Ancient DNA × Climate Signal Explorer: A Methodological Note for Visualizing Genetic and Paleoclimate Time Series

Abstract

The *Ancient DNA × Climate Signal Explorer* is a heuristic research and teaching tool for visually comparing time-binned ancient DNA allele-frequency estimates with paleoclimate proxy records. It is designed for early-stage hypothesis formation, methodological transparency, and structured annotation rather than for deterministic claims about selection, adaptation, migration, or demographic replacement. Ancient DNA has made it possible to observe allele-frequency change across archaeological time, creating new opportunities to examine how genetic variation, ecological stress, diet, mobility, pathogen exposure, and climate regimes may have interacted in past populations (Mathieson et al. 2015; Schraiber, Evans, and Slatkin 2016; Skoglund and Mathieson 2018). At the same time, visual overlap between genetic and environmental series can be misleading if read outside its archaeological, demographic, chronological, and technical contexts. This note defines the tool's purpose, input structure, binning logic, visualizer assumptions, annotation system, interpretive limits, and recommended standards for responsible use. The tool should be understood as a research-design instrument: it helps users identify candidate relationships, sampling gaps, chronological discontinuities, and questions requiring formal population-genetic, archaeological, ecological, and paleoclimatic analysis.

Keywords

Ancient DNA; paleoclimate; allele frequency; selection; climate proxy; archaeological genetics; temporal binning; visual analytics; research methodology; Stella Nova Education

Methodological Note

Purpose and Scope

The *Ancient DNA × Climate Signal Explorer* is intended to support exploratory research at the intersection of ancient genomics, archaeology, and paleoclimatology. Its primary function is to place time-binned allele-frequency estimates and paleoclimate proxy values into a shared visual frame so that users can ask whether a genetic pattern and an environmental pattern deserve closer investigation. The tool does not claim to detect selection, prove adaptation, establish migration, or identify causal relationships. Instead, it creates a structured visual environment for asking better questions.

This distinction is important because ancient DNA has changed what researchers can observe about the past. Rather than inferring all evolutionary history from present-day genomes, ancient DNA allows direct sampling of earlier populations and therefore enables time-series approaches to allele-frequency change (Schraiber, Evans, and Slatkin 2016; Skoglund and Mathieson 2018). Studies such as Mathieson et al. (2015) show that ancient genomic datasets can reveal changes in loci associated with pigmentation, diet,

immunity, and other traits across prehistoric populations. However, the same literature also demonstrates the complexity of interpreting those changes. Allele frequencies can shift because of selection, but they can also shift because of migration, admixture, drift, founder effects, replacement, sampling design, archaeological visibility, or laboratory and capture biases (Mathieson et al. 2015; Skoglund and Mathieson 2018).

The purpose of this tool is therefore methodological rather than diagnostic. It helps users compare patterns, identify periods of apparent change, inspect sparse data, evaluate whether bin width alters interpretation, and record scholarly questions through a no-database annotation system. It is most appropriate as a teaching tool, a preliminary research-planning tool, or a visual supplement to a more formal analytical workflow.

Data Model

The tool is built around two kinds of input: ancient DNA sample data and paleoclimate proxy data.

The ancient DNA data model assumes that each row represents an observed genotype or allele dosage for a dated individual or sample. At minimum, each row should include a sample identifier, region, date in years before present, locus or SNP identifier, and derived-allele dosage. Dosage is treated as a count of derived alleles, normally 0, 1, or 2 for diploid genotype data. Missing, negative, or uncertain calls should be excluded or explicitly coded according to the user's data-preparation protocol. If pseudohaploid ancient DNA calls are used, the user must recognize that the visualized allele count does not carry the same assumptions as high-coverage diploid genotype data.

The climate data model assumes that each row contains a date in years before present, a proxy name, and a proxy value. Paleoclimate data may derive from tree rings, ice cores, corals, speleothems, lake sediments, ocean sediments, pollen records, or other natural archives. NOAA's paleoclimatology archive describes these proxy sources as extending weather and climate information across timescales ranging from hundreds to millions of years, often through geophysical or biological measurements and reconstructed variables such as temperature or precipitation (National Centers for Environmental Information 2026). Because each proxy has its own depositional environment, calibration history, resolution, and uncertainty, proxy values must be interpreted according to the original archive and publication from which they were drawn.

