Salivary amylase gene copy number: Have humans adapted to high starch diets?

E. F. Caldwell, N. von Cramon-Taubadel, M.E. Weale, M.G. Thomas

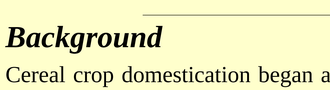
The Centre for Genetic Anthropology, University College London.

Background

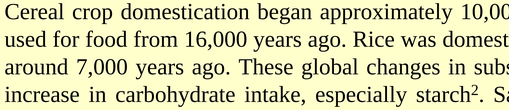
Cereal crop domestication began approximately 10,000 years ago in the Middle East[1]. Wild varieties were already being used for food from 16,000 years ago. Rice was domesticated in China and Maize was domesticated in South America, both around 7,000 years ago. These global changes in subsistence led to an overall reduction in protein and fat intake and an increase in carbohydrate intake, especially starch[2]. Salivary Amylase is an enzyme found in saliva which catalyses the initial stage of starch digestion in the mouth. Humans show variation in salivary amylase (AMY1) gene copy number - between 2 and 18 gene copies in different individuals[3]. Bank et al. (1992)[4] demonstrated that variation in AMY1 gene copy number results in differing levels of salivary amylase enzyme expression . High levels of AMY1 enzyme expression confers an increased ability to hydrolyse starch in the mouth.

AMY1 gene cluster variation

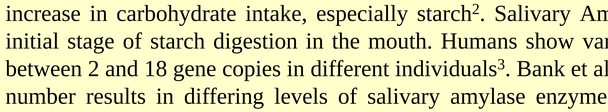
0 AMY1 cluster repeats - AMY*H0 allele



1 AMY1 cluster repeats - AMY*H1 allele



2 AMY1 cluster repeats - AMY*H2 allele



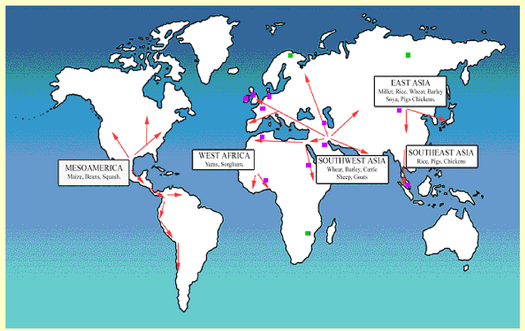


Fig 1:The origins and spread of agriculture. Agricultural populations sampled (purple) – Ireland, Britain, Germany, Ashkenazi Jews, Armenia, Kuwait, Mongolia, Algeria, Ethiopia, Nigeria. Non-agricultural or recent agricultural populations sampled (green) – Saami, Yakut and Malawi.

Hypothesis

•High AMY1 gene copy number became selectively advantageous with the adoption of high starch diets in human populations.

Prediction:

•Populations with a long tradition of high starch diets should have an increased frequency of high AMY1 repeat-number alleles when compared to hunter gatherers or those populations which have adopted cereal agriculture more recently.

•High AMY1 repeat number alleles will have gone from low to high frequency more quickly than expected for neutral alleles.

Methodology

•Two PCR based assays were designed:

1) To amplify an area around a deletion present in the AMY1 genes but not in AMY2;

2) To amplify an area around a deletion present in the AMY2 genes but not in AMY1.

•Each PCR reaction produces 2 fragments of different size, one originating from AMY1 and the other from AMY2 genes.

•ABI GeneScan technology was used to quantify the ratio of AMY1 products to AMY2.

•As the number of AMY2 genes are always constant, the AMY1: AMY2 product ratio indicated the number of AMY1 genes present.

•Experiments were carried out to assess the accuracy of the quantification method and to determine how many times the assay needed to be repeated to obtain reliable results. Each sample is PCRed twice and each PCR is run twice on a GeneScan gel.

AMY Microsatellite Protocol

•A multiplex PCR protocol[5] was designed to amplify 6 microsatellite markers closely linked to the AMY gene cluster.

•ABI GeneScan technology was used to provide accurate sizing of the products, to determine the microsatellite repeat number alleles.

Determining Phase

•In addition to using families to establish AMY1 repeat number alleles, software was developed for establishing haplotypes, as well as for the accurate prediction of AMY1 allele frequencies.

•Microsatellite haplotypes were established by typing family samples.

Identifying Selection (1)

Differences in allele frequencies among populations

The proportion of allele frequency variation due to differences between populations was estimated using F_{ST}. Directional selection on certain alleles in some, but not in other populations, will lead to an increase in F_{ST} for that locus, when compared to values for neutral loci[6] (a null-distribution). Balancing selection leads to a decrease in F_{ST} when compared with neutrally evolving loci. When compared against a published null-distribution[9] of F_{ST} between European and East Asian populations for assumed neutral markers, the highest pairwse F_{ST} obtained for the AMY1 data (Mongolia vs Saami=0.056) was not an outlier on the distribution (See Fig 3).

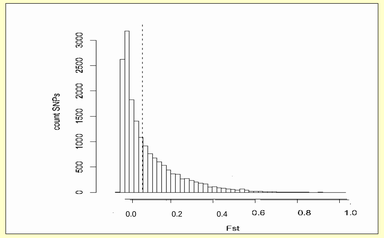


Fig 3: The distribution of F_{ST} values for 20,701 SNPs in comparisons between Europeans and East Asians. The F_{ST} for the AMY1 Mongolian vs Saami comparison is shown by the dotted line.

