

Population-Specific Links Between Heterozygosity and the Rate Human Microsatellite Evolution

William Amos

Received: 3 June 2010 / Accepted: 29 November 2010 / Published online: 15 December 2010
© Springer Science+Business Media, LLC 2010

Abstract Microsatellites form an abundant class of DNA sequences used widely as genetic markers. Surprisingly, the length of human microsatellites varies highly predictably with distance from Africa, apparently following the linear decline in variability that arose as we colonised the world. Such patterns have been used to argue that heterozygosity modulates the rate of microsatellite evolution. Here I test the ensuing prediction that variation in demographic history will cause individual populations predictably either to lead or to lag any given trend in length. I find that they do: larger populations with locally higher heterozygosity have microsatellites that are longer when a locus is expanding and shorter when a locus is contracting. These patterns remain even after controlling for the stepwise way in which heterozygosity and allele lengths decline across the world. This analysis provides support for a strongly discontinuous model for how human genetic variability is distributed and shows how individual populations differ in the average rate their microsatellites are evolving. Such patterns have the potential to provide a new window onto historical demography.

Keywords Microsatellite · Heterozygote instability · Human · Distance from Africa · Genetic variation

Introduction

The recent explosion of data on human polymorphisms (Rosenberg et al. 2005; The International HapMap

Consortium 2007) allows both high resolution studies of population affinities and a re-examination of assumptions about how the markers themselves evolve. One particularly interesting class of sequences is the microsatellites, short tandemly repeated motifs that exhibit high levels of length polymorphism (Goldstein and Schlötterer 1999). Although early studies suggested a rather simple model for microsatellite evolution (Shriver et al. 1993; Valdes et al. 1993), in which most mutations involved the gain or loss of a single repeat unit and as a result allele length drifted randomly up and down over time, more complicated patterns have since emerged (Brandström et al. 2008; Crawford et al. 1998; Kelkar et al. 2008; Schlötterer 2000). Most notably, most microsatellites appear subject to mutation bias in which gain in length mutations outnumber loss of length mutations or vice versa (Amos et al. 1996; Primmer et al. 1998; Vigouroux et al. 2003; Xu et al. 2000).

The widespread existence of mutation bias provides an unusual opportunity to test the constancy of the genome-wide mutation rate because, over time, variation in mutation rate becomes translated into variation in relative mean allele length (Amos et al. 2008). For example, when mutation bias favours expansion, a population with elevated mutation rate will tend to exhibit longer alleles than one where the mutation rate has not changed. Although it might be thought that the genome-wide mutation rate would be approximately constant, evidence is starting to accumulate that heterozygosity itself may influence mutation rate, with heterozygous sites being more mutable than equivalent homozygous sites, the so-called ‘heterozygote instability’ hypothesis (HI) (Amos 2010; Amos et al. 2008; Ninio 1996; Rubinsztein et al. 1995). Perhaps the best evidence for HI come from an analysis that exploited the way in which humans lost variability as they moved out of Africa to colonise the rest of the world,

W. Amos (✉)
Department of Zoology, University of Cambridge, Downing
Street, Cambridge CB2 3EJ, UK
e-mail: w.amos@zoo.cam.ac.uk

creating a well-defined gradient in which heterozygosity declines highly predictably with distance from Africa (Liu et al. 2006; Manica et al. 2007; Prugnolle et al. 2005; Ramachandran et al. 2005). Under HI, this gradient of heterozygosity should create a parallel gradient in mutation rate which, interacting with mutation bias, should lead to predictable variation in microsatellite length. In practice, relative microsatellite length is indeed strongly predicted by distance from Africa (Amos et al. 2008).

Two possible criticisms can be levelled at this test of HI. First, distance from Africa is a catchall and takes no account of the different branches of the migration that may have experienced strongly contrasting demographic histories. For example, populations approximately 7000 km from Addis Ababa may lie either in northern Europe or in southern central Asia. Second, despite efforts to control for shared ancestry statistically (Purvis and Rambaut 1995), there is always a concern that much of the predictability of microsatellite length could be due to macro-geographic patterns. Thus, the five American populations in this study are all major outliers, separated from all others by the whole of North America, yet are also related to each other. These populations therefore tend to exert a dominant influence on the shape of the relationship between distance from Africa and mean allele length.

