Ancient Genome Cross-validation of Modelled Selection Co-efficients in Middle-Eastern & African Populations

Tom Davy [t.davy17@imperial.ac.uk]

December 9, 2017

Main Supervisor: Dr. Jason Hodgson, Imperial College London, j.hodgson@imperial.ac.uk

Key Words: Human Evolution, Selection, Introgression, Population Genetics, Ancient DNA

1   Project Introduction

The first currently surviving human population lineages to have left Africa would have been exposed to an entire suite of environmental pressures in the form of changes to diet, ecosystem & pathogenic landscape, all of which would have acted upon the genomic landscape of these migratory populations, possibly selecting for certain variants from background variation or novel mutation.

Recent analyses [Gresham, 2017] have demonstrated that populations in the Middle-East (ME), after suitable time for a selective process to occur, had the opportunity to admix with African populations, leading to both populations experiencing admixture and conserving a quantifiable amount of introgressed genetic content. Applications of the LSBL test has suggested that such tracts may be under selective pressure, and an overlap in the candidate loci for selection between these populations are under selection within both populations, suggesting that admixture can function as an adaptive event at the population level. In this project, I aim to delineate these selection co-efficients via modelling the population history of these allele frequencies, then applying this workflow to a real dataset consisting of SNPs for 259,257 SNPs from across several African populations [Hodgson et al., 2014].

Importantly, recent publication of ancient African genomes [Skoglund et al., 2017, Llorente et al., 2015] may allow us to validate my Selection co-efficient predictions through time using real empirical data.

2   Proposed Methodology

Firstly, a simulation of the experimental framework will be conducted with forward simulations through SFS-code [Hernandez and Bateman, 2008]. Ancestral and admixed populations, upon which selection would have acted on sites of introgression as seen in [Gresham, 2017] will be scored using the adapted locus-specific branch length (LSBL) test [Gresham, 2017], which can be leveraged to predict selection-co-efficient as a function of daughter population branch width. Presence or absense of predicted genotypes in the admixed across the ancestral population genomes will then be scored for correct predictions, giving us a metric against which to validate this process. Model fitting of differing simulation parameters will be performed via a log-likelihood ratio approach. Simulated VCF files will also be verified in Plink [Chang et al., 2015] via the pair-wise F-statistic before downstream analyses.
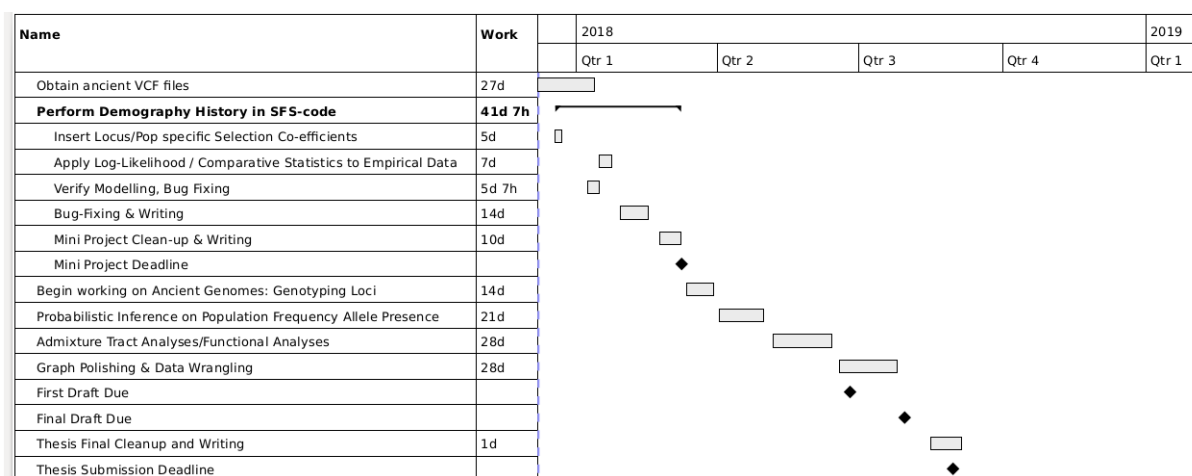
The Horn of Africa & Middle Eastern Genomic datasets from the 260k SNP dataset [Hodgson et al., 2014] will then be similarly utilised via this method. Presence or absense of predicted genotypes across the ancient genomes will then be scored for correct predictions, where a null population is direct mapping of current allelic frequencies to the ancient genome, giving us a metric against

which to validate this process. The Mota man [Llorente et al., 2015] will also be investigated via the adapted LSBL test to these analyses to identify presence of abscense of introgression, which is important in terms of intepreting predicting selection co-efficients across this ancient genome.

