Research Article

Unexpected D-tour Ahead: Why the D-Statistic, applied to Humans, Measures Mutation Rate Variation not Neanderthal Introgression

William Amos[1*] and Eran Elhaik[2]

[1] Cambridge University, Department of Zoology, Cambridge CB2 3EJ, UK

[2] Lund University, Department of Biology, Lund, Sweden, 22362

* Please address all correspondence to William Amos at w.amos@zoo.cam.ac.uk

Keywords: D-statistic, Introgression, Mutation Rate Variation, Recurrent Mutations, Incomplete Lineage Sorting (ILS), Neanderthals

Abstract

It is widely accepted that humans interbred with Neanderthals and other extinct hominins, leaving a lasting genetic legacy. However, much of the supporting evidence was developed using the statistic D, which assumes, without testing, both that mutation rate is constant and that recurrent mutations are vanishingly rare. These assumptions together preclude an alternative explanation based on variation in mutation rates across human populations. Here we critically evaluate the assumptions underlying D and confirm that neither is valid. Over 40% of SNPs in dbSNP carry recurrent mutations. Theory indicates that D does not vary with mutation rate as long as the mutation rate does not vary between populations. In practice, D calculated separately for different sequence motifs varies greatly, implying strongly that mutation rates do vary between populations. We show that most, if not all, D-informative sites result from two mutations rather than the one mutation expected under the introgression hypothesis. Moreover, individual non-Africans carry a signal in more than five times as many genomic windows as can be accounted for by the 2% legacy they are thought to carry, indicating a signal that is radically more diffuse than expected. Remarkably, partitioning the data by whether the chimpanzee or Neanderthal allele is the major allele in humans reveals that the overall reported D-value of ~5% actually comprises two opposing components: one with D ~ 30% and another with D ~ -25%. Tellingly, the positive component is produced by sites where the Neanderthal allele is the *major* allele, the exact opposite of what should be the case under introgression, where introgressed alleles should be rare. We show further that the entire D signal can be accounted for by sites where the Neanderthal allele is fixed outside Africa and the chimpanzee allele is rare inside Africa. Investigating potential mechanisms, we extend the published observation that the mutability of three-base combinations across human populations is influenced by flanking sequence heterozygosity to reveal how genomic regions that lost more heterozygosity out of Africa exhibit higher D-values. This correlation supports a model where loss of heterozygosity slowed the mutation rate, thereby reducing the divergence between Neanderthals and non-Africans. Across independent tests, our findings consistently indicate that the mutation rate variation hypothesis provides a more compelling explanation for the observed patterns in human-Neanderthal genetic relationships than the introgression hypothesis. We argue that the mutation rate variation hypothesis would help settle a number of conflicting patterns in the literature and, hence, that the concept of archaic introgression into humans and its implications for hominin-derived traits warrants reconsideration.

Introduction

The idea that humans outside Africa carry a small but ubiquitous legacy that results from interbreeding with Neanderthals sometime around 40 – 50,000 years ago was first proposed 15 years ago when the draft sequence of the Neanderthal genome was published (Green et al. 2010) and quickly became accepted as fact (Prufer et al. 2014; Sankararaman et al. 2014; Skov et al. 2020). Based on these findings, a wide range of human traits, from adaptations to high altitudes in Tibetans (e.g., Huerta-Sánchez et al. 2014) to anthropometric traits such as lip thickness (Bonfante et al. 2021) and sitting height (Dannemann and Kelso 2017), as well as immune-related phenotypes, including major genetic risk factors for severe COVID-19 (Zeberg and Pääbo 2020), have been claimed to be inherited from Neanderthals or Denisovans. A variety of approaches have been used to evaluate interbreeding, including haplotype-based methods and Bayesian and likelihood maximization of parameterized demographic models (review by Gopalan et al. 2021). However, the most prominent approach is based on summary statistics, which estimate the extent of mixing between species (Durand et al. 2011; Rogers and Bohlender 2014; Peter 2016).

The most widely used summary statistic is the so-called ABBA-BABA test and its associated statistic D (Green et al. 2010; Durand et al. 2011), which has been cited by over 2,300 studies (Google Scholar, see Methods). Durand and colleagues (2011) derived the analytical expectation of D, considering factors such as population sizes, the timing of the human - Neanderthal and African – non-African splits, and the timing and amount of introgression into non-Africans. The D-statistic is applied to four-way alignments involving four taxa ((P1,P2), P3, O): typically, two populations of interest P1 and P2 (e.g., two humans), a test species P3 (e.g., Neanderthals), and an outgroup O (e.g., chimpanzees), arranged in that order. Informative sites are bi-allelic, with the alleles labeled as A and B, where the chimpanzee is always A, the Neanderthal is B, and the two humans carry different alleles, giving the two states, named ABBA and BABA, depending on which human carries the Neanderthal B. D is calculated as the difference between ABBA and BABA counts, normalised by the total number of sites: (ABBA-BABA)/(ABBA+BABA). This ratio quantifies how often (presumed) derived alleles in the Neanderthal are shared between each of the two humans, with any asymmetry in counts being taken as indicative of introgression into the human population that shares more derived alleles with the test species. As such, introgression has been inferred between non-Africans and Neanderthals as well as between Denisovans and East Asians. For example, a D(African, European, Neanderthal, chimpanzee) of about 5% was interpreted as Europeans carrying an average of around 1-2% Neanderthal DNA. Using East Asians instead of Europeans yielded a slightly higher value and an even higher value for D(African, Papuan, Denisovan, chimpanzee) was used to argue for a large Denisovan legacy in Oceanians (Meyer et al. 2012; Jacobs et al. 2019).

There is no dispute that non-Africans are genetically closer to Neanderthals compared with Africans, regardless of how this is quantified. However, this pattern can arise in two different ways (Figure 1). The classic explanation is that the genetic distance between Neanderthals and non-Africans was *reduced* by introgression. Aternatively, the genetic distance between Neanderthals and Africans could have been *increased* by a relatively higher mutation rate in Africans. We refer to the latter as the mutation rate variation hypothesis (MRVH). Effectively every study reporting introgression into humans assumes both that recurrent mutations can be ignored and that mutation rate is constant. These assumptions are usually explicit (Green et al. 2010; Durand et al. 2011; Sankararaman et al. 2016) but are sometimes merely implicit in the definition of derived alleles (the term ‘derived’ becomes ambiguous when more than one mutation has occurred at a site). Consequently, the MRVH is standardly rejected without testing, thereby forcing the conclusion that introgression has occurred. Mutation rate constancy is

assumed so widely throughout population genetic theory that it is easy to forget that it has no empirical basis. Indeed, where evidence for mutation rate variation has been looked for, it has often been found (Kong et al. 2012), including among primate species (Brown et al. 1982; Chintalapati and Moorjani 2020) between human populations (Amos 2013; Mallick 2016; Harris and Pritchard 2017) and, massively, along a chromosome (Rogozin and Pavlov 2003; Nesta, Tafur, and Beck 2021). Moreover, in the few cases where mutation rate variation among populations is considered alongside introgression as competing explanations for why non-Africans are closer to Neanderthals, the data appear to fit the MRVH better than introgression (Amos 2020; Amos 2021).