A further potential complication arises from the recent debate as to whether human populations in different geographic regions have evolved to some extent separately. Some have argued that discontinuities in the distribution of human genetic diversity merely reflect unavoidable sampling gaps due to natural barriers such as mountain ranges and oceans (Serre and Pääbo 2004). A similar view has also been taken by many studies that have examined how variability declines with distance from Africa, where a continuous stepping stone model is used as the default (Liu et al. 2006; Prugnolle et al. 2005; Ramachandran et al. 2005). However, when data from global populations are analysed, clustering tends to reflect strongly the major continental divisions, and the level of differentiation appears to exceed the level expected simply from gaps in the sampling coverage (Rosenberg et al. 2002, 2005). Interestingly, the existence of HI might help to resolve these issues because historical regional isolation would tend to have created regional differences in demographic history that in turn could be translated into regional differences in the relationship between microsatellite length and heterozygosity.

To determine whether individual populations and geographic regions show evidence of heterozygosity-associated variation in their microsatellite dynamics I revisited the large dataset published by Rosenberg and colleagues (Rosenberg et al. 2002, 2005), involving 783 microsatellites genotyped in 1048 samples from 53 worldwide

populations in the Human Diversity Cells Line Panel (Cann et al. 2002). I tested the prediction from HI that, after subtracting macrogeographic trends in the way mean allele length varies across the world, any residual variation in length will not be random, but will instead correlate with modern population size. I show that residual mean allele length is indeed non-randomly distributed, and that individual populations and regions both carry diagnostic signals that appear consistent with the idea that mutation rate is modulated by heterozygosity.

Data and Rationale

Data for 783 microsatellites genotyped in 1048 individuals drawn from 53 worldwide populations were downloaded from <http://rosenberglab.bioinformatics.med.umich.edu/diversity.html> (Rosenberg et al. 2002, 2005). Populations were classified as belonging to seven main geographic regions: Europe (black triangle, $n = 8$), Middle East (grey circles, $n = 4$), Central Southern Asia (open circles, $n = 8$), Oceania (open triangles, $n = 2$), America (black diamonds, $n = 5$), Africa (black circles, $n = 7$) and East Asia (grey triangles, $n = 19$), following Rosenberg et al. (2002). All allele lengths were converted from amplified fragment length to repeat numbers by reference to the original clone sequence and the published primer sets using a custom Visual Basic macro (Amos and Hoffman 2010). Distance from Africa was calculated as the land-only distance from Addis Ababa (Prugnolle et al. 2005) and kindly provided by A. Manica (pers. com.). For consistency, all population size estimates were taken from the Joshua project website <http://www.joshuaproject.net> and no attempt was made to correct for population sub-structure: if samples were described as being collected from country X we use the total population size of country X, while if the samples are described as being collected from region Y in country X the population size of region Y was used.

Under the null model of microsatellite evolution, changes in mean allele length (=‘MAL’) at a locus within a particular population occur purely through a combination of genetic drift and a constant mutation rate (which may or may not be biased), and all variation in MAL should be attributable to a combination of stochasticity and non-independence due to shared coancestry. Under this model, there is no mechanism capable of causing predictable differences in MAL between any given pair of populations. Put another way, MAL at any given locus should be as likely to be greater in population A relative to population B as it is to be greater in population B. In contrast, under HI, trends in MAL between populations will be non-random, accruing through an interaction between mutation bias and variation in mutation rate caused by demographically

induced changes in genome-wide heterozygosity (Amos 2010; Amos et al. 2008). Thus, if two populations differ in their genome-wide heterozygosity, HI should cause the population with higher heterozygosity to have a higher microsatellite mutation rate. Consequently, when a locus is expanding, HI should cause the more heterozygous population to carry predictably longer alleles, while when a locus is contracting, the alleles will be predictably shorter.

