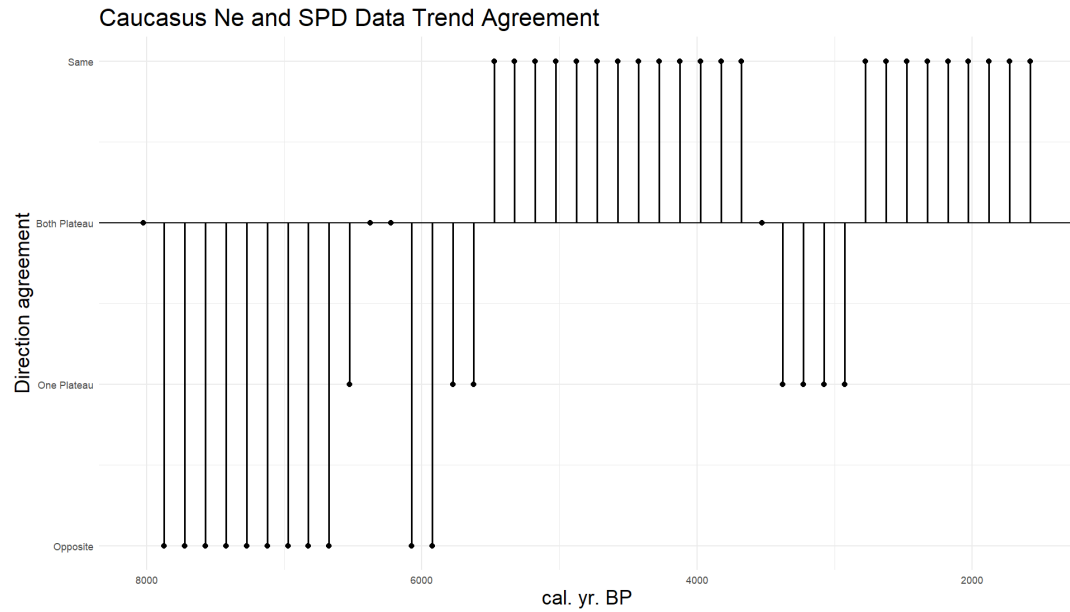


Figure 15: Caucasus direction agreement graph between demographic proxies using genetic data (Ne) and radiocarbon data (SPD).

Caucasus Ne estimations illustrate a decrease in population size between 7800 and 6750 cal. yr. BP (0.75 to 0.5), before a steady increase in population until 2500 cal. yr. BP (0.8). An additional slight decline in population occurs afterwards from 6750 towards 1250 cal. yr. BP (0.8 to 0.74). This matches almost identically to the radiocarbon trend. The SPD trend suggests a slow increase in population size from 10000 to 7000 cal. yr. BP (0 to 0.25) before a slight decrease or plateau in population size from 6250 to approximately 5000 cal. yr. BP. An increase in population size then occurs at approximately the same time as Ne, from 5000 to 3250 cal. yr. BP (0.25 to 0.45), before a similar although somewhat stronger population decline towards 1250 cal. yr. BP (0.45 to 0.12). Between 6250 and 1250 cal. yr. BP, the trends are nearly identical, with an increase in population, slight peak, and then a plateau towards population decrease.

The trend between the two proxies seen in Figure 15 is anticorrelated at the start of the time series and then becomes predominantly in agreement in the later half. This initial anticorrelation spans initially from 8000 cal. yr. BP to 6525 cal. yr. BP, where the SPD trend appears to increase while the Ne trend appears to decrease. It should be noted, however, that the SPD trend at 8000 cal. yr. BP is in part affected by its previous data points in a smoothed model, which began at 10000 cal. yr. BP, while the Ne trend begins at 7725 cal. yr. BP. In addition, there are only two samples until 6075 cal. yr. BP, leading to overall low confidence in the estimation. Indeed, shortly after this period at 5475 cal. yr. BP, the two proxy trends are almost entirely in agreement both as they increase and then as they decline towards zero. Overall, the

two trends agree half the time (50%), with a weak disagreement 16% of the time and a strong disagreement 25% of the time.

7.33 Iran

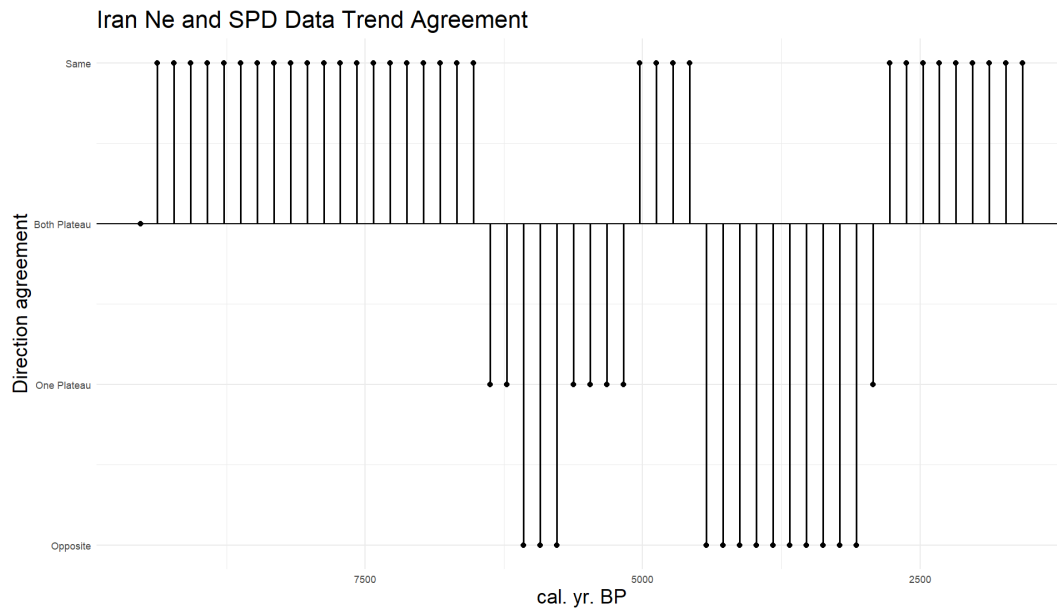


Figure 16: Iran direction agreement graph between demographic proxies using genetic data (Ne) and radiocarbon data (SPD).

Iran Ne estimations illustrate an increase in population size between 10000 and 6500 cal. yr. BP (0 to 0.8) before a sharp trough in population just after 5000 cal. yr. BP (0.5). As the time series trends towards 2800 cal. yr. BP, the population increases once more, peaking at 2800 cal. yr. BP (0.75), before declining towards the present (0.37). The initial peak in population size for the Ne proxy demonstrates at 6500 cal. yr. BP occurs at nearly the same time as the peak in the radiocarbon trend at 6500 cal. yr. BP. Indeed, the trends of both proxies appear similar from 10000 to 4800 cal. yr. BP. The SPD trend illustrates a slow increase in population size from 10000 to 7000 cal. yr. BP (0.45 to 0.7) before a slight decrease or plateau in population size from 6250 to approximately 5000 cal. yr. BP. (0.7 to 0.6). This trend occurs around the same time as Ne, as population increases from 10000 cal. yr. BP to 6500 cal. yr. BP (0 to 0.8) before initially declining until 5000 cal. yr. BP (0.5). These trends, however, begin to differentiate after 5000 cal. yr. BP. As the Ne trend begins to increase again until 2800 cal. yr. BP (0.7), the SPD trend continues to decline towards zero, only matching the Ne trend again after Ne begins its final decline after 2800 cal. yr. BP.

This relationship can be seen in the agreement graph in Figure 16. Both trends are in complete agreement until 6525 cal. yr. BP. They then become predominantly weakly or strongly anticorrelated from 6525 until 2925 cal. yr. BP, with a small

brief agreement between 5025 and 4500 cal. yr. BP during a brief period of decline of both trends. The trends then agree again for the remainder of the time series after 4500 cal. yr. BP where both trends decrease. Overall, both trends agree in trend direction 61% of the time. A weak disagreement occurs 13% of the time, while a strong disagreement occurs 25% of the time.