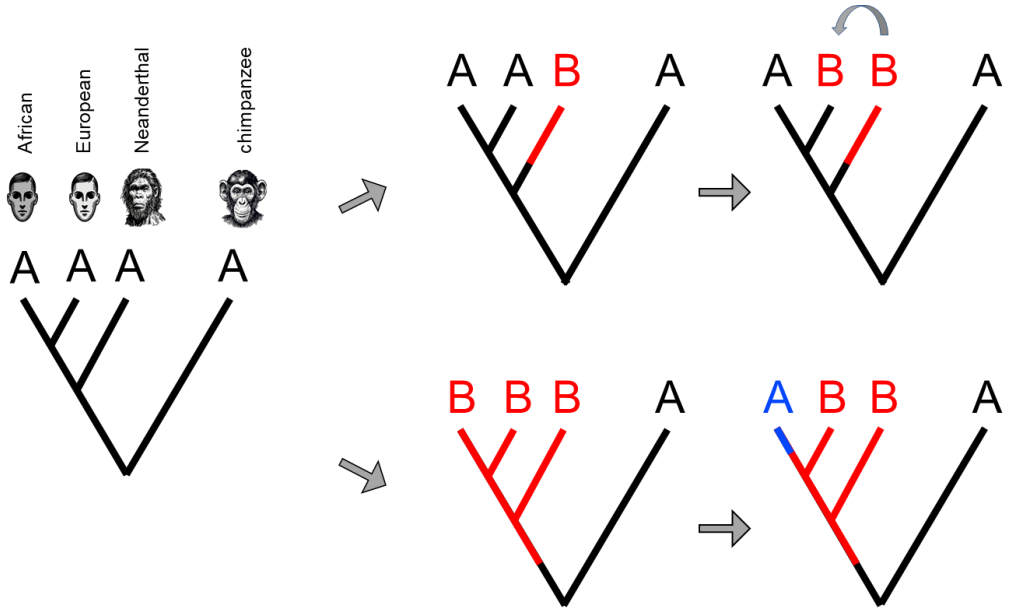

Figure 1. Schematic diagram comparing the two hypotheses, starting with the ancestral state where all taxa carry A. Under introgression (top row), a derived mutation in Neanderthals enters non-Africans (represented by Europeans) via inter-breeding. Under mutation rate variation (bottom row), a first mutation creates the common BBBA state then, in humans, recurrent B->A mutations create ABBAs and BABAs but these occur at a higher rate in Africans than in non-Africans. Lineages carrying the first mutation are in red and those carrying the second mutation are in blue. Branch lengths are not drawn to scale.

Analytical expressions have been derived for the expected numbers of ABBAs and BABAs (Durand et al. 2011). With the assumption that mutation rate is constant, in the final expression for the expectation of D, the mutation rate terms cancel. Thus, in any given pairwise population comparison, D is not expected to vary regardless of the mutability of the sites used to calculate it. It has also been assumed that each polymorphism is created by a single mutation and that recurrent mutations are rare enough to be ignored. Under these assumptions, variation in mutation rate will not drive variation in D. Rather, it will act merely to modulate the density of derived variants along a chromosome, with higher mutation rates increasing both the number of informative introgressed alleles that may be detected and also the number of sites subject to incomplete lineage sorting (ILS). ILS refers to a state where the root of the human gene tree predates the human-Neanderthal split, allowing one or more human lineages to root within Neanderthals rather than humans. More importantly, under these assumptions, introgression is the only possible mechanism capable of creating asymmetric ABBA and BABA counts.

Until the MRVH is formally tested and rejected, it is incorrect to consider that introgression has been proved. Interestingly, published data already show that both the critical assumptions that

underpin D are contravened. Thus, databases of human variation (e.g., dbSNP) reveal that an appreciable proportion of SNPs carry three or even four alleles, showing that recurrent mutations are far from vanishingly rare within humans (e.g., Phillips et al. 2020). Moreover, the frequency of recurrent mutations is expected to be appreciably larger if 'silent' mutations are included, i.e., recurrent mutations where the same alleles are created twice independently and over the greater evolutionary timescale spanned by the human-hominin-chimpanzee clade. Similarly, certain three-base combinations have highly significantly different mutation rates in different human populations, most notably between Africans and non-Africans, though also between other geographic regions (Harris 2015; Harris and Pritchard 2017). Any triplet where the mutation rate differs between Africans and non-Africans has the potential to generate non-zero D, even in the absence of introgression.

In view of these clear contraventions of the assumptions that underpin D, we revisited the D statistic and explored more rigorously the extent to which its assumptions are violated and the possible impact this might have on the interpretation of D. For the first time, to the best of our knowledge, we compare the mutation rate variation and introgression hypotheses to test which mechanism better explains the genetic distances between humans and Neanderthals. We use published, high-coverage genome sequences to conduct a series of tests designed to expose any possible influence that mutation rate variation might have on rates of allele sharing between the taxa of interest. We find that, time and time again, the MRVH fits the data better, often to the complete or near-complete exclusion of a signal due to introgression. Given the complexity of our analyses, we summarise the expectations of the two hypotheses and our findings in Table 1 below.

	Introgression Hypothesis	**Mutation Rate Variation Hypothesis (MRVH)**	**Finding**
Cause of variation	Interbreeding between populations or species	Variability in mutation rates among human populations.	
Proximate mechanism driving positive D values	Genetic variants were introduced into human populations when humans and Neanderthals interbred.	A higher mutation rate caused some populations to diverge faster from related taxa such as Neanderthals.	
Underlying assumptions	Among human populations, the mutation rate is constant, and recurrent mutations are so rare that their potential impact can be ignored.	Among human populations, the mutation rate is variable, and recurrent mutations are common enough to influence D.	Recent human mutation rates are generally higher inside Africa compared with outside, and recurrent mutations are found at almost 40% of SNPs.
Number of mutations per informative site as reflected in the transition: transversion ratio (TS:TV)	All sites contributing to D are created by a single formative mutation, regardless of whether they reflect an introgressed allele or were generated by incomplete lineage sorting.	Recurrent mutations are required and these scale with mutation rate squared, predicting a TS:TV ratio of ~4.	Sites contributing to D have a TS:TV ratio of ~2.3, consistent with a mix of 90% of sites due to ILS with with a single