The tool does not standardize climate proxies automatically. A positive value in one proxy series may indicate warming, while a positive value in another may indicate aridity, precipitation, isotopic enrichment, vegetation change, or another environmental process. For that reason, the tool should never be used without a written explanation of what the selected proxy measures, where it was collected, how it was dated, and what directionality its values imply.

Temporal Binning Procedure

The visualizer groups ancient DNA samples into user-selected temporal bins. Within each bin, the tool sums the observed derived-allele dosages and divides that sum by the total number of observed chromosomes or allele observations. The resulting value is treated as an estimated derived-allele frequency for that bin. These binned estimates are then plotted against time, allowing the user to examine whether allele-frequency change appears gradual, abrupt, discontinuous, or unstable.

Temporal binning is useful because ancient DNA datasets are often unevenly distributed across time. Some periods may contain many individuals, while others may contain only one or two samples. Binning can reduce visual noise, but it can also create false stability or false transitions. A large bin may merge distinct archaeological periods or populations into a single estimate. A narrow bin may appear precise while relying on too few observations to support meaningful interpretation. Users should therefore test multiple bin widths and report whether the visual pattern is robust to those changes.

The tool's confidence band is only a simple binomial display aid. It is meant to warn users when a bin is based on few observations, not to provide a full uncertainty model. It does not account for radiocarbon uncertainty, temporal autocorrelation, kinship, relatedness, capture bias, pseudohaploid calling, genotype likelihoods, population structure, changing ancestry proportions, or uncertainty in the climate chronology. It is therefore inappropriate to describe the visual band as a formal confidence interval for selection, demographic continuity, or environmental causation.

Paleoclimate Matching

For display purposes, the tool matches each genetic time bin to the nearest paleoclimate proxy observation. This nearest-neighbor approach is transparent and easy to understand, but it is not a substitute for paleoclimate modeling. Proxy records often have their own sampling resolutions, chronological uncertainties, smoothing effects, depositional lags, and local environmental constraints. A genetic sample dated to a particular archaeological period may not correspond exactly to a climate observation in the same calendar range, especially if either chronology has wide uncertainty.

Because of this, users should treat proxy matching as a visual alignment, not as a claim of synchronous causality. If an apparent relationship emerges, the next step should be to examine the original paleoclimate archive, the proxy's dating model, the archaeological chronology, and the spatial distance between the sampled human population and the environmental record. A proxy taken from a distant lake core, marine sediment sequence, or ice core may describe regional or global climate trends but may not capture the local ecological pressures experienced by a specific archaeological population.

Visualizer Element

The visualizer is designed to make three elements visible at once: the binned allele-frequency trajectory, the number of allele observations supporting each bin, and the selected paleoclimate proxy series. Marker size or tabular summaries can be used to indicate the number of observations in each bin, reminding users that an apparent genetic trend may be supported by uneven data. The accompanying summary panel reports the number of temporal bins, allele observations, median frequency, and oldest-to-youngest change.

The visualizer is deliberately framed as an interpretive aid. Its purpose is to help the user notice questions such as:

Does the allele-frequency trajectory change during the same broad period as a climate proxy shift?

Is the apparent change driven by one sparsely sampled bin?

Does the pattern remain visible when bin width changes?

Are there gaps in the ancient DNA record that make the visual pattern unreliable?

Could the apparent transition reflect population replacement rather than selection within a continuous population?

Does the archaeological context support a plausible mechanism linking the locus, phenotype, ecology, and proxy?

These questions are more important than the image itself. A visually persuasive overlap between an allele trajectory and a climate proxy can be methodologically dangerous if treated as proof. The visualizer should slow interpretation down, not accelerate it beyond the evidence.

No-Database Annotation System

The tool includes a no-database self-postback annotation section. Rather than requiring MySQL or a server-side database, user comments are submitted through the same PHP page and appended to a local JSON file. This design is appropriate for lightweight shared-hosting environments and for classroom or prototype use where the goal is to preserve methodological questions, corrections, source suggestions, and interpretive cautions.

The annotation section should be treated as a scholarly field notebook rather than as a social comment thread. Useful annotations may include notes about sample provenance, concerns about bin width, suggestions for alternative proxies, questions about locus function, warnings about population discontinuity, or references to relevant archaeological evidence. Because the annotation system writes to a local JSON file, it should be protected with ordinary shared-hosting safeguards: file permissions should be reviewed, spam controls should be considered, and the JSON file should be backed up before public deployment.

