

Project title: Genetic insights into the cultivation history of poppy

Lead supervisor: Surabhi Ranavat, Postdoctoral Researcher in the Hidden Frontiers Project, Department of Archaeology, surabhi.ranavat@york.ac.uk

Project description

Overview: The domestication history of the opium poppy is a major question for archaeologists, botanists and the medical community. Ancient DNA research provides a way to explore the human past using genomic sequencing data. Here, DNA from poppy accessions will be used to explore the pace of domestication and its use for humans in the past. During this internship, data from poppy specimens will be processed with a bioinformatic pipeline for archaeological crops. The wider goal of this project is to set the foundation for expanded analysis of ancient poppy specimens that are being processed in York's Ancient DNA laboratory.

Computational requirements: This project will use the University of York's Viking high-performance compute facility. Published genomic data is available for 220 modern and 100 herbarium accessions of *Papaver* species, amounting to ~4 terabytes of raw DNA sequencing data. These data will be sequentially downloaded from public repositories and mapped to the reference genome so that hundreds of thousands of genetic markers can be characterised. This analysis will be done with a Nextflow pipeline developed in our group, in part thanks to a previous internship overseen by Dr Ranavat and Senior Lecturer Dr Nathan Wales. Based on ongoing work in grapevine, the project will require the standard "nodes" partition and ~1000 hours of compute time to process the raw data.

Necessity of HPC compute resources: As is the case with most bioinformatic processes, ancient DNA or "palaeogenomic" research requires parallel processing, ideally in a SLURM system. The datasets cannot be readily processed on a personal computer but the national supercomputers are not necessary. York's Viking HPC provides an ideal arrangement to process large genomic datasets. Additionally, Dr Lacy guided us with the best practices for Nextflow for an N8-CIR internship in our lab last summer.

Approximate schedule: As we do not yet have a student selected for the project, there is some uncertainty in the amount of time needed to train the student in submitting SLURM jobs and Nextflow, however, based on past internships, we anticipate the following arrangements.

- Week 1: Explore literature on poppy domestication and set up Viking account
- Week 2: Nextflow training and Zoom meeting with colleagues genotyping historic poppy
- Week 3: Download first raw dataset and map to reference genome
- Week 4: Continue downloading data and mapping
- Week 5: Depending on speed of processing, complete remaining samples or include archaeological samples into the analysis
- Week 6: Produce PCA figure to find links between archaeological and modern poppies
- Week 7: Document pipeline so that additional samples can be included
- Week 8: Create presentation so that results can be shared with diverse audiences

Expected outcome: This internship will test our Nextflow pipeline with a new species. The use of poppy's analgesic and psychotropic properties by past human populations is an area ripe for research, and the genetic patterns we identify will undoubtedly feature in high-impact publications.