	TS:TV ratio will be ~2.		mutation plus 10% of sites carrying recurrent mutations.
Distribution of D-asymmetric sites across the genome	Highly clustered, being restricted *within a single individual* to the 1-2% of the genome thought to be inherited from Neanderthals.	Uniformly distributed across the human genome, though locally concentrated in mutation hotspots and at more mutable sites.	We estimate conservatively that more than 11% of the genome of a non-African individual carries a detectable signal.
Variation in D across individuals	Informative variants at D-asymmetric sites occur mainly in non-Africans, so D calculated using different African individuals should vary *less* than D calculated using different non-Africans	Informative variants at D-asymmetric sites occur mainly in Africans, so D calculated using different African individuals should vary *more* than D calculated using different non-Africans	Rotating African individuals results in significantly more variation in D than equivalent rotation of non-Africans
Frequency and identity of alleles that drive positive D values	Almost the entire signal should be carried by low-frequency, derived Neanderthal B alleles in non-Africans, and these alleles are also absent from Africa.	Signal will mainly be carried by rare chimpanzee A alleles in Africans, these being generated by recurrent mutations at a higher rate compared with equivalent sites outside Africa.	Positive D is dominated by sites where the Neanderthal B allele is *fixed* outside of Africa and common within.
Variation in D across classes of sites with different mutational characteristics.	Theory predicts that D should not vary across classes of variants even if they differ in mutation rate *unless* mutation rate varies between populations.	D will vary across classes of variants wherever these exhibit mutation rate differences between populations.	D varies massively across different variant classes of sites, from less than zero to greater than 0.2.
Impact of heterozygosity	Introgressed DNA will tend to increase the heterozygosity of the recipient population.	Heterozygosity modulates the mutability of different motifs, and the large loss of heterozygosity out of Africa may reduce the overall mutation rate in non-Africans.	Genomic regions where non-Africans show the largest deficit in heterozygosity also exhibit the largest values of D.

Table 1. **Comparing the properties and predictions of the introgression hypothesis (blue) and the mutation rate variation hypothesis (pink)**. Results are summarised in the column 'Finding', highlighted in the colour of the hypothesis supported by the data.

Results

To be explicit, here and throughout this paper, we refer to all sites counted as ABBAs and BABAs as 'D-informative'; to the component of D generated by ILS as 'D-symmetric' sites; and the sites that drive non-zero D as 'D-asymmetric'. Note, that there is no easy way to tell individual D-symmetric and D-asymmetric sites apart. Many of our analyses leverage the differences in mutation rate between transitions and transversions or between different sequence motifs. We use 'class' to refer to any such grouping, regardless of whether this is a mutation type, a motif, or a combination of the two. For motifs, we follow Harris and colleagues in defining the 'mutation spectrum' as three-base sequences where an inferred mutation has occurred at the central base (Harris and Pritchard 2017).

Testing the assumptions of D

The two fundamental assumptions of D are that mutation rate is constant and that recurrent mutations are so rare that any possible impact they might have can safely be ignored. We point out above that both assumptions are questionable and test them more quantitatively.

Recurrent mutations are not vanishingly rare and should not be ignored when calculating D

It is unclear why recurrent mutations are generally ignored: it may just be for simplicity or because their potential importance was under-estimated. It may also have been based on their expected frequency, based on a single constant mutation rate equal to the genome-wide average of around $2 * 10^{-8}$ per base per generation. Such a calculation is naive because it ignores the way mutation rate varies massively depending on the genomic region and the identity and sequence context of the mutating base, with recurrent mutations becoming disproportionately more likely where the mutation rate is high. Whatever the reason, it is well-known that many SNPs in dbSNP carry three or even four bases. Across the much greater evolutionary distances covered by the whole chimpanzee-hominin clade, the greater time available for mutations to occur means the rate of recurrent mutations will be substantially higher.

For a more quantitative look into the distribution of multi-allelic SNPs, we downloaded the data from dbSNP (build 156). We find that an average 50kb window harbours 183 biallelic SNPs, 96 triallelic SNPs and 34 tetrallelic SNPs. In other words, over 41% of the SNPs carry more than two alleles (Figure 2). Since the number of alleles discovered at a site increases with the number and diversity of samples assayed, the dbSNP data will exaggerate the situation found in single studies. Repeating this analysis on the high-coverage HGDP data, we find that an average of 50kb window harbours 421 (96.6%) biallelic SNPs, 14 (~3.24%) triallelic SNPs, and 0.66 (0.15%) tetrallelic SNPs.

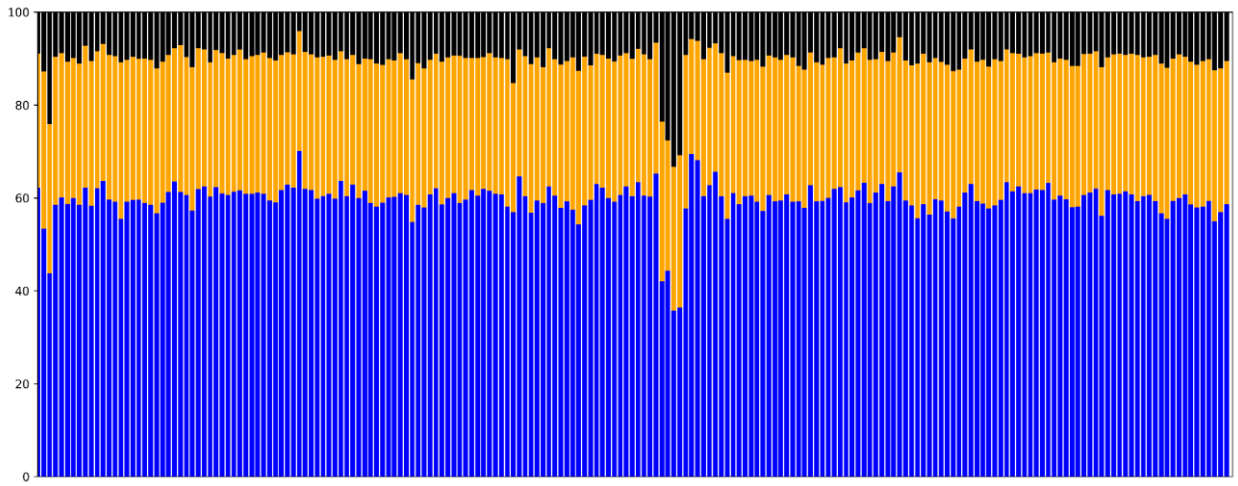

Figure 2. Numbers of SNPs carrying two, three and four alleles in non-overlapping 50kb windows along human chromosome 1. The y-axis shows the percentage of variable SNPs carrying two (blue), three (orange), and four (black) different bases in 50kb non-overlapping windows (x-axis) along chromosome 1. The results for all other autosomes are in Figure S1.