To test this prediction, I first sought to remove broad-scale variation in allele length. For this I analysed each of the 783 loci, to each fitting a second order polynomial to the relationship between MAL in each of 53 populations and distance from Africa. The rationale here is that while at many loci MAL varies in a strongly non-linear way across the world, appearing variously ‘humped’ or ‘dipped’ and making a simple linear regression inappropriate, very few loci reveal patterns that would be captured better by a third or higher order polynomial. In each population I then calculate both residual length about the fitted line and the instantaneous slope of the line at the appropriate distance from Africa. Under HI, these two quantities should be correlated (see Fig. 1). Specifically, if a locus is expanding, the greater heterozygosity nearer to Africa should mean populations nearer to Africa have greater MAL than more

distant populations, and the local slope will be negative (local slope indicates the direction of evolution). At the same time, the same microsatellite in an unusually large population will be evolving even faster, and should tend to have a large residual length. Conversely, a positive slope would indicate a locus that is contracting, and the same large population would this time have a large negative residual. Unusually small populations would show the converse pattern, tending to lag rather than lead the trend. By contrast, if HI does not operate, there is no a priori expectation that populations nearer to Africa will have accumulated more microsatellite mutations than populations that are more distant. Consequently, a negative slope should be as likely to indicate a locus that has contracted more in populations further from Africa as one that has expanded more in populations lying closer to Africa. If slope is uninformative about the direction of evolution, then even if microsatellites in particular populations were evolving at different rates, there is no obvious way by which local slope and residual MAL can become correlated.

Results

Do Individual Populations Differ In Their Relative Rates of Microsatellite Evolution?

To test whether residual allele length varies randomly, for each locus I fitted second order polynomial of MAL against distance from Africa and then, for each population, calculated both residual MAL about the fitted line and the instantaneous slope. For each population I then calculated the rank correlation between these two quantities, slope and residual length, across all 783 microsatellites. Under the null model, a correlation should not exist. However, under the HI hypothesis, a correlation could exist in some or many populations if two key conditions are met. First, slope must indicate the direction of any mutation bias (and hence direction of evolution), as should be true if the higher heterozygosity towards Africa drives a higher mutation rate, such that loci that are contracting become shorter near to Africa and those that are expanding become longer near to Africa. Second, populations must vary in their genome-wide mutation rates above and beyond the simple trend expected from the way heterozygosity declines with distance from Africa. This second property could be satisfied if the differences in demographic history of the various populations since leaving Africa are sufficient to drive detectable differences in heterozygosity and hence mutation rate. In practice, 42 of 53 populations exhibit significant rank correlations [$=$ “RC”, $|r_s| > 0.07$, $n = 783$ loci, $P < 0.05$; observed range $r_s = -0.467$

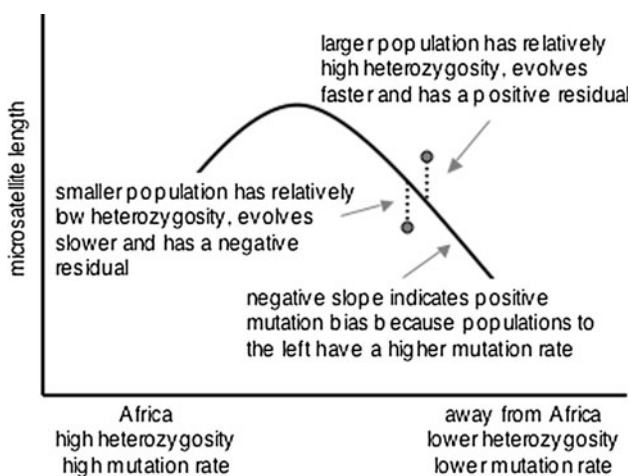


Fig. 1 Predicted impact of heterozygote instability (HI) on mean allele length in large and small populations relative to a general global trend. Under HI, the greater heterozygosity in Africa causes generally higher mutation rates compared with microsatellites in non-Africa populations. The *solid line* summarises a fitted relationship of how mean allele length at a particular microsatellite varies across the world. Since populations nearer to Africa evolve faster, a negative slope indicates positive mutation bias (microsatellite is longer nearer to Africa) and vice versa. Relatively large/small populations will be ahead of/lag behind the local trend, respectively. Large populations will therefore have positive residuals around negative slopes and negative residuals around positive slopes. Across many loci with diverse slopes, large populations should therefore exhibit a negative correlation between local slope and residual mean allele length. The converse is true of small populations