7.34 *Levant*

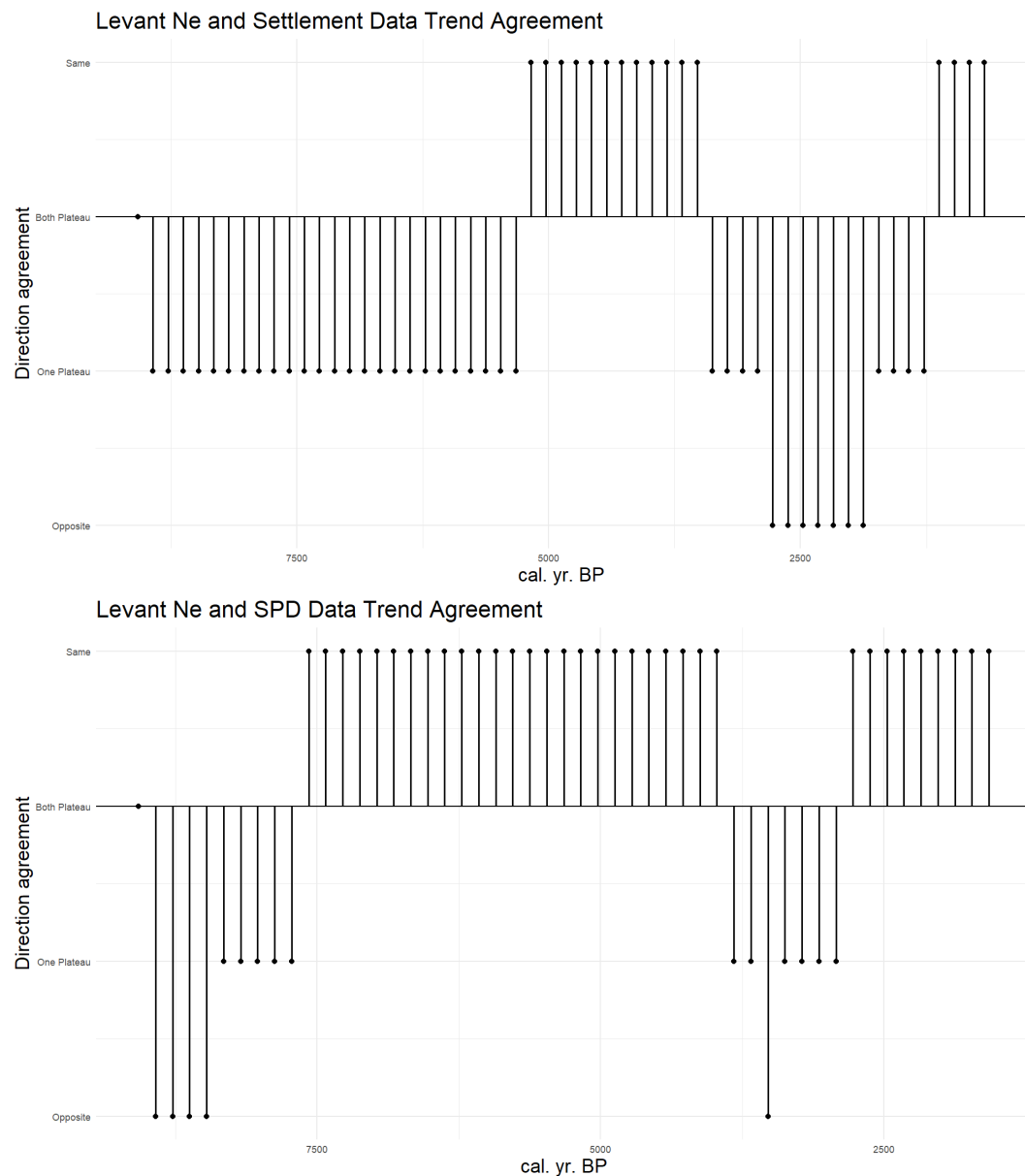


Figure 17 and 18: Levant direction agreement graphs between demographic proxies using genetic data (Ne) and radiocarbon data (SPD) (left) and settlement data (right).

Levant Ne estimations illustrate an oscillating pattern of trend disagreement at the beginning of the time series, then a stretch of agreement, then a second set of disagreement, ending with a final stretch of agreement. The Ne time series trend

suggests a steady increase in population size from 10000 to 3250 cal. yr. BP (0.27 to 0.8), a slow decline or plateau from 3250 to 2000 cal. yr. BP (0.8 to 0.76) until a sharp decline from 2000 cal. yr. BP to the present (0.76 to 0.3). When this trend is compared to radiocarbon and settlement data proxies, it agrees broadly in the same way to radiocarbon data as it does to the settlement data, where the settlement data relationship to Ne resembles that of the SPD trend shifted forward in time. Indeed, both the SPD and settlement trends increase in population size slowly from a plateau before peaking broadly, followed by a decrease. When SPD and settlement data trends are compared to one another, the agreement is most similar to that of the Ne and settlement data, particularly pertaining to the weak disagreement in the first half of the trend, agreement in the middle, and strong disagreement for the remaining portion of the trend (Figure 19).

Ne and radiocarbon data agree approximately 67% of the time, with a weak disagreement 21% of the time and a strong disagreement 10% of the time (Figure 17). The relationship between the two proxies begins in strong disagreement from 10000 to 8400 cal. yr. BP and then shifts to a weak disagreement from 8400 to 7725 cal. yr. BP as the SPD trend shifts from decreasing to a plateau. The two trends then correlate until 3975 cal. yr. BP as they both increase before shifting to a weak disagreement until 2925 cal. yr. BP as the SPD trend begins to decrease while the Ne trend remains increasing until plateau. Both trends then end in agreement as they decrease towards the present.

Ne and settlement data agree approximately 28% of the time, with a weak disagreement 58% of the time and a strong disagreement 12% of the time (Figure 18). Ne and settlement trends initially are in weak disagreement as the settlement data remains predominantly plateaued until 5175 cal. yr. BP, from where the relationship shifts to agreement as both trends increase until 3525 cal. yr. BP. The relationship then shifts again from a weak disagreement to a strong disagreement from 3525 to 1275 cal. yr. BP as the Ne trend begins to decline while the settlement trend continues to increase until plateau. They then shift again into agreement as both decline until the present.

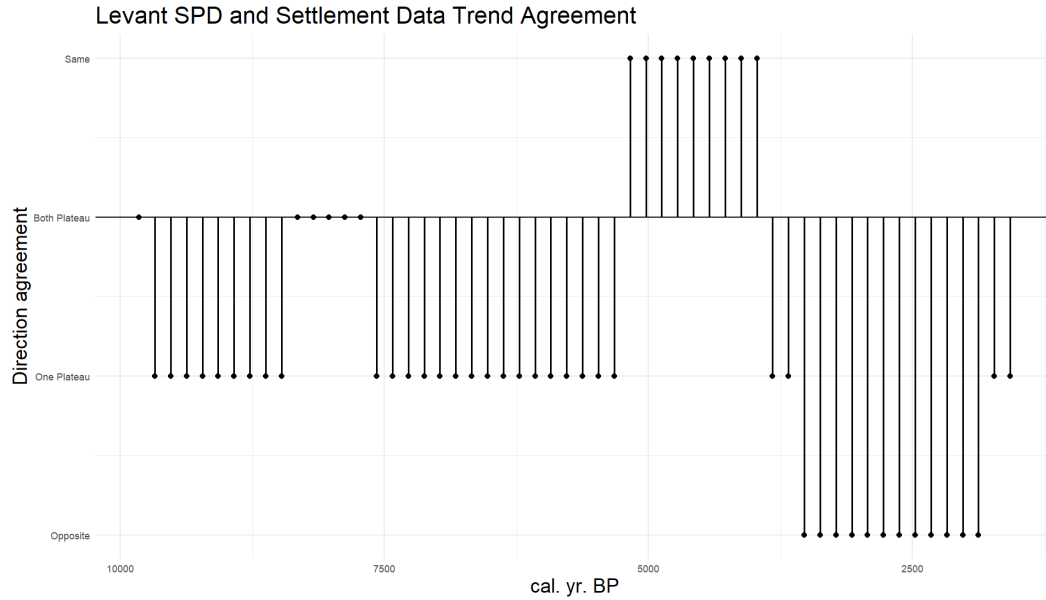


Figure 19: Levant direction agreement graphs between demographic proxies using radiocarbon data (SPD) and settlement data.