These numbers can be compared with the numbers of D-informative sites. Taking data from the original Green et al. (2010) study, about 200,000 D-informative sites are scored across 0.83Gb of aligned sequences, giving an average of about 12 D-informative sites per 50kb. Of these, only 8,265 are D-asymmetric, equivalent to 0.5 sites per 50kb. Thus, there are typically >1.15 triallelic sites per D-informative site and 28.5 triallelic sites per D-asymmetric site, the sites thought to be associated with introgression. Moreover, triallelic sites are generated by one transition and one transversion. Given that transitions occur at around twice the rate of transversions (Segurel, Wyman, and Przeworski 2014), we can expect two times the observed number of triallelic sites will carry 'silent' recurrent mutations, where either the same transition has occurred twice or a back-mutation has occurred. Thus, even ignoring the much larger numbers of recurrent mutations that would be found across hominins or across the chimpanzee-hominin clade, recurrent mutations within humans outnumber D-asymmetric sites by a ratio of 85.5:1 (there are 28.5 'visible' triallelic sites and for each of these there will be about two 'silent' triallelic sites, giving a total number of 3 x 28.5 = 85.5). It is clear that recurrent mutations cannot be treated as negligibly rare!

Recent human mutation rates are generally higher in Africa compared with outside

The key observation that Neanderthals are closer to non-Africans than to Africans does not distinguish between the European branch being shorter than expected *versus* the African branch being longer than expected. However, the mechanisms differ in their expectations if the outgroup is changed (Figure 3). Under the introgression hypothesis, the branch length asymmetry is created by the transfer of derived Neanderthal alleles to humans, so the European branch is shorter only with respect to Neanderthals (and their closest relatives). Consequently, if the Neanderthal is replaced by the chimpanzee (bottom panel) and evolutionary rate is constant across the tree, the asymmetry should disappear. Under the MRVH, the branch length asymmetry is an endogenous property of humans, so the asymmetry should be present regardless of whether the Neanderthal or chimpanzee is used as the outgroup.

To assess the dependence of the branch length asymmetry on which outgroup is used, we identified all derived alleles in Africans and Europeans with a derived allele defined as an allele that differs from the outgroup and is present in only one of the two human population groups. To illustrate, if Neanderthal = A, all Europeans are homozygous A, and Africans carry a mixture of As and Ts, the T is counted as a derived African allele. When treating the chimpanzee as the outgroup, we conservatively required the alleles in all three archaic hominins to match the chimpanzee allele, thereby avoiding sites carrying derived Neanderthal alleles that might otherwise complicate the interpretation. We ignored allele frequencies and instead counted how many derived African and derived European alleles occur in each non-overlapping 50kb window across the genome. On average, each window carries 7.2 and 1.6 derived African and European alleles, respectively, with the Neanderthal as the outgroup and 197 and 80 derived African and European alleles, respectively, with the chimpanzee as the outgroup. As a measure of any branch length asymmetry, we calculate a measure analogous to D, which we call D′, using Eq 1:

Eq 1. $D'_X = (nAf - nEu) / (nAf + nEu)$

Where X signifies the outgroup (N = Neanderthal, C = chimpanzee), nAf is the number of sites where Africans carry a derived allele, and nEu is the complementary quantity for Europeans. $D'_X>0$ represents more derived alleles unique to Africans and suggests a longer African branch, whereas $D'_X<0$ represents more derived alleles unique to Europeans. $D'_X=0$ suggests symmetry, with no significant difference in branch lengths. Therefore, D' provides a relative measure of the lengths of the branches. Note that robust absolute measures of mutation rate are impossible because there is a strong dependence on sample size and sample structure: larger sample sizes are more likely to include rare derived alleles, while relatedness between individuals within a sample reduces the effective sample size.

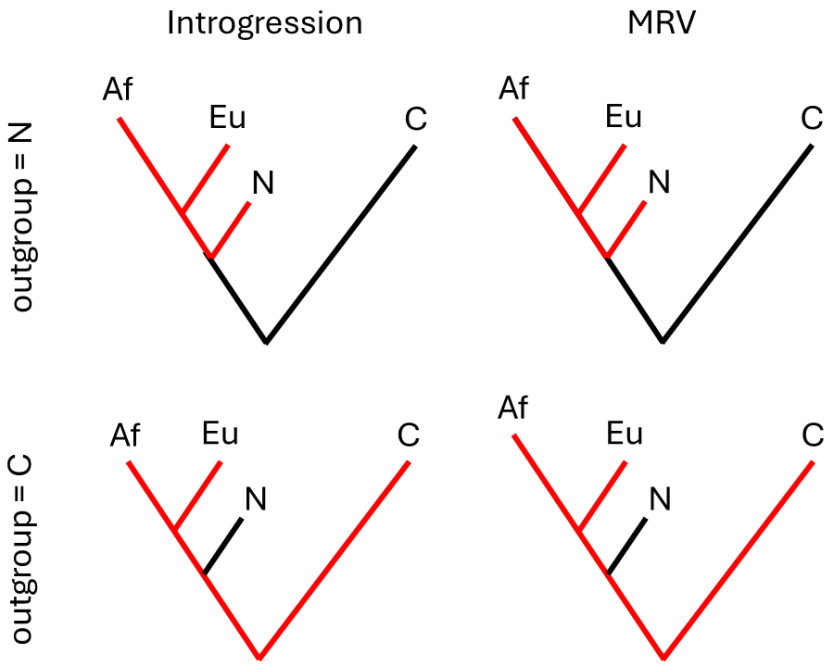

Figure 3. Simplified diagram to illustrate the expected branch length relationships under two competing hypotheses. The taxa represented are: Af (Africans), Eu (Europeans), N (Neanderthals) and C (chimpanzee). Branch lengths are not drawn to scale and the comparison of interest is highlighted in red. Introgression (left-hand panels) predicts that the observed African - European branch length asymmetry will disappear if the outgroup is changed from Neanderthal (top row) to chimpanzee (bottom row). Under mutation rate variation (MRV) the longer African branch is due to a higher mutation rate, so it is endogenous to humans and the asymmetry will be the same regardless of which outgroup is used.

Binning the data by D'_C, we plotted the corresponding average D'_N values on the vertical axis (Figure 4a). The two measures are clearly strongly positively correlated, indicating that in windows where Europeans are relatively closer to Neanderthals, they are also relatively closer to the chimpanzee. This demonstrates that the observed branch length asymmetry captured by D is not specific to comparisons involving the Neanderthal but instead appears to be an intrinsic feature of human evolution. Under the introgression hypothesis, these two statistics should be effectively uncorrelated. To confirm this inference, we then replotted the data, this time replacing D'_N with D(African, European, Altai, chimpanzee) (Figure 4b). As expected, a strong positive correlation exists, though with more scatter due to the relatively large reduction in number of contributing sites. Thus, across the genome, sites where Africans are relatively less similar to the chimpanzee compared with Europeans also tend to have higher D. This relationship has no sensible biological basis under the introgression hypothesis, but it is the expectation under the MRVH, where the branch asymmetry is driven by the African branch being inherently longer.