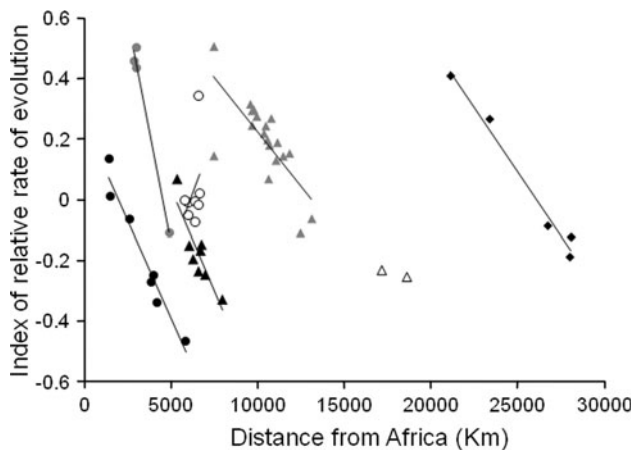


Fig. 2 Global variation in the population-specific rate of microsatellite evolution. Data are from 783 microsatellites genotypes in 1048 samples drawn from 53 populations representing seven main geographic regions: Europe (*black triangle*, $n = 8$), Middle East (*grey circles*, $n = 4$), Central Southern Asia (*open circles*, $n = 8$), Oceania (*open triangles*, $n = 2$), America (*black diamonds*, $n = 5$), Africa (*black circles*, $n = 7$) and East Asia (*grey triangles*, $n = 19$) (Rosenberg et al. 2005). For each population, the relative rate of microsatellite evolution (tendency for a population to carry longer alleles when a locus is expanding and shorter alleles when a locus is contracting) is estimated as the rank correlation between the instantaneous slopes of the relationship between mean allele length and distance from Africa, and their associated residuals. Assuming evolution occurs fastest in Africa, lagging and leading populations will yield positive and negative correlations, respectively. Thus positive values indicate slower evolution and negative values indicate faster evolution. Here and elsewhere, distance from Africa is taken as land-only distance to Addis Ababa

(Mandenka, $P < 10^{-42}$) to $r_s = 0.507$ (Xibo, $P < 10^{-52}$]. Thus, the residuals are highly non-random, with some populations consistently tending to be ahead of the trend and others consistently lagging. Interestingly, within each of the primary population groups, the correlation coefficients tend to decline with increasing distance from Africa, and in most cases the decline is significant (Fig. 2). The primary exception is central southern Asia, where all the populations are tightly clustered except for a large outlier, the Hazara, which appears to sit more comfortably with the East Asia populations.

The data in Fig. 2 appear to go *against* the trend predicted by HI. If high heterozygosity accelerates microsatellite evolution, the RC values should become more positive with increasing distance from Africa, not, as seen, more negative. However, the way the populations exhibit trends within but generally not between the main continental groupings suggests that these groupings have functional significance, and that each region is evolving to some degree separately (Rosenberg et al. 2005), raising the question of whether the decline in heterozygosity is indeed perfectly linear. To test this assumption, I extracted the

