

# BIO00056I

## Directed study 11 (worksheet): Population genomics

Daniel Jeffares

2025-12-03

### Table of contents

|          |                                     |          |
|----------|-------------------------------------|----------|
| <b>1</b> | <b>Learning objectives</b>          | <b>1</b> |
| <b>2</b> | <b>Introduction</b>                 | <b>2</b> |
| 2.1      | The hypothetical scenario . . . . . | 2        |
| 2.2      | Background Information . . . . .    | 3        |
| 2.2.1    | The data . . . . .                  | 3        |
| 2.2.2    | The problem . . . . .               | 3        |
| 2.3      | The questions . . . . .             | 4        |
| <b>3</b> | <b>Model answers</b>                | <b>4</b> |

! This document now contains answers to questions.

### 1 Learning objectives

This worksheet is designed to prepare you for exam questions on population genomics. Upon completing this worksheet your knowledge will be enhanced in these areas:

- Understanding how population genomic data can be collected
- Appreciating how population genomic data can be applied to various biological questions
- Interpret various patterns that arise from population genomic data analysis

## 2 Introduction

This exercise describes a scenario where you can use population genomics to study ancient history. One of the most ancient small cities is [Çatalhöyük](#) which is in modern day Turkey. This little city dates to about 7000 BCE. We know that this civilization had either goats or sheep, but we do not know which. The archeologists also suspect that the people of Çatalhöyük had a shared contemporaneous culture and traded with people at Aşıklı Höyük and some other settlements in the same part of the world (Figure 1).

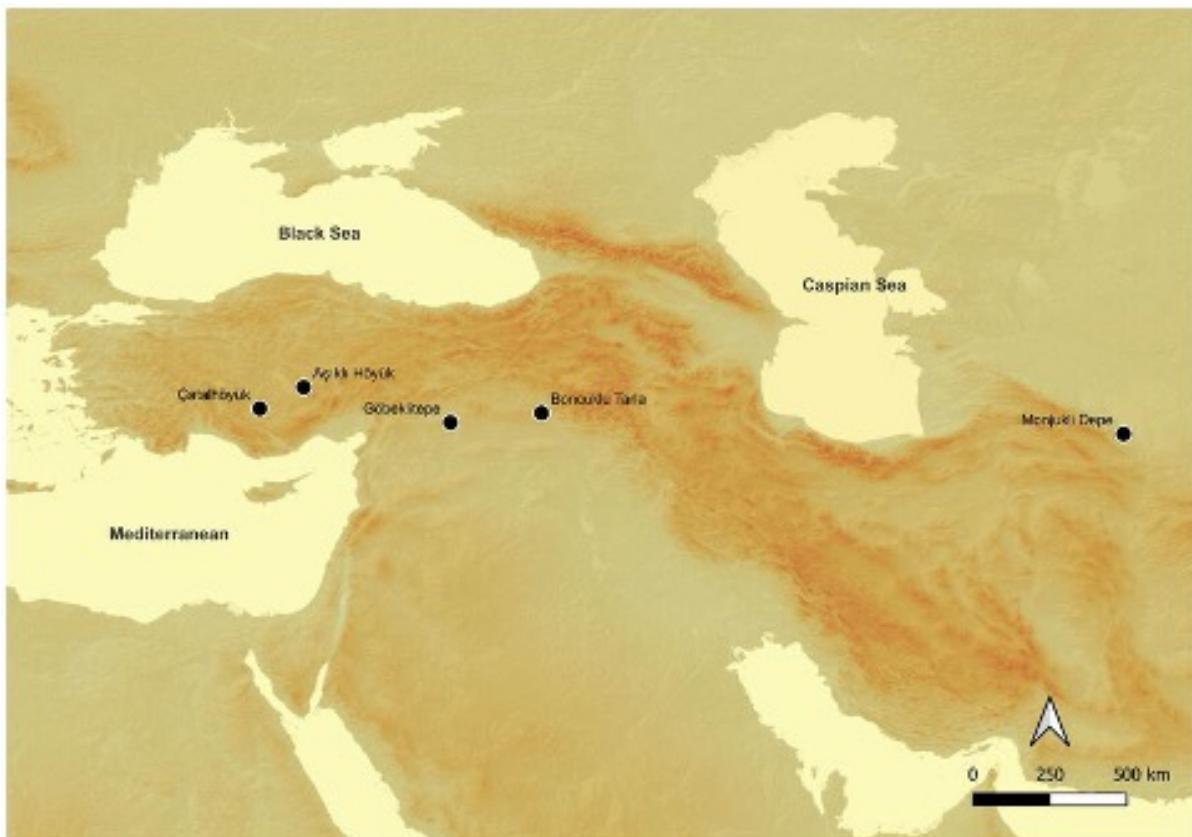


Figure 1: Figure 1. Archaeological sites that are contemporaneous with Çatalhöyük

### 2.1 The hypothetical scenario

The remainder of this problem is hypothetical. Imagine that archaeologists disagree about whether Çatalhöyük and Aşıklı Höyük regularly trade. There is disagreement about whether this civilization domesticated goats (if they were goats) or if they merely caught and ate wild goats.

Luckily, in this scenario, we can occasionally obtain small amounts of DNA from the bones of sheep (or goats) at Çatalhöyük and Aşıklı Höyük. The samples we consider are described below. Note that while this is a hypothetical example, studies of ancient civilizations do use DNA and population genomics.

For example, the study of [Population genomics of the Viking world](#) used population genomics extensively. Çatalhöyük is older, so in our scenario, we assume the quality and quantity of DNA is much lower.