8. Discussion

We have explored hapROH as a method for estimating genetic effective population size as well as how it agrees with other proxies for population size. Across the Near East between 10000 and 1500 years before present, all proxy trends show a relatively high degree of agreement. In periods where they do not agree, we explore further as to why this may be the case, and which method may be more reliable at a given time. We now discuss them below.

8.1 Binning and Confidence Interval Reduction Exploration

Generalized linear models suggest that only the number of ROH within a binned sample is indicative of a decrease in the confidence interval. It could be believed that, with the increase in individuals with more ROHs in a bin, the confidence interval would also decrease, but this is in fact not statistically significant - nor is the temporal size of the bin. In this case, this may be in part due to the proportion of samples with no ROH $> 4\text{cM}$ within a population sample. No-ROH samples make up over half of the sample size of most regions. As these samples indicate a large population size with more distant ancestry, they are given the maximum N_e estimation of 50,000 for hapROH with a confidence interval width of 100,000. This is a severe limitation of hapROH, as it infers that these ancestral populations are large, but is unable to assess exactly how large.

As the proportion of individuals with 0 ROH $> 4\text{cM}$ within a population increases, so does the width of the confidence interval ($p\text{-value} = 8.82\text{e-}05$) (Figure 20). As binning occurs, it may occur that an individual with no ROH will be added to the

bin, increasing the confidence interval size and counteracting a reduction that may occur by adding a sample with many ROH. This is a crucial consideration in the case that the confidence interval is used as an indicator of N_e estimation accuracy. Ideally, investigation into the feature would need to be done, especially in tandem with other features of the dataset such as group or bin size. As it stands, with small N_e , we can investigate changes in N_e more precisely as the confidence intervals around these estimates are smaller – plus, they can remain or become smaller with the inclusion of more individuals into a group. With larger N_e 's, even if accurately inferred, it is difficult to say exactly how large the population size is – hapROH simply caps it at 50,000. In this case, it may be more useful to rely on a qualitative or comparative manner that N_e is larger, rather than investigate specific quantitative changes.

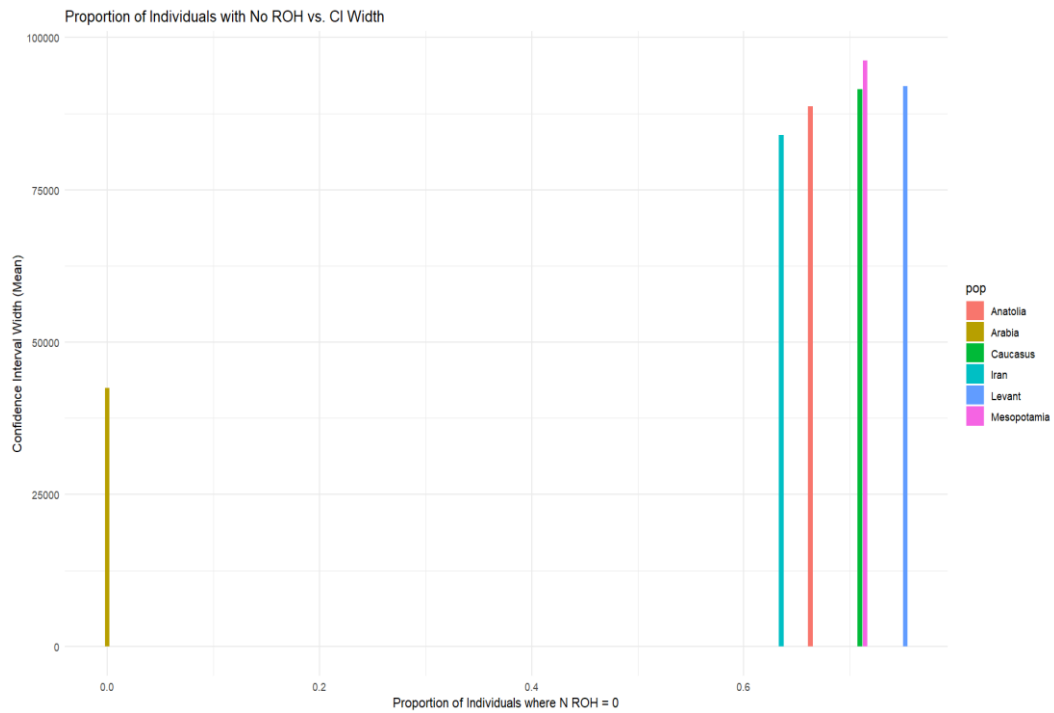


Figure 20: Proportion of individuals with no ROH in a population dataset and average confidence interval width for 0 generations.

The increase in generational bins size does not guarantee that more individuals will be grouped into that bin, while the inclusion of more individuals within a group only becomes meaningful if they have ROH 4-8 cM to contribute. The reduction of the confidence interval likely depends on the quality and ability of the individuals within these groupings themselves to contribute meaningful N_e estimates, predominantly for populations under 50,000. For populations of larger size, specifically over 50,000, the use of ROH-based N_e as a reliable proxy of population change should be carefully assessed, especially in the case of low data quality or data availability as they may make the detection of ROH more difficult.

8.2 Analysis of the N_e Trend and Comparative Analysis of Proxies

According to the N_e estimation, the effective population size appears to increase steadily for most populations for most of their time series, beginning from 10000 cal. yr. BP. Despite some small discrepancies, they generally follow the same demographic trend, increasing in the beginning of the time series, peaking in the middle, and then decreasing towards the end. Below, we explore these relationships in detail and provide additional historical and archaeological context that may explain these trends and relationships.

Anatolia

Anatolia's archaeological settlement and radiocarbon data show the same population trend, where they increase slowly in the first half, peak between 5000 and 3000 cal. yr. BP, and then decline towards the present. However, they are shifted slightly, where the SPD peaks about 1500 years before the settlement proxy. This general agreement is not unusual, as several studies have shown broad agreement in demographic trends by the SPD and other archaeological indices (Palmisano et al, 2017; French and Collins, 2015). The lag/shift between the two proxies, on the other hand, might be due to a temporal offset as opposed to a real difference. An increase in census size may be seen first in the increased use of organic material, which is a more immediate manifestation of population change. Settlement data, however, is composed of estimated site counts and size. This would require more time to manifest in the archaeological record, as it physically requires more time to expand.

The genetic proxy of N_e agrees with both the settlement and radiocarbon data proxies until approximately 4000 cal. yr. BP, where it continues increasing while the others begin to decrease. It is hypothesized by Palmisano and colleagues (2021) that the decrease in the SPD and settlement proxies at this time is broadly correlated to a 4200 cal. yr. BP event of climate drying. However, this event and its impact is under particular scrutiny, with some arguing that it was responsible for massive population decline, while others are less convinced. Moreover, there is significant variation across the Near East to the degree that this is manifested (Palmisano et al, 2021). The population doesn't increase nor decrease after this point, hypothesizing that if this region was indeed impacted by this climatic event, it may have been coped with through technological advancements and adaptive strategies such as drought-tolerant crop species, switching to different resources, or food preservation techniques (Palmisano et al, 2021).

It is also important to consider how the radiocarbon data is interpreted in this context. As radiocarbon data is composed of organic material of several origins including that of seeds and bones, a change in the SPD may indicate a change in production, subsistence and material use, whether due to climatic changes or otherwise. In addition, climatic change or site disturbance may have long term

effects on site preservation, resulting in less usable data. The interpretation of radiocarbon data and its use as a population proxy has been debated since at least 2017, where its proposed use should be as an estimate for energy consumption as opposed to a direct proportional proxy for population size (Freeman et al, 2017). Thus, while it may be hypothesized that this drying phase impacted the environment and organic material, its human population size may be less certain, especially considering the steady increase in Ne during that time.