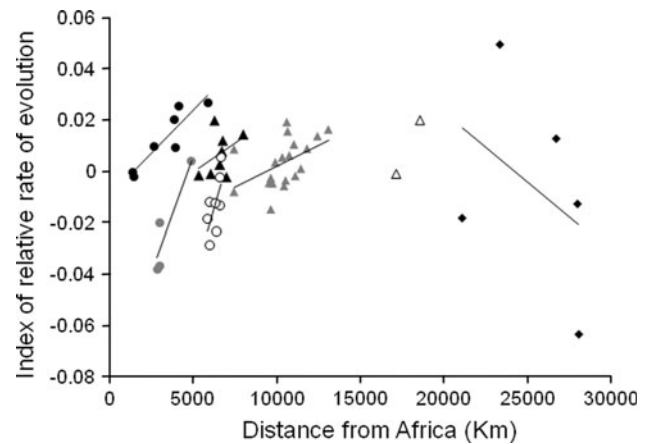


Fig. 3 Regional variation in residual heterozygosity after accounting for the global trend. Residuals are taken about a linear trend fitted to mean observed heterozygosity. Data and symbols are given in the legend to Fig. 1

residuals from the global linear trend between heterozygosity and distance from Africa, and plotted these against distance from Africa (Fig. 3). I find a previously undescribed pattern in which residual heterozygosity within each group tends to increase with distance from Africa. The explanation appears simple. Although the continental groups vary in mean heterozygosity in a way that declines remarkably smoothly with mean distance from Africa ($r^2 = 0.971$, $n = 7$ groups, $r^2 = 0.99$ excluding the two Oceania populations) within each group the relationship between heterozygosity and distance from Africa is almost perfectly flat (data not shown). The only exception is the American group which shows a downward trend, but there is a lot of scatter and the relationship is non-significant. Clearly, within each region, if heterozygosity varies little, residuals about a fitted slope will tend to show the converse trend.

Does Heterozygosity Drive Population-Specific Microsatellite Evolution?

The above analysis suggests that populations differ in their genome-wide average mutation rates but also that it is dangerous to assume that global trends apply to populations within a given geographic region. Indeed, such an assumption can generate spurious and misleading trends. Consequently, I recalculated the RC values but this time, instead of using residual allele length based on a second order polynomial fitted to all 53 global populations (X -axis = distance from Africa, Y -axis = MAL), I used residual length based on a seven separate linear regressions (axes as previously), each fitted to all populations within one of the main geographic regions (defined in Methods). As before, I then looked for a correlation in each

population across all 783 loci between local slope and residual length. When this is done, significance is generally reduced compared with when a single second order polynomial is fitted. However, 24 populations still yield significant rank correlations, and many of these are highly significant [in all cases, $n = 783$ loci, observed range $r_s = -0.34$ (Uygur, $P < 10^{-22}$) to $r_s = 0.305$ (Hazara, $P < 10^{-17}$)]. In this particular analysis, for the instantaneous slope of the relationship between allele length and distance from Africa I used slopes based on the global pattern, allowing some degree of non-linearity. However, very similar results are obtained if the locally fitted linear slopes are substituted.

The analysis above indicates both that the slope of the relationship between MAL and distance from Africa is indicative of the direction of evolution of a microsatellite, and that populations within a region differ in the average rate at which their microsatellites appear to be evolving. The HI hypothesis makes a clear prediction that these differences should reflect differences in heterozygosity. Therefore, I next explored the relationship between the calculated RC values and two quantities that should reflect/measure relative heterozygosity among populations within a region (a) log (modern population size) and (b) residual genome-wide heterozygosity, calculated by fitting separate linear regressions within each region for mean heterozygosity across all 783 markers. In particular, we expect populations in which microsatellites appear to be evolving faster than average to be larger and to have higher heterozygosity. Such populations should have negative RC values (positive residuals when a locus is expanding, indicated by a negative slope of MAL against distance from Africa, and negative residuals when a locus is contracting, indicated by a positive slope). In practice, our crude estimates of modern population size are reasonable predictors of local residual heterozygosity ($r^2 = 0.22$, $n = 53$, $P = 0.0004$), and both quantities also exhibit significant negative relationships with RC [log (modern population size) vs RC, $r^2 = 0.089$, $n = 53$, $P = 0.03$; local residual heterozygosity vs RC, $r^2 = 0.14$, $n = 53$, $P = 0.006$]. Furthermore, given the limited time available for mutation-driven microsatellite evolution, one would expect stronger relationships to be found at faster compared with more slowly evolving loci, and this appears to be the case. If only the least variable loci are used, those with average global heterozygosity <0.6 ($n = 77$ loci), the significance of the relationship between RC and log (modern population size) is lost ($r^2 = 0.019$, n.s.), while if only the most variable loci are used, those with mean heterozygosity >0.8 ($n = 56$ loci), the relationship strengthens noticeably (Fig. 4, $r^2 = 0.28$, $P = 5.3 \times 10^{-5}$). Thus, larger modern populations have locally higher heterozygosity and their microsatellites tend to be evolving relatively faster than average.