Interpretive Limits

The most important methodological limit is that correlation is not causation. Apparent covariation between allele frequency and climate may reflect many processes. A locus may rise in frequency because of natural selection, but it may also rise because a new population entered the region, because sample ancestry changed, because a burial context overrepresents a particular group, because of drift in a small population, because of kinship among sampled individuals, or because the time bins merge unrelated communities.

Formal methods exist for inferring selection from allele-frequency time series, including Bayesian and likelihood-based approaches (Schraiber, Evans, and Slatkin 2016; Stern, Wilton, and Nielsen 2019). Those methods are designed to estimate selection parameters or allele-frequency trajectories with more explicit assumptions than a visualizer can provide. Even then, demographic history can bias inference if not modeled carefully (Schraiber, Evans, and Slatkin 2016). The Stella Nova tool does not implement those methods. It should therefore not be described as a selection test.

The tool also does not model cultural behavior. Climate may shape subsistence, settlement, mobility, disease ecology, and technological practice, but humans respond to environmental pressure through social systems, not biology alone. Archaeological evidence is therefore essential. A genetic-climate association becomes more plausible only when it can be situated alongside material culture, settlement history, dietary

evidence, isotopic data, pathogen evidence, mobility studies, and regional paleoenvironmental reconstruction.

Recommended Workflow

A responsible workflow should begin with provenance. Users should document where the ancient DNA data came from, whether the samples are published, whether permissions and ethical guidelines allow reuse, how the samples are dated, and whether individuals are archaeologically and genetically appropriate to compare. The next step is genotype review. Users should identify whether the data are diploid genotypes, pseudohaploid calls, imputed genotypes, or another format, and they should document missing data handling.

After provenance and genotype review, users should test the visualization across multiple bin widths. If a pattern appears only under one binning scheme, it should be treated cautiously. If the pattern persists across several reasonable bin widths, it may deserve deeper analysis. Users should then examine demographic alternatives. This includes checking whether ancestry proportions change across the same period, whether the region experienced known migration or population replacement, and whether the sampled individuals plausibly represent a continuous population.

The climate proxy should then be reviewed independently. Users should cite the original proxy publication and archive, identify the proxy type, explain the proxy's environmental meaning, and discuss chronological uncertainty. Only after these steps should a candidate relationship be described. Even then, the preferred wording should be modest: the tool identifies a candidate temporal association that requires formal testing and archaeological corroboration.

Academic Standard for Use

The appropriate academic claim from this tool is not "this allele changed because the climate changed." The appropriate claim is closer to the following: "A preliminary visual comparison suggests that this locus and this proxy show temporal change across overlapping periods; however, the relationship remains exploratory and requires demographic modeling, proxy-specific interpretation, and archaeological corroboration."

This standard matters because ancient DNA is powerful but easily overinterpreted. It can show that allele frequencies changed through time, but explaining why they changed requires a broader evidentiary frame. The tool is strongest when used as a bridge between visual exploration and formal research design. It helps organize uncertainty, not eliminate it.

References Cited

Mathieson, Iain, Iosif Lazaridis, Nadin Rohland, Swapan Mallick, Nick Patterson, Songül Alpaslan Roodenberg, Eadaoin Harney, et al.
2015 Genome-Wide Patterns of Selection in 230 Ancient Eurasians. *Nature* 528:499–503.

National Centers for Environmental Information
2026 Paleoclimatology. National Oceanic and Atmospheric Administration, National Centers for Environmental Information. Electronic document, accessed June 26, 2026.

Schraiber, Joshua G., Steven N. Evans, and Montgomery Slatkin
2016 Bayesian Inference of Natural Selection from Allele Frequency Time Series. *Genetics* 203(1):493–511.

Skoglund, Pontus, and Iain Mathieson
2018 Ancient Genomics of Modern Humans: The First Decade. *Annual Review of Genomics and Human Genetics* 19:381–404.

Stern, Aaron J., Peter R. Wilton, and Rasmus Nielsen
2019 An Approximate Full-Likelihood Method for Inferring Selection and Allele Frequency Trajectories from DNA Sequence Data. *PLOS Genetics* 15(9):e1008384.