Caucasus

Caucasus Ne estimations match almost identically to the radiocarbon trend, both broadly corresponding to known historical events. The overall increasing trend in population size correlates well with the onset of agricultural practices, while a steeper increase in population size at 3500 cal. yr. BP, is a known period characterized by increased social complexity and agricultural production thanks to technological advancements such as irrigation (Palmisano et al, 2021). Much like Anatolia, there also appears to be a decrease in population size after 4200 cal. yr. BP, potentially reflecting the same climatic drying shift also in this region.

These two methods do, however, differ at the onset of Ne at 8000 cal. yr. BP, where Ne initially decreases while SPD increases. This initial high estimation in Ne, however, is supported by only one individual, with no other data available until after 7000 cal. yr. BP. Regardless, this decrease in population in Ne may still be loosely correlated with the 8200 cal. yr. BP event described by Palmisano and colleagues 2021, characterized by a decrease in temperatures and increase in snow cover. This appears little supported by the SPD proxy, however, which increases steadily during this time. In addition, no genetic data is so far available for the Caucasus between 10000 and 8000 cal. yr. BP, making it impossible to compare the trends at this point. The inclusion of additional genetic data before 8000 cal. yr. BP, if available, may help in further understanding the agreement between the two proxies further.

Iran

Among any of the regions, the relationship between the genetic and radiocarbon proxies of Iran is the most in disagreement. While both Ne and SPD curves show an initial increase in population size in the beginning until about 6250 cal. yr. BP, as well as a decrease at the very end after 3000 cal. yr. BP, they are predominantly in disagreement between these two points. After the initial peak in population at 6000 cal. yr. BP, the SPD trend decreases for the rest of the time series. However, while the Ne trend also initially begins to decrease at around the same time or slightly before, it only does so until 4600 cal. yr. BP, from which it then increases once more, peaking at 3000 cal. yr. BP. Therefore, while SPD has one peak in population, Ne has two.

The initial decline in population size at 6000 cal. yr. BP suggested by both proxies is affirmed in archaeological literature. Site prevalence and size in the Qazvin plain, for example, is known to have decreased significantly during this time, where a shift is noted from proto-urban sites of the fourth millennium to a more small-scale, village agricultural lifestyle during the third millennium BCE (Nashil et al, 2022). This event does not appear to correspond well to any broad climatic events, falling between the 8200 and 4200 cal. yr. BP climatic events, although Palmisano and colleagues attribute population booms and busts in this region to wet and dry climatic oscillations during the same time. In addition, population appears to increase for Ne when the 4200 cal. yr. BP event should occur, potentially hypothesizing the reduced impact of this event on the region due to coping through technological advancements.

Despite the lack of evidence of climatic improvement, the population growth in Ne after 4600 cal. yr BP is also supported by archaeological records, such as the monumental expansion of the Iranian city Shahr-i Sokhta of south-east Iran. Between 2800-2500 BCE (4750–4450 BP), this city expanded to its largest size of 80 hectares from its original founding 16 hectares between 3200-2800 BCE (5150-4750 BP) before undergoing a general collapse until 2450 BCE (4400 BP) (Talebian, 2014). At its peak, it is estimated that this city had between 5000 and 10000 inhabitants (Ascanlone et al, 2017).

As we only have two proxies to compare for this region, it is important to reiterate that the radiocarbon data provided is only effective until 3000/2500 cal. yr. BP, as after this point, the predominant population size proxy becomes archaeological data (Palmisano et al, 2021). Considering this, it is possible that this population increase seen in Ne may be a real indication of population recovery, at least in some regions in Iran, given the archaeological and historical support illustrated.

Levant

In the Levant, all three population proxies share a relatively similar demographic trajectory, although interestingly, all are temporally shifted to some capacity and do not line up one-to-one. In the first half of the time series, the proxies show a slow increase, peaking in the middle at some time between 3750 and 1250 cal. yr. BP, before declining towards the present. The brief initial period of population decline in the SPD trend is accounted for by an initial population bulge before 9500 cal. yr. BP, then a decline, potentially correlating with overall dryer conditions (Palmisano et al, 2021). The population decline that begins post-5000 cal. yr. BP for both Ne and SPD may also be attributed to climatic changes, in this case due to more pronounced dry-wet cycles superimposed onto a larger drying trend during this time (Palmisano et al, 2021). However, settlement data continues to grow through this period, only beginning to decline extremely close to the present. Palmisano and colleagues do mention an additional demographic peak during the Iron Age, which

may be reflected in the settlement data and to some degree N_e , while not in the SPD due to lack of effectiveness of this data after 3000 cal. yr. BP.

As stated, the three population proxies are similar in demographic trajectory, but are shifted slightly. A similar pattern can be seen in the proxies of Anatolia – where the SPD trend peaks before the settlement trend. This may be due to the same reason, where the increase in population has a more immediate impact on organic materials and thus manifests in the archaeological record before the increase in settlement sizes and quantity. The N_e trend, however, peaks more-so between the SPD and settlement proxies and has a generally smoother appearance. This may be due to this variable changing on a more moderate time span than the other two – as SPD changes occur more immediately, and settlement changes may take generations, a single human generation is about thirty years. Thus, while this change in N_e is seen quickly, it is not immediately represented in the record.

8.3 Limitations, Future Directions, and Conclusions

The use of ROHs has the remarkable ability to estimate the size of a population generations before the life of the individual sampled, allowing us to peek into the population demographics of the past for time periods of which we may otherwise have limited data. Ideally, the use of longer ROH lengths, such as those between 8-12 and 12-20 cMs, could greatly improve the confidence of the demographic trends throughout time. This would extend the reach of a single individual's data by producing more than one N_e estimation at different points in the time series, allowing for the estimations of different ROH lengths from different individuals to overlap and support the trend. In addition, as the length of ROH gives inference to “N” generations into the past, it is more informative for deeper ancestry, and it may be more difficult to infer more recent N_e as we get closer to the present, especially if there are few ROH of longer lengths. As a result, the inclusion of modern genetics should also be considered when the time series in question nears towards the present.

It should be noted, however, that such a proxy is not without flaws or limitations. The availability of genetic data is directly dependent on the human remains available for each region, our ability to sequence them, and if enough was able to be sequenced for analysis. In addition, many of the limitations pertaining to archaeological proxies, such as academic bias or lack of funding, apply in much the same way to genetic data. Head counts and how they relate to geographic areas should be considered as well in the final interpretation. In this analysis, samples from a broad area were used and placed under large regional labels. However, as these are only samples from an area, they may not reflect the entire region in question, and smaller areas within a region may have a more complex history than the averaging taking place.

The time series of the genetic data also directly depends on the initial date estimation of the individuals sampled. If the estimated date of the individual has a large standard deviation, then the date is less precisely estimated, which may in turn impact the overall estimation of the N_e when binning multiple individuals together. The N_e estimation is further dependent on the quality and quantity of the genetic data, both from how many samples are within a region, and the quality of the samples themselves. Genetic data degrades exponentially, resulting in highly fragmented and chemically modified DNA molecules that are also at risk of contamination. This is especially the case under poor conservation conditions, which could result in individuals of poorer or less reliable quality (Bonfigli et al, 2023). Preservation is a particularly concerning issue with the data presented in this thesis, as it is known that DNA preservation in the regions studied here is typically poor, further restricting the few data available. With few individuals of poor quality, the N_e estimate may not be well supported. In addition, individuals with no ROH at all, while biologically relevant, provide no ability to truly estimate effective population size and result in massive N_e estimations with massive confidence intervals. When included in binning, this could potentially counteract the decrease in the confidence interval that is the result of including individuals with large amounts of ROH, resulting in a less certain estimation despite the inclusion of additional data. It is therefore important to further evaluate the reliability or even relevancy of N_e estimations through hapROH for assumed large population sizes.