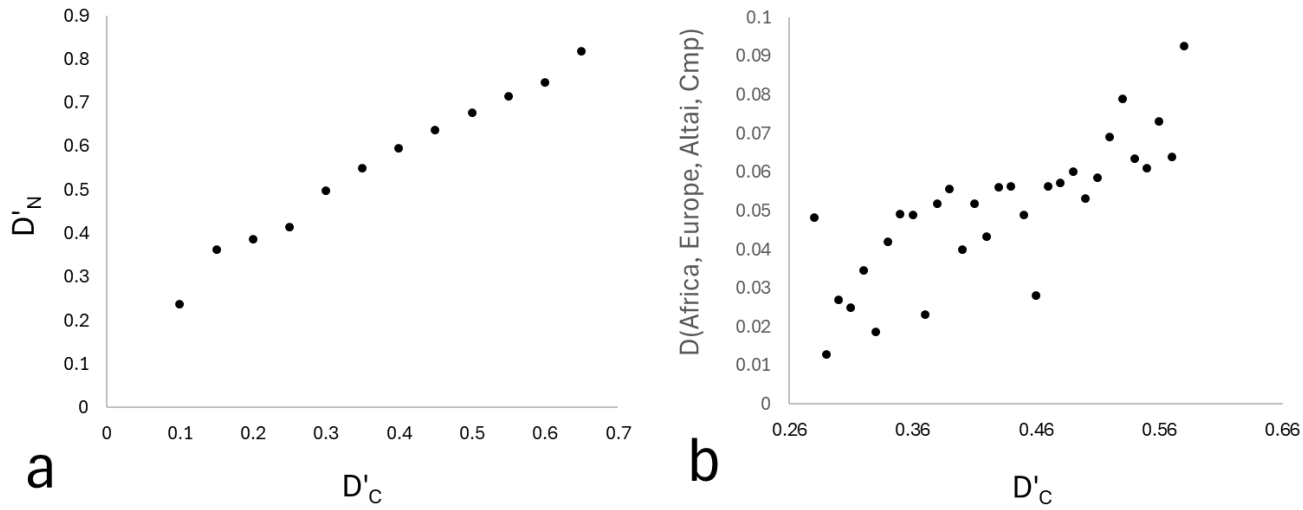

Figure 4. The human - Neanderthal branch asymmetry is driven by a longer African branch not a short European branch. We compared the relative lengths of branches separating the chimpanzee from Africans and Europeans, D′(C-Af, C-Eu) and written as D'_C, with the equivalent measure for Neanderthals, D′(N-Af, N-Eu), written D'_N. Panel (a) shows how these two measures are strongly positively correlated (r = 0.99, d.f. = 10, P = 2.3 x 10^{-10}, for clarity, the smallest D'_C bin, containing 0.2% of all values, is excluded because, although the corresponding D'_N value supports the trend, it is a large outlier), even though introgression should have little, if any, impact on D'_C. Panel (b) shows that D'_C also predicts D(Yoruba, French, Altai Neanderthal, chimpanzee) (sparsity of data in the distribution tails causes a large increase in variance, therefore graph depicts the middle 94.8% of all individual values, r = 0.8, df = 29, P = 6.6 x 10^{-8}). Each data point represents the average of all genomewide 50kb windows in the same D'_C bin.

As confirmation, we also conducted a spiking analysis. We analysed all pairwise comparisons between individuals from the 22 Yoruba samples, between all combinations of Yoruba and French individuals and between pairs of Yoruba individuals where a random 2% of the second Yoruba sample was replaced by Altai Neanderthal bases. The latter simulates introgression, with the Yoruba chosen as the African population because this population is often cited as the group with the least or no evidence of within-Africa introgression (e.g., Green et al. 2010; Mallick 2016; Skov et al. 2020). In each of these three comparisons, we analysed all 50kb non-overlapping autosomal windows and, within each, counted the numbers of bases of form BAAA, ABAA, BABA, ABBA, where the taxa are arranged as Yoruba1, X, Altai Neanderthal, chimpanzee with X being Yoruba2, French, or spiked Yoruba2. The relationship between D'_C and D'_N is presented in Figure 5.

As expected under both hypotheses, the unspiked Yoruban individuals (blue) reveal no relationship and very little variation in either variable. When the second Yoruban is spiked with Altai bases to simulate true introgression (red), almost all values for both variables are positive and, overall, a significant positive relationship is found (Spearman's rho = 0.97, N = 18, P < 8.4 x 10^{-6}), even though the range of values on the X-axis is still small (range -0.01 to 0.09, see inset). Around $D'_C = 0$, D'_N values are small, positive and approximately constant but increase strongly as D'_C increases. This pattern makes intuitive sense. Whether due to a different evolutionary rate, different generation length, early extinction, more sequencing errors or a combination of factors, it is reasonable that the Neanderthal - chimpanzee branch is not the same length as the human - chimpanzee branch, hence the residual, small positive value of D'_N. Consequently, in windows most affected by spiking, both D'_C and D'_N increase, but the much greater impact is on D'_N, the source of the spiking material.

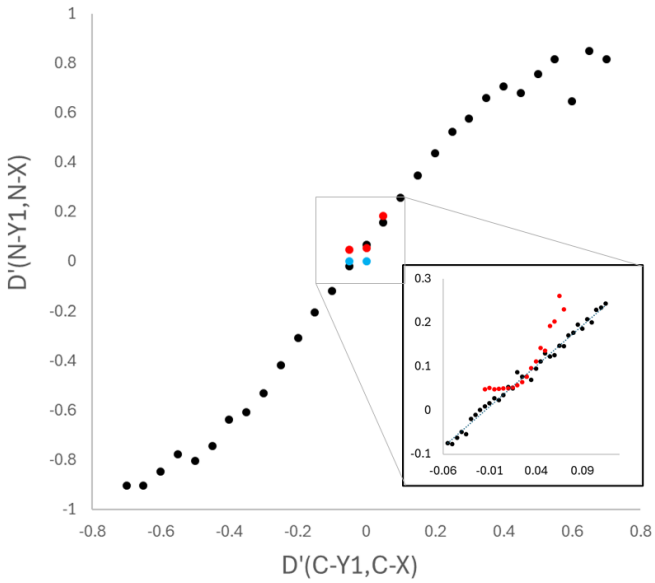

Figure 5. Testing whether the human - Neanderthal branch asymmetry is driven by a longer African branch or a shorter European branch. Plotting D′(C-Y1,C-X) against D′(N-Y1,N-X), where C = chimpanzee, N = Altai Neanderthal, Y1 = a single Yoruban individual and X = either a second Yoruban (blue), a second Yoruban spiked with 2% Altai Neanderthal bases (red), or a French individual (black). The first two scenarios act as controls for no difference between the population and what should be seen if introgression had occurred, respectively. The third scenario, although showing a significant positive correlation (Spearman's rho = 0.97, N = 18, $P < 8.4 \times 10^{-6}$), does not approach the real African - non-African comparison in terms of strength of correlation and range of values seen. These results confirm that the human - Neanderthal branch asymmetry is driven by a longer African branch, not a shorter European branch.