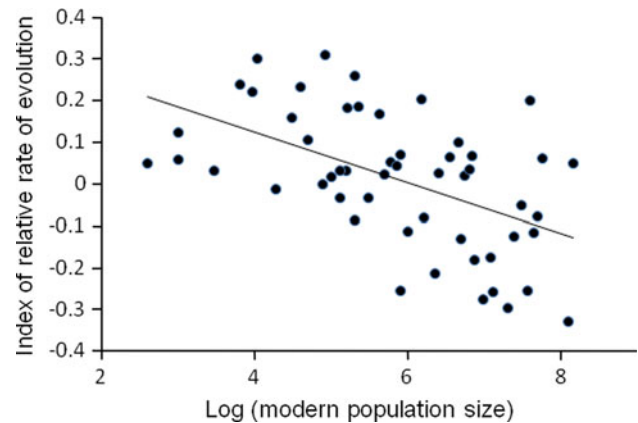


Fig. 4 Relationship between log(modern population size) and a measure of the relative rate of microsatellite evolution for 53 worldwide populations. Estimates of modern population size are taken directly from the Joshua project website and do not correct for population structure or other complexities. Relative rate of microsatellite evolution is as described in the legend to Fig. 2. Note, positive values indicate slower evolution and conversely. Only loci with worldwide average heterozygosity >0.8 are used

Discussion

Simple models of microsatellite evolution may by chance yield patterns in which allele length at a locus increases or decreases across the world, but lack any obvious mechanism capable of causing systematic variation in the average size or direction of residuals associated with particular populations across many unlinked loci. I show that residual allele length is highly non-random, and is instead strongly predicted by the local direction of the slope of the relationship between mean allele length and distances from Africa across 53 worldwide populations. Moreover, different populations give signatures, some indicative of microsatellites that are evolving faster than average, and some where evolution is slower. Modern population size and excess heterozygosity both predict which populations are evolving faster/slower in a way that is consistent with a model where mutation rate is modulated by heterozygosity.

Large datasets such as those published by Rosenberg and colleagues (Rosenberg et al. 2002, 2005) are remarkably informative about microsatellite evolution because of the way averaging across literally hundreds of loci can reveal subtle patterns that would otherwise be masked by the stochasticity of sampling and drift. When this is done, surprisingly complicated patterns are laid bare (Pemberton et al. 2009), particularly when the data are plotted according to distance from Africa (Amos et al. 2008; Prugnolle et al. 2005; Ramachandran et al. 2005). Previous studies have shown that both mean allele length and heterozygosity are both strongly predicted by distance from Africa (Amos et al. 2008). I now confirm that the

microsatellite length—distance from Africa trends are not random because individual populations have highly coloured residuals: in some cases residual length being predictably positive when overall allele length is increasing with distance from Africa and predictably negative when allele length is decreasing; while in others the reverse patterns hold. Quantified in terms of the rank correlations (RC) between local slope and residual mean allele length, these trends imply two things. First, the local slope of the fitted relationship of allele length against distance from Africa is indicative of whether allele length is, at that point in the world, either increasing or decreasing. If not, it is difficult to conceive how a relationship between slope and residual direction could arise. Second, some populations are predictably ahead of the overall trend, while others are predictably lagging.