Binning as a method also has its flaws. While it can be useful to gain a more confident N_e estimation over a determined time series, this averaging may not reflect a true biological reality when taking into consideration the generational range of ROH lengths. The length of ROHs and the N_e estimation garnered from them correspond to an estimated range of generations in the past. This analysis uses ROH which are between 4 and 8 centimorgans. While this dissertation uses a simple equation based on the recombination rate of the human genome, estimating these ROH coalesce between 6.25 to 12.5 generations, both Ringbauer and colleagues (2021) and Fernandes and colleagues (2021) propose that ROH of shorter lengths may coalesce 30 or even 50 generations in the past (Ringbauer et al, 2021; Fernandes et al, 2021). This is a considerable difference in estimation, which clearly depends on the statistical method used. Therefore, it is important to further investigate how we can generationally interpret these ROH ranges before making conclusions on demography.

Returning to our 6.25 to 12.5 generational interpretation of ROH 4-8 cM, we recall that this is a range of approximately 175 years where the N_e estimation can be placed in the time series, of which the true location is unknown. This uncertainty becomes further amplified when binning. When analysing two unrelated individuals within the same time bin and obtaining a single N_e estimation, it is

assumed that this estimation is reflective of the same historical population of both individuals at the same time point, some 6 to 13 generations back. When considering the 4-8 cM ROH from each individual independently, it is impossible to tell whether these ROH are, for example, reflective of 6 generations back for one individual, and 13 for the other. A single bin estimate assumes that these are the same. This thus widens the gap of temporal uncertainty of demographic reconstruction, assuming the N_e from two people reflects the same time frame when they may be up to 300 years apart (Figure 21). Thus, the binned N_e estimation may not be representative of either individual's ancestral population. Many events such as war, plague or famine can substantially impact a population over such a large time range and thus may be missed during demographic reconstruction by assuming these individuals and their ancestral populations are of the same period. The predominant way to increase the confidence of an N_e and its respective time point would be to do away with binning entirely or at least reduce bin sizes. However, as it stands, this is often not possible due to poor sample sizes or less informative genetics, as even a binning size of two generations significantly reduces the confidence interval of N_e estimations.

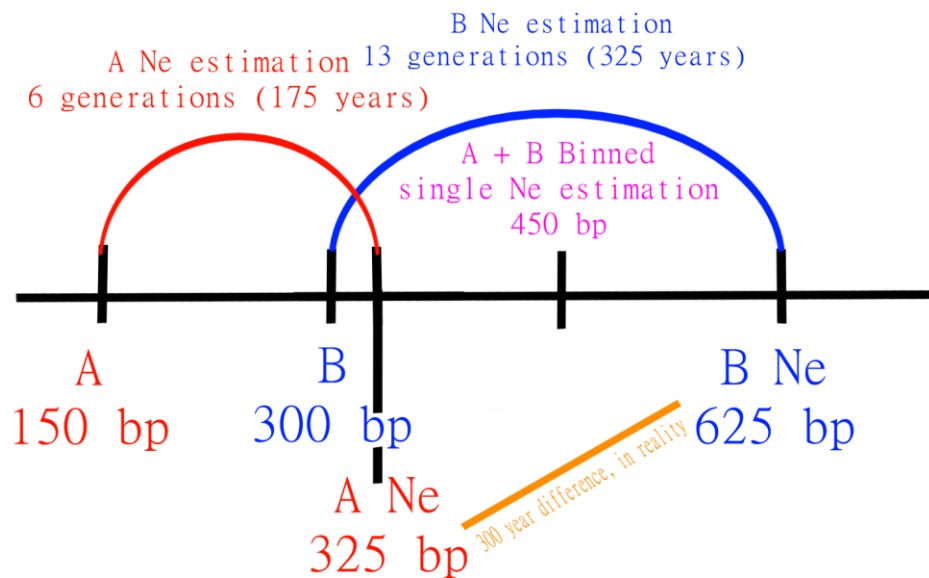


Figure 21: A hypothetical binning scenario. Individual A and B are considered as part of the same population in the same 5-gen time bin when estimating N_e , even though they are 150 years apart. The resulting single N_e estimation is for that binned population at a single time point in the past. For ROHs between 4-8 cM, the ancestral backgrounds for the ancestral populations of these two individuals could be up to 300 years apart. Thus, this single point may not be reflective of the true N_e for either ancestral population.

We also encourage the exploration and use of other genetic approaches in reconstructing N_e . genetic markers such as SNP inheritance due to linkage disequilibrium have just begun to be used to reconstruct N_e , which show promise, at the very least with regards to contemporary samples (Santiago et al, 2023). More complex and novel models, such as sibship assignment analysis and LDNE, which are also based on linkage disequilibrium, prove that the field of population genetics is ever expanding (Wang, 2009; Waples and Do, 2008; Waples, 2026). However, the methods for ROH detection are still being investigated and improved, such as through the use of simulations to optimise the detection of ROH (Shi et al, 2026). Completely novel approaches besides hapROH, such as ROHan, are also being developed and improved, which may improve upon some of the limitations present in hapROH (Renaud, 2019; Taylor, 2025).

As it stands, the estimation of effective population size (N_e) from ROH may prove useful as a tool for historical demographic reconstruction. While this thesis primarily investigates four regions of West Asia, being Anatolia, the Caucasus, Iran, and the Levant, their demographic trends estimated by N_e are consistent with other proxies, such as radiocarbon and archaeological settlement proxies, and correspond well with historical and climatic events. The inclusion of a well-supported proxy can support existing demographic hypotheses, while it may also be able to pick up on demographic signals that are missed by other proxies. While the estimation of N_e through ROH does make inherent assumptions and is not perfect, we can encourage the use of this proxy in tandem with others to create a more complete and supported demographic reconstruction.

9. Acknowledgements

Thank you to Hélios and Dr. Pagani for their welcoming and supportive demeanour and availability in the lab.

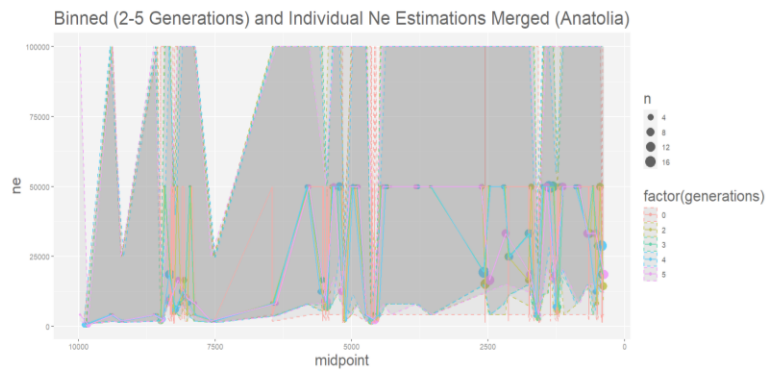
Special thanks to Francesco Santi and Jadwiga Wisniewska for their friendship and bright minds, helping me work through my ideas with their curious and enthusiastic natures.

10. Supplementary Materials

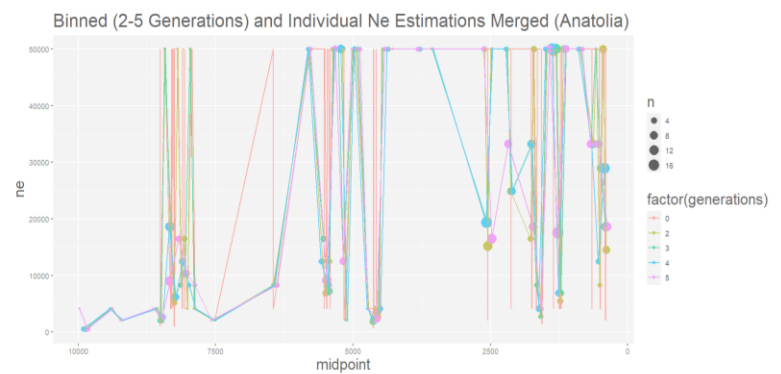
The below six images are larger versions of Figure 2 from section 7.1 to account for possible resolution/size issues during the exportation of this dissertation. In order: Anatolia, Arabia, Caucasus. Iran, Levant, Mesopotamia.