3 Anticipated Outcomes and Outputs

This project aims to input into a manuscript of high research-impact upon completion.

4 Project Feasibility

Name	Work	2018				2019
		Qtr 1	Qtr 2	Qtr 3	Qtr 4	Qtr 1
Obtain ancient VCF files	27d					
Perform Demography History in SFS-code	**41d 7h**					
Insert Locus/Pop specific Selection Co-efficients	5d					
Apply Log-Likelihood / Comparative Statistics to Empirical Data	7d					
Verify Modelling, Bug Fixing	5d 7h					
Bug-Fixing & Writing	14d					
Mini Project Clean-up & Writing	10d					
Mini Project Deadline						
Begin working on Ancient Genomes: Genotyping Loci	14d					
Probabilistic Inference on Population Frequency Allele Presence	21d					
Admixture Tract Analyses/Functional Analyses	28d					
Graph Polishing & Data Wrangling	28d					
First Draft Due						
Final Draft Due						
Thesis Final Cleanup and Writing	1d					
Thesis Submission Deadline						

5 Budget

Being purely computational, budget can be put towards the attendance of relevant meetings, such as to present and discuss such work. Candidates include ESHE 2018 `http://www.eshe.eu/meetings.html` A presumptial budget for attendance at this meeting follows.

Table 1: ESHE 2018 Attendance Budget

Item	Cost(GBP)
Transport	163.00
Registration	44.00
Accomodation	150.00
Poster Printing	15.00
Total	373.00

References

[Chang et al., 2015] Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation plink: rising to the challenge of larger and richer datasets. *Giga-Science*, 4(1):7.

[Gresham, 2017] Gresham, S. (2017). Post-admixture selection has played a major role in african and middle eastern modern human evolution. Master's thesis, Imperial College London.

[Hernandez and Bateman, 2008] Hernandez, R. D. and Bateman, A. (2008). A flexible forward simulator for populations subject to selection and demography. *BIOINFORMATICS APPLICATIONS NOTE*, 24(23):2786–278710.

[Hodgson et al., 2014] Hodgson, J. A., Mulligan, C. J., Al-Meeri, A., and Raaum, R. L. (2014). Early Back-to-Africa Migration into the Horn of Africa. *PLoS Genetics*, 10(6).

[Llorente et al., 2015] Llorente, M. G., Jones, E. R., Eriksson, A., Siska, V., Arthur, K., Arthur, J., Curtis, M., Stock, J. T., Coltorti, M., Pieruccini, P., et al. (2015). Ancient ethiopian genome reveals extensive eurasian admixture in eastern africa. *Science*, 350(6262):820–822.

[Skoglund et al., 2017] Skoglund, P., Thompson, J. C., Prendergast, M. E., Mittnik, A., Sirak, K., Hajdinjak, M., Salie, T., Rohland, N., Mallick, S., Peltzer, A., Heinze, A., Olalde, I., Ferry, M., Harney, E., Michel, M., Stewardson, K., Cerezo-Román, J. I., Chiumia, C., Crowther, A., Gomani-Chindebvu, E., Gidna, A. O., Grillo, K. M., Helenius, I. T., Hellenthal, G., Helm, R., Horton, M., López, S., Mabulla, A. Z., Parkington, J., Shipton, C., Thomas, M. G., Tibesasa, R., Welling, M., Hayes, V. M., Kennett, D. J., Ramesar, R., Meyer, M., Pääbo, S., Patterson, N., Morris, A. G., Boivin, N., Pinhasi, R., Krause, J., and Reich, D. (2017). Reconstructing Prehistoric African Population Structure. *Cell*, 171(1):59–71.e21.

6 Supervisor Budgetary Approval

5

"I have seen and approved the proposal and the budget"

Jason A. Hodgson

8/10/17