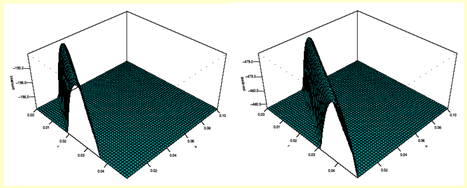


Fig 4: A plot of the likelihood of the observed the variability at the 6 microsatellites found on a background of the AMY1*H0 allele (left) and of the AMY1*H1 allele (right) in European populations, for different values of selection and population growth.

Identifying Selection (2)

Intra-allelic variability

Intra-allelic variability is the joint distribution of the frequency of an allele and the extent of variability at closely linked loci. Under the neutral model, the frequency of an allele should be related to its age, because it takes a long time for rare (including new) alleles to drift to high frequencies in populations. Therefore, if a young allele is at high frequency in a population then this suggests that selection has been operating[7]. The relative ages of different alleles can be estimated by looking at the variation in closely linked markers (e.g. microsatellites). Preliminary analysis of the AMY1 repeat alleles and variance at six closely linked microsatellites indicates that there has been some selection in the AMY1*H1 allele compared to the AMY1*H0 allele in European populations .

Chimpanzee AMY1 alleles?

The AMY1 quantification protocol was redesigned to quantify chimpanzee AMY1 variation. The 5 chimps tested were found to be homozygous for the AMY1*H0 allele. These results suggest that AMY1 gene copy number variation in humans arose after the split between humans and chimpanzees.

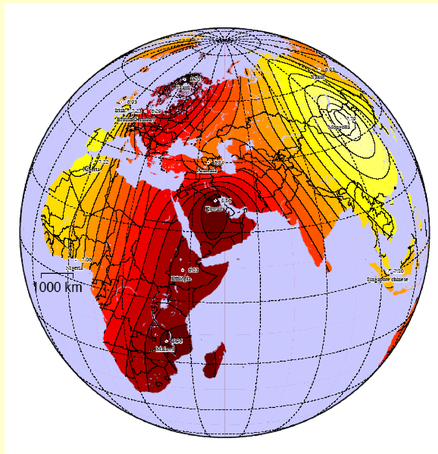


Fig 2: Geographical frequency distribution of mean number of AMY1 repeats per individual.

Summary

Based on the comparison of F_{ST} values among different populations it was not possible to identify different selection forces acting on AMY1 alleles in different populations. This could be because selective pressure has not been acting on the AMY1 alleles. Alternatively, a lack of power in the method and the data may also explain our observations. The F_{ST} method is known to be relatively insensitive. Using data from six closely linked human microsatellites in an analysis of intra-allelic variability, some evidence was found for positive selection acting on the AMY1*H1 allele in Europeans. Further work is in progress to estimate a date for the emergence of the various AMY1 repeat number alleles. The chimpanzee data suggests that the most frequent allele in humans (AMY1*H1) may not be the ancestral allele in humans as was previously assumed.

Acknowledgements

•All DNA samples were collected and extracted by The Centre for Genetic Anthropology - UCL.

•This research would not have been possible without the assistance of Dr Neil Bradman, Prof. Dallas Swallow, Dr Ben Fletcher, Prof. Levon Yepiskopsyan, Dr Ayele Tarekegn, Amanda Bradman & Abigail Jones.

•Thank you to Prof. David Goldstein for access to Singaporean Chinese family samples.

•Samples of known AMY1 genotype were kindly provided by Prof. Jan Pronk (Vrije Universiteit), Amsterdam.

•Many thanks to Dr. M. Stumpf for the development of MstMulti (interallelic variability analysis software)

•**This project is funded under the Wellcome Trust BioArchaeology Scheme.**

References

1. Moore A, (1989) The transition from foraging to farming in Southwest Asia: present problems and future directions. In Harris & Hillman (eds) *Foraging and farming: the evolution of plant exploitation.* London Unwin Hyman. Pp 620-631
2. Neel JV (1982) The thrifty genotype revisited. In Kobberling J, Tattersall R (eds) *The Genetics of Diabetes Mellitus.* London Academic Press. Pp283-293.
3. Groot P, Bleeker M, Pronk J Arwert F, Mager W, Planta R, Eriksson A, Frants R (1989) The Human alpha amylase multigene family consists of haplotypes with variable numbers of genes. *Genomics* **5**: 29-42
4. Bank R, Hettema E, Muijs M, Pals G, Arwert F, Boomsma D, Pronk J (1992) Variation in gene copy number and polymorphism of the human salivary amylase isoenzyme system in Caucasians. *Human Genetics* **89**:213-222
5. Thomas MG, Bradman N, Flinn HM (1999) High throughput analysis of 10 microsatellite markers and 11 diallelic polymorphisms in the human Y chromosome, *Human Genetics* **105**: 577-581.
6. Akey J, Zhang G, Zhang K, Jin L, Shriver M (2002) Interrogating a High-Density SNP map for signatures of natural selection. *Genome Research* **12**:1805-1814.
7. Slatkin M, Bertorelle G (2001) The use of intraallelic variability for testing neutrality and estimating population growth rate. *Genetics* **158**: 865-874.
8. Kayser M, Brauer S & Stoneking M(2003) A genome scan to detect candidate regions influenced by local natural selection in human populations. *Mol. Biol. Evol* **20**(6):893-900.
9. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schultz B, Cook L, Davenport R, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstein B, Baldwin J, Stange-Thomann N, Zody MC, Linton L, Lander ES, Altschuler D; International SNP working group (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409: 928-933.