Anatolia

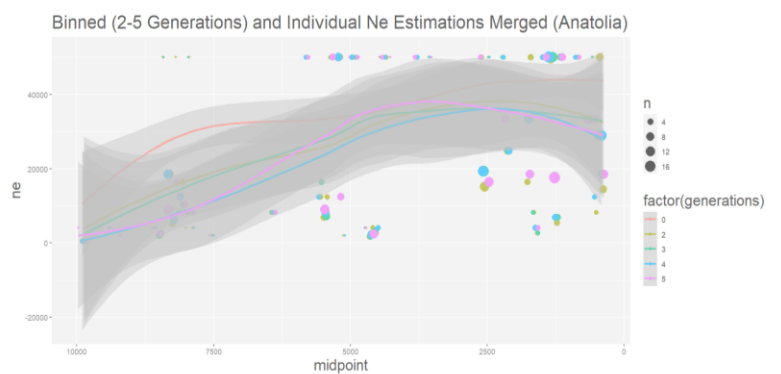
Plotted With Confidence Intervals



Plotted W/O Confidence Intervals

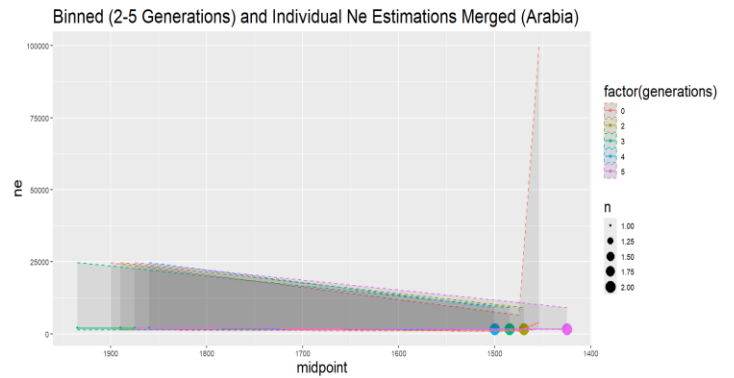


Plotted W/ Smoothing (Loess)

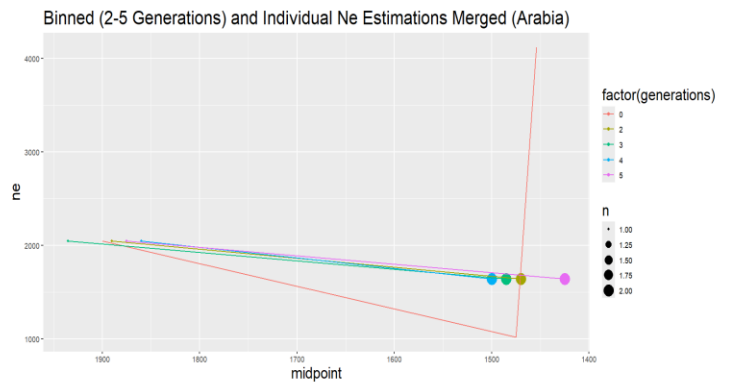


Arabia

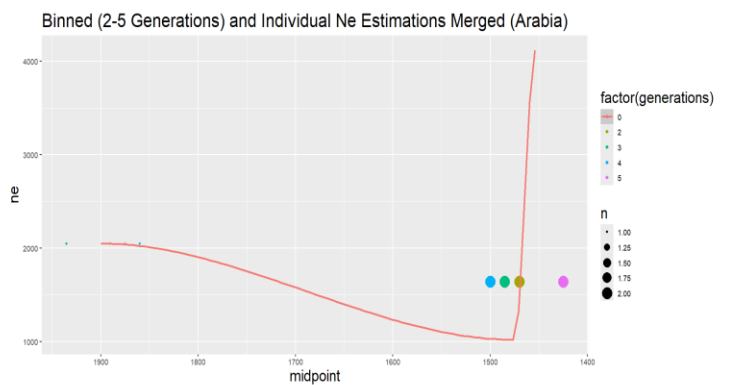
Plotted With Confidence Intervals



Plotted W/O Confidence Intervals

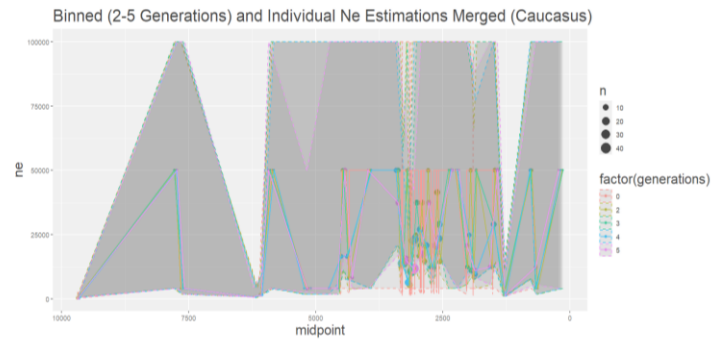


Plotted W/ Smoothing (Loess)

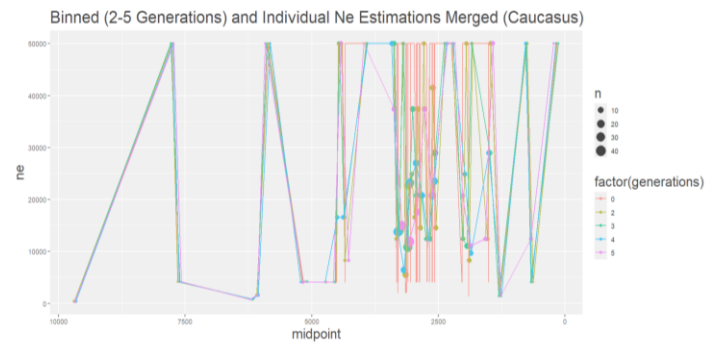


Caucasus

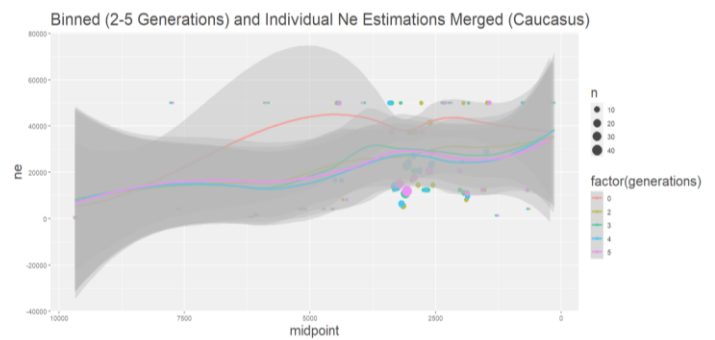
Plotted With Confidence Intervals



Plotted W/O Confidence Intervals

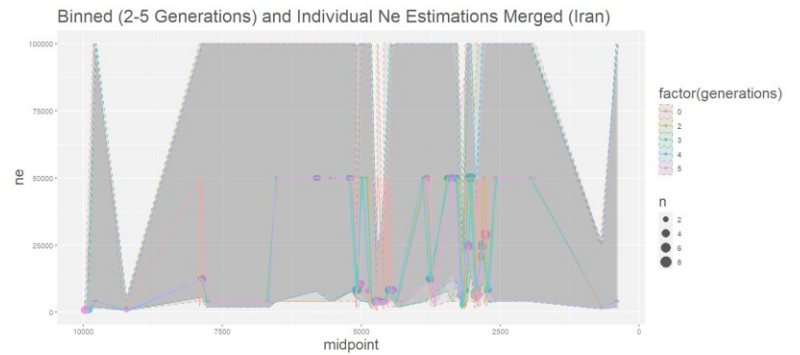


Plotted W/ Smoothing (Loess)