Compared with both the spiked data and the negative control, the Yoruba - French comparison (black) exhibits a radically different relationship, with a strong, approximately monotonic correlation (Spearman's rho = 0.99, N = 29, $P < 2.2 \times 10^{-16}$) across a vastly wider range of values on both axes. These data bear no relation to the simulated introgression, either in the shape of the relationship or in the extent of variation of the two variables. Of course, the simulated introgression is, by no means perfect, but it should approximate the impact of one individual carrying 2% Neanderthal DNA compared with a second individual who does not, and it makes intuitive sense that the impact of such a small legacy will be modest at best. Indeed, the appreciable difference between profiles in the inset suggests that introgressed material in Europeans would distort the D'_C and D'_N correlation in the real data. That no distortion is apparent and the real comparison shows such a contrasting pattern compared with the spiked data confirms that, even if a signal from introgression is present, it is completely dwarfed by the signal from another process that drives far greater variation in both variables and a more or less monotonic relationship.

D varies extensively across bases with different mutational properties.

According to theory, and if the underlying assumptions hold, D should not vary between sites with different mutation rates (Durand et al. 2011). To test this prediction, we considered classes of sites across the mutation spectrum (see Methods). Following Harris and Pritchard (2017), the 64 possible mutating triplets were collapsed to 32 by converting triplets with central G or T to their opposite strand equivalents. Instead of considering all possible base changes, we partitioned the data simply into

transitions and transversions, giving 64 classes of mutation. For each class, we calculated D(Yoruba, French, Altai, chimpanzee). Other African and non-African population combinations yield essentially identical results. We repeated this analysis twice with two controls. As a negative control, we conducted all pairwise comparisons among individual Yoruban samples, with the expectation that D would be zero across all site classes. As a positive control, all pairwise Yoruban comparisons were conducted after spiking the second Yoruban with 2% Altai Neanderthal bases (as above). If D is driven primarily by introgression, the Yoruba-French comparison should generate profiles identical to the Yoruba-spiked Yoruba comparison.

As expected, the negative control profiles (black and gray) show little variation around zero (Figure 6). The positive controls, where the genome of one of each pair of Yorubans is spiked with Neanderthal DNA to simulate introgression (blue), exhibit modest variation in D, the mean value and variation about the mean both being greatest in panel (a), where non-human hominins are polymorphic. These profiles contrast strikingly with the red (transversions) and yellow (transitions) profiles of real comparisons between pairs of Yoruban and French individuals. Here we find massive variation in D, with several triplets yielding zero or even negative D values, particularly when the three archaics are monomorphic, and other triplets yielding values up to and slightly exceeding 0.2. These results again expose a large difference between simulated introgression and real data and show that D varies between different site classes, strongly consistent with previous observations that the mutation rates of different triplets vary between human populations (Harris and Pritchard 2017). It is noticeable that the most negative D-value is for triplet TCC (3rd from the right), the triplet identified as having a significantly higher mutation rate in non-Africans compared with Africans. Since D is not expected to vary if mutation rate is constant across human populations, even if introgression has occurred, the large variation in D we uncover suggests strongly that human population do differ extensively in the mutation rate of different motifs.

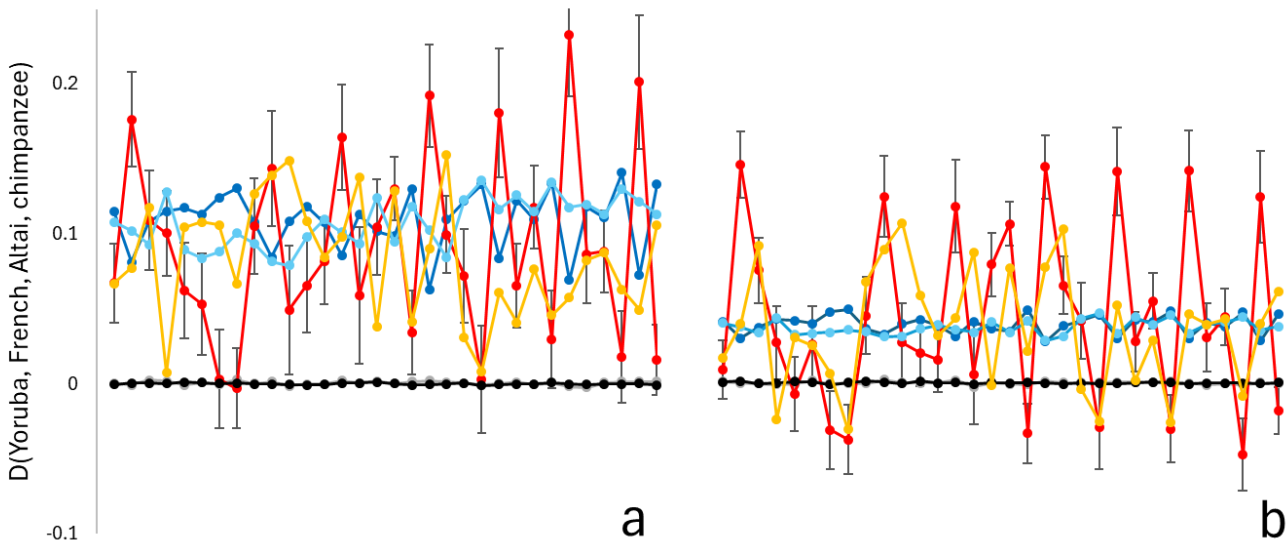

Figure 6. Variation in the introgression statistic D across the mutation spectrum. Data were partitioned according to the mutating triplet (x-axis) and whether the inferred mutation was a transversion (TV), or a transition (TS). Each panel depicts D values from three different population comparisons: Yoruba-French (TV=red, TS = yellow), a negative control of pairwise comparisons between individual Yorubans (TV = black, TS = gray) and a positive control of Yoruba-spiked Yoruban (TV = dark blue, TS = light blue), where 'spiked Yoruban' means that a random 2% of the Yoruban bases were replaced by their Altai homologues. Panel (a) is based on sites where the Altai allele differs from either the Vindija or the Denisovan. Panel (b) is based on sites where all three archaics carry the same allele. For clarity, error bars are included only for the main series (Yoruba-French transversions). Other error bars are similar but generally a little smaller. Triplet order is AXA, AXC, AXG, AXT,

CXA, CXC, CXG, CXT, GXA, GXC, GXG, GXT, TXA, TXC, TXG, TXT, where X = A for the first 16 triplets and X = C for the second 16.

How are D-informative sites distributed across the genome?