The initial results appeared rather striking and clear-cut, with each of the major population groups showing a separate trend in which RC values became more negative with increasing distance from Africa. Such clustering lends strong support to the notion that these seven continental regions harbour fundamental differences in the way their genetic diversity is distributed (Rosenberg et al. 2005). In other words, the assumption that allele length and heterozygosity both vary more or less continuously across the world (Liu et al. 2006; Prugnolle et al. 2005; Ramachandran et al. 2005) appears false. Instead, heterozygosity declines in a series of steps, each corresponding roughly to one of the main population groupings. Within each group, the relationship between heterozygosity and distance from Africa is effectively flat. The trends for allele length are somewhat more complicated, in that the slopes of local regressions (within each geographic area) tend to correlate with the slopes predicted by the global pattern (data not shown). One can imagine that the relative homogeneity within regions could be due either to initial gradients being subsequently eroded by mixing and movement within each region, or that gradients within a region never formed because each region was colonised in a single founder event from which local colonisation occurred without much further loss of variability. Distinguishing between these models will require further work.

Having recalculated the RC values, this time using trends in allele length fitted separately to populations within each of the seven geographic regions, I find that there remains strong evidence of non-randomness in the residuals. Almost half the populations carry microsatellites that consistently either lag or lead the overall trend. A relationship between local slope and residual length is consistent with the heterozygote instability hypothesis, HI, because it implies a driven, non-random process governed by some inherent characteristic of the population, beyond simple distance from Africa. However, on its own, a link

between residual length and local slope does not relate directly to heterozygosity. I therefore examined how RC values vary with local residual heterozygosity and modern population size. I find that modern population size predicts local residual heterozygosity (the extent to which, within each region, a population has higher than average genome-wide heterozygosity), and that higher heterozygosity and higher population size both predict accelerated microsatellite evolution. Furthermore, this relationship is strongest among microsatellites with highest average heterozygosity, as would be expected of a process that reflects recent variation in mutation rate. Importantly, a correlation between RC and population size would be unlikely if significant RC values arose purely through undetected heterogeneities such as the original assumption that heterozygosity declines in a single, continuous gradient or, perhaps, that movement of populations has caused some to be classified as nearer or further from Africa than they 'should' be. Moreover, it is worth remembering that the estimates of modern population size are inevitably extremely crude, taking no account of either population substructure, historical mixing or the rate of recent population expansion. As such, their correlation with effective population size will be weak, implying that if better estimates could be obtained, the relationship with RC could be substantially stronger.

Despite the above, HI remains a hypothesis in need of further testing. However, if it does prove correct, the slope-residual length correlations should be informative about population history. Specifically, populations with negative RC values, indicative of accelerated microsatellite evolution, should be found to have had demographic histories that acted to increase heterozygosity, either through population expansion or through mixing with other populations. Equally, populations with positive RC values would be viewed as having had reduced heterozygosity through either founder effects of having remained persistently small. Importantly, since changes in microsatellite length take time to accumulate, these trends inform on historical demography more than present-day measures of population size, and therefore may provide anthropologists with an extra tool for elucidating the events that shaped early modern human evolution.

In summary, this analysis emphasises that we still have much to learn about microsatellite evolution but at the same time illustrates how cryptic but strong signals can be extracted from large datasets. On the one hand, my results provide further support for the notion of heterozygote instability. On the other, the clustering of populations reveals both strong discontinuities between geographic regions and also the presence of outlier populations that may be informative about these population's origins.