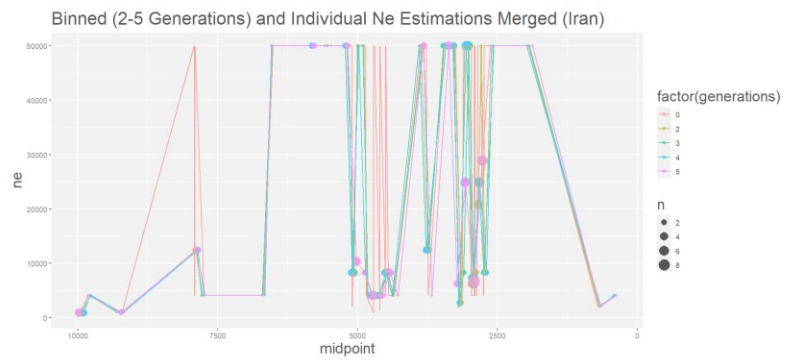


Iran

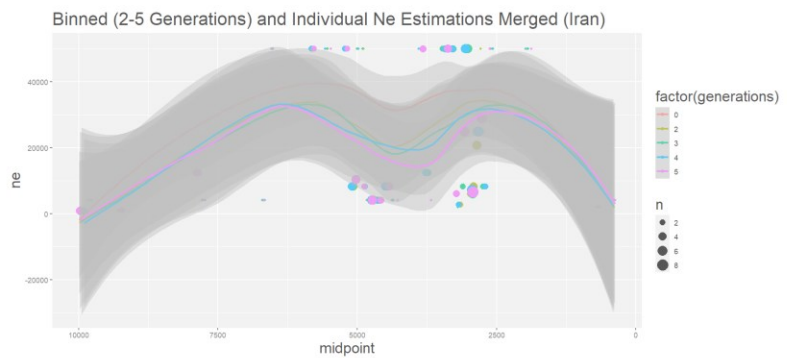
Plotted With Confidence Intervals



Plotted W/O Confidence Intervals

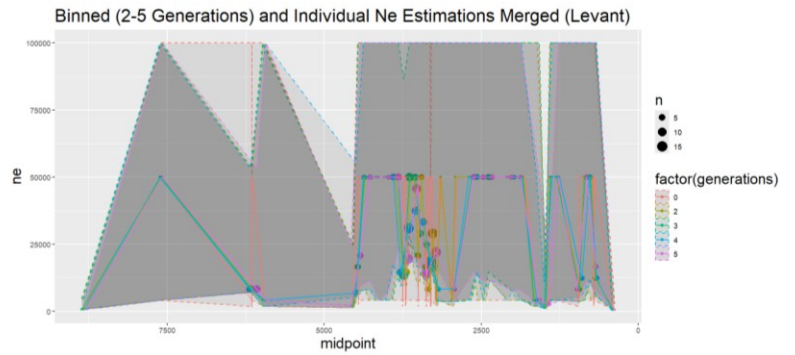


Plotted W/ Smoothing (Loess)

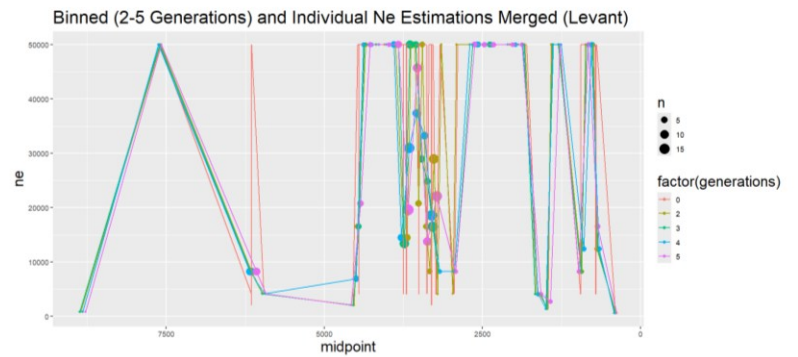


Levant

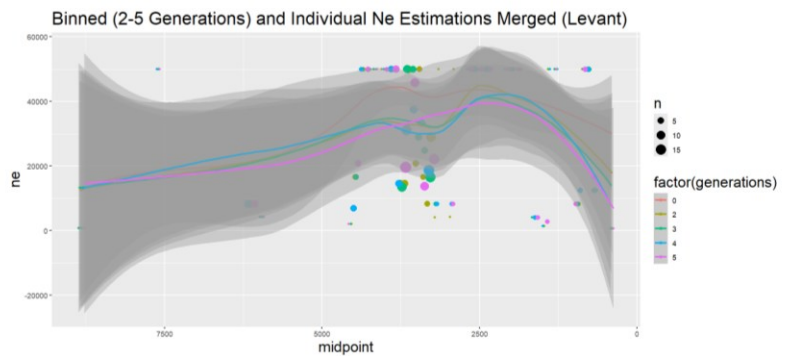
Plotted With Confidence Intervals



Plotted W/O Confidence Intervals

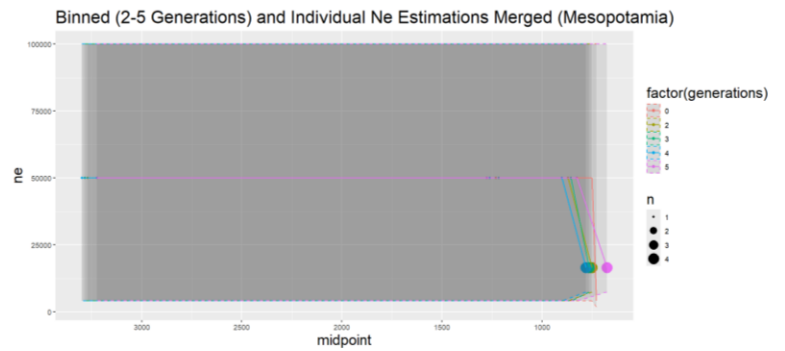


Plotted W/ Smoothing (Loess)

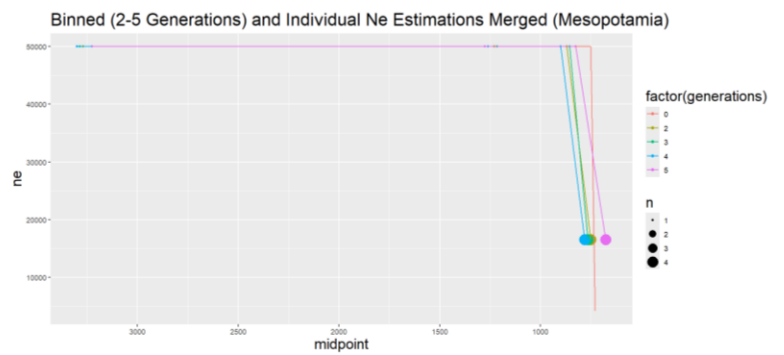


Mesopotamia

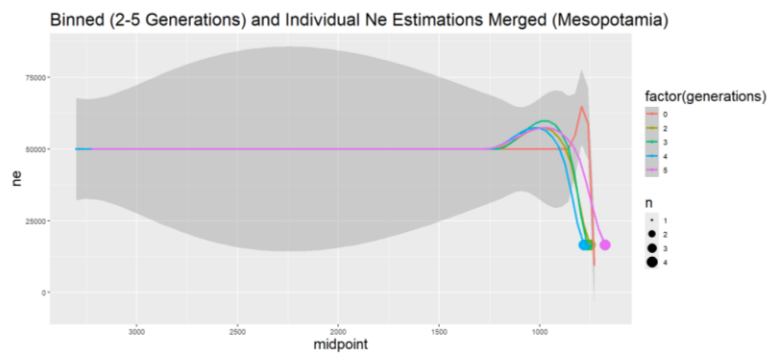
Plotted With Confidence Intervals



Plotted W/O Confidence Intervals



Plotted W/ Smoothing (Loess)