Having established that neither of the two key assumptions underpinning D are met, we next examined the distribution of D-informative sites across the genome. Under introgression, within an individual, D-asymmetric sites will be confined to around 2% of a given non-African genome, the inferred size of an average legacy. Under the MRVH, informative sites can occur anywhere, suggesting a more diffuse distribution.

Sites that contribute to D are not as tightly clustered as expected under the introgression hypothesis

To determine a realistic distribution of introgressed fragment sizes, we used Monte Carlo simulations based on the high-resolution recombination map of Halldorsson et al. (2019). Specifically, we used the cumulative recombination fractions and randomly selected one recombination event per Morgan per generation for 2,000 generations. The resulting set of recombination breakpoints defines the locations and size distributions of fragments that would have been generated by recombination since any introgression occurred. From this set, we randomly selected fragments totalling 2% of the genome to represent the modern legacy. Next, the chosen fragments were superimposed on a set of non-overlapping 50kb windows to determine the expected number of windows that would contain introgressed material. Repeating this procedure 100 times, we find an average of 2.4% of windows (95% CI: 1.4%–3.4%) contain at least some introgressed material. However, not all introgressed fragments will carry a D-informative site, so the number of windows containing fragments detectable through D will be even smaller. This result highlights the sparsity of sites that would carry any signal of introgression and the challenges facing any attempt to use D to detect introgression across the genome.

To determine the empirical distribution of 50kb windows that contribute to D-asymmetric sites, we calculated counts of ABBAs and BABAs for D(Yoruba, French, Altai, chimpanzee) in all non-overlapping 50kb windows. After ordering the counts from the most negative value of D to the most positive, we determined the point where the summed ABBA and summed BABA counts yielded a D value of exactly zero. These sites account for the D-symmetric sites that likely arose from ILS. The remaining windows are deemed responsible for the D-asymmetric sites. Since these windows have been chosen from an ordered list, they carry the most extreme ABBA-BABA asymmetries and hence represent the minimum number of windows needed to explain the observed positive D. We find these remaining windows constitute a total of 11% of all widows, at least five-fold greater than the number expected under introgression. However, this already conservative ratio should still be higher because a simulated introgressed fragment that spans two or more windows is deemed to be detectable in any of these windows, thus inflating the count of introgressed events. In reality, a large majority of fragments will carry either zero or one D-asymmetric site, so they will usually be detectable through the presence of excess BABAs in only one of these windows.

Variants driving variation in D tend to be found in Africans

Under the introgression hypothesis, variation in D among individuals is driven by the size and number of introgressed haplotypes carried by individual non-Africans, with Africans acting as passive controls. A testable prediction is that, among a number of equivalent pairwise comparisons, rotating the non-African individual will have a greater impact on D than rotating the African. To test this prediction, we

focused on the 22 Yoruban and 28 French individuals in the HGDP data set, conducting all pairwise African – non-African comparisons and calculating D in every autosomal, non-overlapping 100kb window. We use 100kb here rather than 50kb used elsewhere to increase the number of informative sites detected per window. For each individual, we then calculated the variance in D across all valid comparisons. To avoid any possible sample size bias, each Yoruban was compared against a random 22 French individuals. Considering only windows where all D values can be calculated (i.e., no individual comparison yields zero ABBAs and zero BABAs), we find that 2558 / 2793 windows (91.7%) have a higher variance when African individuals are rotated than when non-Africans are rotated, and this mean difference is highly significantly greater (excess variance = 0.021 +/- 0.0003 1 s.e.m, t = 62.5, d.f. = 2792, P=0). Thus, whatever drives non-zero D varies more between Yoruban individuals, thought to contain no introgressed material, than between French individuals, thought to carry 2% Neanderthal DNA. As elsewhere, this is the exact opposition of what is expected under the introgression hypothesis.

How many mutations have typically occurred at sites that contribute to the positive signal captured by D?

The introgression hypothesis requires only a single mutation to generate a derived variant in Neanderthals that then enters non-Africans via inter-breeding. The MRVH requires two mutations to have occurred, one creating a polymorphism outside humans followed by recurrent mutation within humans. To determine the number of mutations that have impacted sites that are informative for D, we exploited the fact that the probability of a single mutation occurring will scale linearly with mutation rate (μ), while the probability of two mutations occurring at the same site is expected to scale with μ^2.

We began by looking at transition:transversion ratios (TS:TV). In terms of single mutations, TS:TV is typically around two, i.e., $\mu_{TS} \sim 2\mu_{TV}$. At sites where two mutations have occurred, double transversions will occur at a rate that scales with μ_{TV}^2, while two transitions will occur at a rate that scales with $\sim 4\mu_{TV}^2$. In other words, sites carrying two mutations should exhibit a TS:TV ratio that is approximately double that seen at sites where only a single mutation has occurred, providing a simple test for whether D-informative sites generally carry one mutation (introgression) or two mutations (MRVH).

To determine the number of mutations characteristic of D-asymmetric sites, we calculated TS:TV ratios for 12 scenarios, chosen to represent instances where the likely history of the site features one or two different mutations (Table 2). Where only a single mutation is likely, the TS:TV ratio is around two. Since ILS between humans and chimpanzees will be negligible, the state PBBBA (P = polymorphic in humans) is almost certainly generated by two different mutations. In line with expectations, the TS:TV ratio is a little over four. Finally, sites that are polymorphic within both humans and non-humans can be generated by ILS, recurrent mutations, or introgression. We find TS:TV ratios ranging from 2.2 - 2.46, all of which are highly significantly greater than 2.08, the ratio for single mutations within humans (chi-squared tests, all $P < 10^{-11}$). Interpreting these values as mixtures of single mutation sites (TS:TV = 2.081, based on polymorphic sites in humans) and double mutation sites (TS:TV = 4.12), the mid-value of 2.28 would equate to 90% of sites being generated by ILS with one mutation and the remaining 10% all carrying two mutations. We note that under the introgression hypothesis, all sites are expected to carry a single mutation, implying a ratio around two. Moreover, the values from Green et al. of 8,265 D-asymmetric sites in around 200,000 indicate that 95.9% of sites are due to ILS, so our estimate of 90% leaves no room for a signal due to introgression.

Scenario	Humans	N μ	TV	TS	ratio

AAABA	mono	1	110,634	207,480	1.875
AABAA	mono	1	18,557	34,370	1.852
ABAAA	mono	1	17,899	34,576	1.932
BAAAA	mono	1	2,064	4,641	2.249
ABBAA	mono	1	39,859	78,231	1.963
PAAAA	poly	1	10,256,280	21,344,476	2.081
PAABA	poly	2	114,311	261,010	2.283
PABAA	poly	2	16,082	35,664	2.218
PBAAA	poly	2	8,050	19,819	2.462
PBBBA	poly	2	252,011	1,038,149	4.12
PBBAA	poly	2	142,553	315,757	2.215

Table 2. TV:TS ratios by type of site for an alignment of (in order): humans, Vindija Neanderthal, Altai Neanderthal, Denisova, chimpanzee. The two Neanderthals are treated either as separate taxa or as a single taxon (i.e., PBBAA). Only bases where all taxa have been called and where each individual except the humans is homozygous are considered. Bases that are polymorphic in humans are indicated by a P. The likely or possible numbers of mutations are given as N μ. The number of sites involving TV, TS and their ratio is latly given.