References

- Amos W (2010) Heterozygosity and mutation rate: evidence for an interaction and its implications. *BioEssays* 32:82–90
- Amos W, Hoffman JI (2010) Evidence that two main bottleneck events shaped modern human genetic diversity. *Proc R Soc Lond B* 277:131–137
- Amos W, Sawcer SJ, Feakes R, Rubinsztein DC (1996) Microsatellites show mutational bias and heterozygote instability. *Nat Genet* 13:390–391
- Amos W, Flint J, Xu X (2008) Heterozygosity increases microsatellite mutation rate, linking it to demographic history. *BMC Genet* 9:72
- Brandström M, Bagshaw AT, Gemmell NJ, Ellegren H (2008) The relationship between microsatellite polymorphism and recombination hot spots in the human genome. *Mol Biol Evol* 25:2579–2587
- Cann HL, de Toma C, Cazes L, FLeGrand MF, Morel V, Piouffre L, Bodmer J, Bodmer WF, Bonne-Tamir B, Cambon-Thomsen A, Chen Z, Chu J, Carcassi C, Contu L, Du R, Excoffier L, Ferrara GB, Friedlander JS, Groot H, Gurwitz D, Jenkins T, Herrera RJ, Huang X, Kidd J, Kidd KK, Langaney A, Lin AA, Mehdi SQ, Parham P, Piazza A, Pistillo MP, Qian Y, Shu Q, Xu J, Zhu S, Weber JL, Greely HT, Feldman MW, Thomas G, Dausset J, Cavalli-Sforza LL (2002) A human diversity cell line panel. *Science* 296:261–262
- Crawford A, Knappes SM, Paterson KA, de Gotari MJ, Dodds KG, Freking RT, Stone RT, Beattie CW (1998) Microsatellite evolution: testing the ascertainment bias hypothesis. *J Mol Evol* 46:256–260
- Goldstein DB, Schlötterer C (1999) Microsatellites: evolution and applications. Oxford University Press, Oxford
- Kelkar YD, Tyekucheva S, Chlaromonte F, Makova K (2008) The genome-wide determinants of human and chimpanzee microsatellite evolution. *Genome Res* 18:30–38
- Liu H, Prugnolle F, Manica A, Balloux F (2006) A geographically explicit genetic model of worldwide human-settlement history. *Am J Hum Genet* 79:230–237
- Manica A, Amos W, Balloux F (2007) The effect of ancient population bottlenecks on human phenotypes. *Nature* 448:346–348
- Ninio J (1996) Gene conversion as a focusing mechanism for correlated mutations: a hypothesis. *Mol Gen Genet* 251:503–508
- Pemberton TJ, Sandefur CI, Jakobsson M, Rosenberg NA (2009) Sequence determinants of human microsatellite variability. *BMC Genomics* 10:612
- Primmer CR, Saino N, Møller AP, Ellegren G (1998) Unravelling the process of microsatellite evolution through analysis of germ line mutations in barn swallows *Hirundo rustica*. *Mol Biol Evol* 15:1047–1054
- Prugnolle F, Manica A, Balloux F (2005) Geography predicts neutral genetic diversity of human populations. *Curr Biol* 15:R159–R160
- Purvis A, Rambaut A (1995) Comparative analysis by independent contrasts (CAIC): an Apple Macintosh application for analysing comparative data. *Comput Appl Biosci* 11:247–251
- Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, Cavalli-Sforza LL (2005) Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc Natl Acad Sci USA* 102:15942–15947
- Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW (2002) Genetic structure of human populations. *Science* 298:2381–2385
- Rosenberg NA, Maghajan S, Ramachandran S, Zhao C, Pritchard JK, Feldman MW (2005) Clines, clusters, and the effect of study design on the inference of human population structure. *PLoS Genet* 1:660–671
- Rubinsztein DC, Amos W, Leggo J, Goodburn S, Jain S, Li SH, Margolis RL, Ross CA, Ferguson-Smith M (1995) Microsatellites are generally longer in humans compared to their homologues in non-human primates: evidence for directional evolution at microsatellite loci. *Nat Genet* 10:337–343
- Schlötterer C (2000) Evolutionary dynamics of microsatellite DNA. *Chromosoma* 109:365–371
- Serre D, Pääbo S (2004) Evidence for gradients of human diversity within and among continents. *Genome Res* 14:1679–1685
- Shriver MD, Jin L, Chakraborty R, Boerwinkle E (1993) VNTR allele frequency distributions under the stepwise mutation model: a computer simulation approach. *Genetics* 134:983–993
- The International HapMap Consortium (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449:851–861
- Valdes AM, Slatkin M, Freimer NB (1993) Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics* 133:737–749
- Vigouroux Y, Matsuoka Y, Doebley J (2003) Directional evolution for microsatellite size in maize. *Mol Biol Evol* 20:1480–1483
- Xu X, Peng M, Fang Z, Xu X (2000) The direction of microsatellite mutations is dependent upon allele length. *Nat Genet* 24:396–399