11. References

- Ascalone, E. (2022). E. Ascalone, Preliminary Report on the 2017 Excavations in Area 33 in Shahr-i Sokhta: Stratigraphy, Finds and Pottery. In E. Ascalone & S.M.S. Sajjadi (Eds.), *Excavations and Researches at Shahr-i Sokhta 2 (= ERSS 2)*. Pishin Pajouh, Tehran, pp. 189–323. ISBN: 978622972972.
- Baum, C.F., Scicchitano, S., Zaiceva, A. et al. (2025). Population change: effects on the environment, society, and economy. *Eurasian Econ Rev*, 15, 915–922. <https://doi.org/10.1007/s40822-025-00348-5>
- Bonfigli, A., Cesare, P., Volpe, A.R., Colafarina, S., Forgione, A., Aloisi, M., Zarivi, O. & Poma, A.G.M. (2023). Estimation of DNA Degradation in Archaeological Human Remains. *Genes (Basel)*, 14(6), 1238. doi: 10.3390/genes14061238. PMID: 37372418; PMCID: PMC10298407.
- Buresch, J.M., Medgyesi, D., Porter, J.R. & Hirsch, Z.M. (2024). Understanding how population change is associated with community sociodemographics and economic outcomes across the United States. *Front. Hum. Dyn.*, 6, 1465218. doi: 10.3389/fhumd.2024.1465218.
- Cleveland, W.S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association*, 74(368), 829–836. JSTOR. <https://doi.org/10.2307/2286407>. Accessed 22 Aug. 2026.
- Contreras, D.A. & Meadows, J. (2014). Summed radiocarbon calibrations as a population proxy: a critical evaluation using a realistic simulation approach. *Journal of Archaeological Science*, 52, 591–608. ISSN 0305-4403. <https://doi.org/10.1016/j.jas.2014.05.030>
- Deffner, D., Kandler, A. & Fogarty, L. (2022). Effective population size for culturally evolving traits. *PLoS Comput Biol.*, 18(4), e1009430. doi: 10.1371/journal.pcbi.1009430. PMID: 35395004; PMCID: PMC9020689.
- Derex, M. & Mesoudi, A. (2020). Cumulative Cultural Evolution within Evolving Population Structures. *Trends in Cognitive Sciences*, 24(8), 654–667. ISSN 1364-6613. <https://doi.org/10.1016/j.tics.2020.04.005>
- Fazeli Nashli, H., Sarlak, S., Hoseinzadeh, J., Yazdani, S., Rafiei-Alavi, B. & Carolin, S.A. (2022). The North and South Central Plateau of Iran During the Third and Second Millennium BCE (3200–1500 BCE). *Journal of Archaeological Studies*, 14(2), 153–180. doi: 10.22059/jarcs.2022.348692.143160.
- Fernandes, D.M., Sirak, K.A., Ringbauer, H. et al (2021). A genetic history of the pre-contact Caribbean. *Nature* 590, 103–110. <https://doi.org/10.1038/s41586-020-03053-2>

- Fogarty, L., Wakano, J.Y., Feldman, M.W. et al. (2017). The Driving Forces of Cultural Complexity. *Hum Nat*, 28, 39–52. <https://doi.org/10.1007/s12110-016-9275-6>
- Freeman, J., Byers, D.A., Robinson, E. & Kelly, R.L. (2018). Culture Process and the Interpretation of Radiocarbon Data. *Radiocarbon*, 60(2), 453–467. doi: 10.1017/RDC.2017.124.
- French, J.C. & Collins, C. (2015). Upper Palaeolithic population histories of South western France: a comparison of the demographic signatures of 14 C date distributions and archaeological site counts. *J. Archaeol. Sci.*, 55, 122e134.
- Howrigan, D.P., Simonson, M.A. & Keller, M.C. (2011). Detecting autozygosity through runs of homozygosity: A comparison of three autozygosity detection algorithms. *BMC Genomics*, 12, 460. <https://doi.org/10.1186/1471-2164-12-460>
- Kirin, M., McQuillan, R., Franklin, C.S., Campbell, H., McKeigue, P.M. et al. (2010). Genomic Runs of Homozygosity Record Population History and Consanguinity. *PLOS ONE*, 5(11), e13996. <https://doi.org/10.1371/journal.pone.0013996>
- Palmisano, A., Bevan, A. & Shennan, S. (2017). Comparing archaeological proxies for long-term population patterns: An example from central Italy. *Journal of Archaeological Science*, 87, 59–72. ISSN 0305-4403. <https://doi.org/10.1016/j.jas.2017.10.001>
- Palmisano, A., Lawrence, D., de Gruchy, M.W., Bevan, A. & Shennan, S. (2021). Holocene regional population dynamics and climatic trends in the Near East: A first comparison using archaeo-demographic proxies. *Quaternary Science Reviews*, 252, 106739. ISSN 0277-3791. <https://doi.org/10.1016/j.quascirev.2020.106739>
- Ramussen, C.E. & Williams, C.K.I. (2006). *Gaussian Processes for Machine Learning*. MIT Press, 2006. ISBN-10 0-262-18253-X; ISBN-13 978-0-262-18253-9.
- Renaud, G., Hanghøj, K., Korneliussen, T.S., Willerslev, E. & Orlando, L. (2019). Joint Estimates of Heterozygosity and Runs of Homozygosity for Modern and Ancient Samples. *Genetics*, 212(3), 587–614. <https://doi.org/10.1534/genetics.119.302057>
- Ringbauer, H., Novembre, J. & Steinrücken, M. (2021). Parental relatedness through time revealed by runs of homozygosity in ancient DNA. *Nat Commun*, 12, 5425. <https://doi.org/10.1038/s41467-021-25289-w>
- Santiago, E., Caballero, A., Köpke, C. & Novo, I. (2024). Estimation of the contemporary effective population size from SNP data while accounting for mating structure. *Molecular Ecology Resources*, 24, e13890. <https://doi.org/10.1111/1755-0998.13890>
- Shi, M., Li, H., Shafer, A.B.A. & Lan, T. (2026). Optimising Genome-Wide Detection of Runs of Homozygosity: Impacts of Reference Genome Quality and Sequencing

- Parameters on Inbreeding Assessment. *Molecular Ecology Resources*, 26(1), e70084. <https://doi.org/10.1111/1755-0998.70084>
- Talebian, M.H. (2014). Shahr-i Sokhta: Nomination File 1456. Iranian Cultural Heritage, Handicrafts and Tourism Organization (ICHHTO), Islamic Republic of Iran. UNESCO World Heritage Centre.
- Taylor, R.S., Manseau, M. & Wilson, P.J. (2026). Accurate Runs of Homozygosity Estimation From Low Coverage Genome Sequences in Non-Model Species. *Molecular Ecology Resources*, 26(1), e70083. <https://doi.org/10.1111/1755-0998.70083>
- Tenesa, A., Navarro, P., Hayes, B.J., Duffy, D.L., Clarke, G.M., Goddard, M.E. & Visscher, P.M. (2007). Recent human effective population size estimated from linkage disequilibrium. Cold Spring Harbor Laboratory Press. <https://genome.cshlp.org/content/17/4/520>
- UNESCO World Heritage Centre. (2014). Shahr-i Sokhta. UNESCO World Heritage List.
- van Schaik, C.P., Pradhan, G.R. & Tennie, C. (2019). Teaching and curiosity: sequential drivers of cumulative cultural evolution in the hominin lineage. *Behav Ecol Sociobiol*, 73, 2. <https://doi.org/10.1007/s00265-018-2610-7>
- Wang, J. (2009). A new method for estimating effective population sizes from a single sample of multilocus genotypes. *Molecular Ecology*, 18, 2148–2164. <https://doi.org/10.1111/j.1365-294X.2009.04175.x>
- Waples, R.S. & Do, C. (2008). Ldne: a program for estimating effective population size from data on linkage disequilibrium. *Molecular Ecology Resources*, 8, 753–756. <https://doi.org/10.1111/j.1755-0998.2007.02061.x>
- Waples, R.S. (2026). Combining Multiple Genetic Estimates of Ne. *Mol Ecol Resour.*, 26, e70114. <https://doi.org/10.1111/1755-0998.70114>
- Weitzel, E.M. & Codding, B.F. (2022). The Ideal Distribution Model and Archaeological Settlement Patterning. *Environmental Archaeology*, 27(4), 349–356. <https://doi.org/10.1080/14614103.2020.1803015>
- Williams, A.N. (2012). The use of summed radiocarbon probability distributions in archaeology: a review of methods. *Journal of Archaeological Science*, 39(3), 578–589. ISSN 0305-4403. <https://doi.org/10.1016/j.jas.2011.07.014>