## 2.2 Background Information

There is a high-quality reference genome for sheep and goats. Studies have shown that it is possible to map reads from sheep and goats and vice versa. But it is better to map reads within species, rather than between.

### 2.2.1 The data

You can obtain small amounts of DNA from these ancient sheep/goat samples:

- 10 samples from Çatalhöyük
- 10 samples from Aşıklı Höyük

There are thousands of genomes of modern goats and available at the European Nucleotide Archive (ENA) from the [VarGoats project](#). This includes the genomes of *domesticated* goats from Pakistan, Spain & France, the Mediterranean, Northern Europe and most importantly, *wild* goats from Turkey and Iran.

There are also more than 100 ancient genomes from Eurasian domestic and wild sheep samples spanning the past 12,000 years [from this article](#), and 100s of modern sheep genomes available at ENA.

#### Note

You do not need to read these articles, but they show us what population genomic data is available.

### 2.2.2 The problem

The archaeologists want to know whether Çatalhöyük and Aşıklı Höyük regularly traded sheep and/or goats.

An experienced ancient DNA technician has extracted DNA for you from Çatalhöyük and Aşıklı Höyük. There isn't much, but you can sequence it. You do, and it is disappointing -

the sequence reads map to only ten tiny regions in 0.5% of the goat genome and 0.4% of the sheep genome. They don't have the funds or the samples to obtain any more.

The challenge is to explain what this might mean if the SNPs show the samples are more like sheep, more like goats, or in between, and if they traded sheep and goats with each other, or merely caught wild sheep and goats.

### 2.3 The questions

**Question 1.** Explain how you would use population genomics data to determine if the people of Çatalhöyük and Aşıklı Höyük kept sheep, goats, or both? (10 marks).

In your answer, explain a) what samples you would use from ENA and how many (3 marks), b) what steps you would need to take to process the genome sequence data to obtain SNPs (3 marks), c) once you have SNP calls, what kinds of plots or analysis you would conduct to determine if the Çatalhöyük and Aşıklı Höyük samples were sheep or goats and what the different displays would show you (4 marks).

**Question 2.** Explain how you could use the 10 samples from Çatalhöyük and the 10 samples from Aşıklı Höyük to determine if the people of Çatalhöyük and Aşıklı traded sheep and/or goats (3 marks). Hint: Trading would appear like migration of goats.

**Question 3.** There is a proposal that the people of Çatalhöyük and Aşıklı domesticated goats, and this domestic breed gave rise to modern domestic breeds in Pakistan and the Mediterranean. Explain how you could use population genomics data to test this proposal. In this answer, assume we have SNP calls from these locations and explain how you would visualize and interpret the data (6 marks).

- The 10 ancient goat genomes from each of Çatalhöyük and Aşıklı
- Hundreds of modern domestic goats from Pakistan and the Mediterranean
- A few dozen wild goats from Turkey and Iran

## 3 Model answers

**Question 1.** Explain how you would use population genomics data to determine if the people of Çatalhöyük and Aşıklı kept sheep, or goats or both? (10 marks).

Ultimately, a PCA plot would be the most useful way to determine if the samples are more like sheep or goats. To generate this plot, we would need to follow these steps:

- a) Download genomes of modern wild and domestic sheep and goats from ENA. About 50 samples of each species should be sufficient to make a PCA plot that separates the sheep from the goats. If possible, we should include sheep and goats from Turkey and nearby regions (3 marks).
- b) To process the genome sequence data to obtain SNPs, we would obtain a reference sheep genome (or reference goat genome) and map the reads from ENA and the ancient samples to the reference genome using a read mapper like BWA. Next, we would use a variant caller to call SNPs across all samples, generating a file containing all the SNPs from all the samples (3 marks).
- c) Once we have SNP calls, we would use a tool like PLINK to generate the PCA coordinates. We would plot the first two principal components. If the Çatalhöyük and Aşıklı samples cluster with the sheep samples, they are likely sheep. If they cluster with the goat samples, they are likely goats. If they fall in between, it may suggest they are hybrids or that there is some uncertainty (4 marks).

**Question 2.** Explain how you could use the 10 samples from Çatalhöyük and the 10 samples from Aşıklı Höyük to determine if the people of Çatalhöyük and Aşıklı Höyük traded sheep and/or goats (3 marks). Hint: Trading would appear like migration of goats.

Let's assume they kept goats. Since trading of goats would appear like *migration*, we would expect to observe goats in Çatalhöyük that are genetically similar to goats in Aşıklı Höyük or vice versa (1). A PCA plot may show this, but only if these two places had different breeds (like different populations), and we had a *fairly recent* goat trade in the data set of 10 ancient samples from each place (1). Such a recent traded goat would also be visible on a phylogenetic tree, as a Çatalhöyük goat in the Aşıklı clade or vice versa (1).

**Question 3.** Imagine there is a proposal that the people of Çatalhöyük and Aşıklı Höyük domesticated goats, and their domestic breeds gave rise to modern domestic breeds in Pakistan and the Mediterranean. Explain how you could use population genomics data to test this proposal. (6 marks)

If this were the case, then we would expect the modern domestic breeds in Pakistan and the Mediterranean to be genetically similar to the ancient goats from Çatalhöyük and Aşıklı Höyük. Alternatively, if domestication occurred again from wild goats, the modern domestic samples would look different. A PCA plot would help here (3);

- We would plot the ancient Çatalhöyük and Aşıklı samples along with the modern domestic goats from Pakistan and the Mediterranean, and the wild goats from Turkey and Iran. If the modern domestic goats cluster closely with the ancient Çatalhöyük and Aşıklı samples, it would support the proposal that they are descended from these ancient domesticated goats (2). If they cluster separately, it would suggest a different origin (1).