Across the mutation spectrum, more mutable triplets are more likely to be subject to recurrent mutations and also contribute disproportionately to positive D

We next sought to extend the TS:TV analysis (Table 2) by considering the entire mutation spectrum. Here it is important to distinguish between *absolute* and *relative* mutation rates. The absolute mutation rate is the probability that a given class of sites mutates in a given generation, while the relative rate is the proportion of all mutated bases that are attributed to that class. These quantities differ because sites with a high absolute mutation rate, although more likely to mutate, also tend to be rare precisely because they mutate rapidly to other states. Consequently, the classes of sites with the highest mutation rates actually contribute rather little to D in terms of the overall number of variants they generate. Since we are interested in learning which sites contribute most to D, we will focus on the *relative* mutation rate.

We counted the numbers of each of the 32 triplets that mutated on the chimpanzee-hominin tree, splitting the data into transitions and transversions. We also categorised sites based on whether they had likely experienced one or two mutations (see above). Figure 7 depicts relative mutation rates for two classes of single mutation sites (AABAA and PAAAA) and for unambiguous double mutation sites (PBBBA). The six profiles are dominated by four peaks corresponding to transitions at the triplets ACG, CCG, GCG, and TCG, i.e., sites where the central base is a C in a CG doublet (henceforth referred to as 'XCG' sites). Interestingly, although XCG transitions have the highest relative mutation rates at double mutation sites, they also contribute some of the lowest relative mutation rates at sites where only one mutation has occurred. This is because XCG sites are inherently rare and only become *relatively* common among the small number of sites generated by recurrent mutations, where the most mutable sites are disproportionately represented.

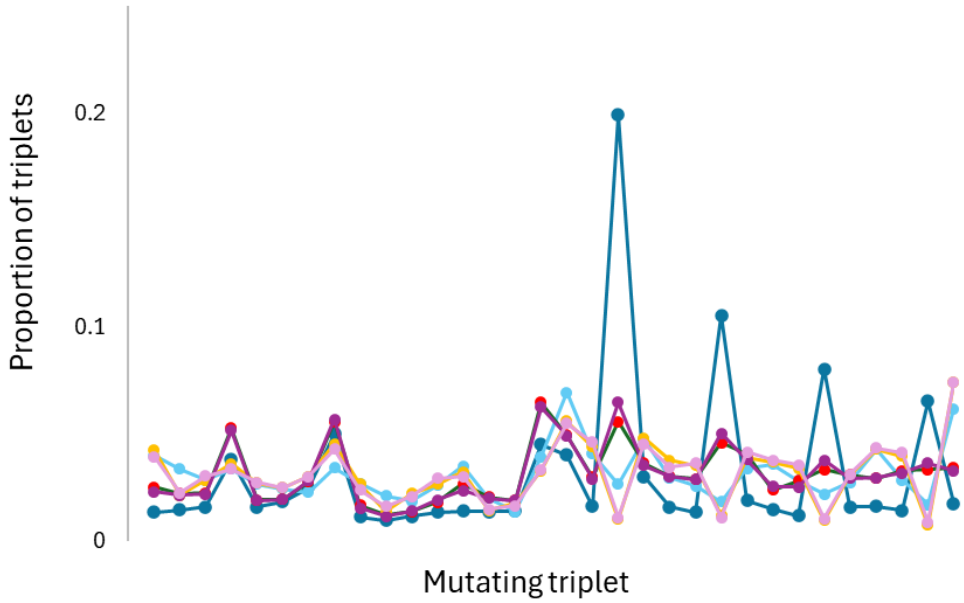

Figure 7. Proportions of triplets that have likely mutated once versus twice. Three site classes are illustrated, each partitioned into transitions (TS) and transversions (TV): sites that likely have a single mutation in the Altai Neanderthal (AABAA; TS = red, TV = yellow), sites with likely a single mutation in humans (PAAAA); TS = purple, TV = pink, the red P indicates polymorphism), and sites with likely two mutations, one between the chimpanzee and hominins and one within humans (PBBBA; TS = dark blue, TV = light blue). The proportion of triplets is the number of each triplet type divided by the number of all 32 equivalent triplets that have mutated in the same class. Triplet order is as in Figure 6.

Having established that triplets vary greatly in their relative mutation rates across the spectrum, we next considered how this variation impacts their relative contribution to D. Non-zero D can be split into two distinct components: D-symmetric sites and D-asymmetric sites. Under the introgression hypothesis, both components are generated by single mutations, and recurrent mutations are typically ignored. Under the MRVH, the asymmetrical component is driven entirely by recurrent mutations, predicting a disproportionate contribution from the most mutable triplets. To test this prediction, we derived a simple measure, M1, of the relative contribution of each triplet to the asymmetrical component:

Eq 2. M1 = piABBA - piBABA

Where piABBA is the proportion of all ABBA counts contributed by triplet *i*. Under the introgression hypothesis, recurrent mutations are not involved, and all triplets contribute equally to both ABBAs and BABAs. Under the MRVH, the excess state (here ABBAs, though in Green et al. and a few other instances it is BABAs, there is as yet no convention) will include a proportion of sites with recurrent mutations, so the more mutable triplets will contribute more to ABBAs than to BABAs, generating positive M1 scores. In a directly parallel way, we can also generate a second measure, M2, that captures a measure of the relative mutation rate of each triplet by calculating:

Eq 3. M2 = piPAAAA - piPBBBA

Where piPAAAA is the number of times triplet *i* is polymorphic only in humans expressed as a proportion of all such sites, while piPBBBA is the number of times triplet i is polymorphic in humans

at sites where the chimpanzee also differs from all three archaic hominins, again expressed as a proportion of such sites. Since piPAAAA involves sites carrying a single mutation while piPBBBA involves sites most likely to arise by recurrent mutations, M2 will tend to be large for triplets that contribute relatively more to double-mutation sites than to single-mutation sites. Plotting M1 against M2 reveals a strong and highly significant positive correlation (Figure 8, r = 0.88, df = 30, P = 2.9 x 10^{-11}), indicating that the higher the *relative* mutation rate of a triplet, the more that triplet contributes to the asymmetric component of D. Thus, non-zero D is dominated by sites with relatively higher mutation rates. Note, it is critical to use measures based on a difference between two related states. Any direct measure of mutation rate is open to the criticism that a high mutation rate generates more informative sites and hence a stronger signal, but the relative representation of each triplet in two related states, such as in ABBA and BABA sites, should control for factors such as the relative abundance of each motif in the genome.

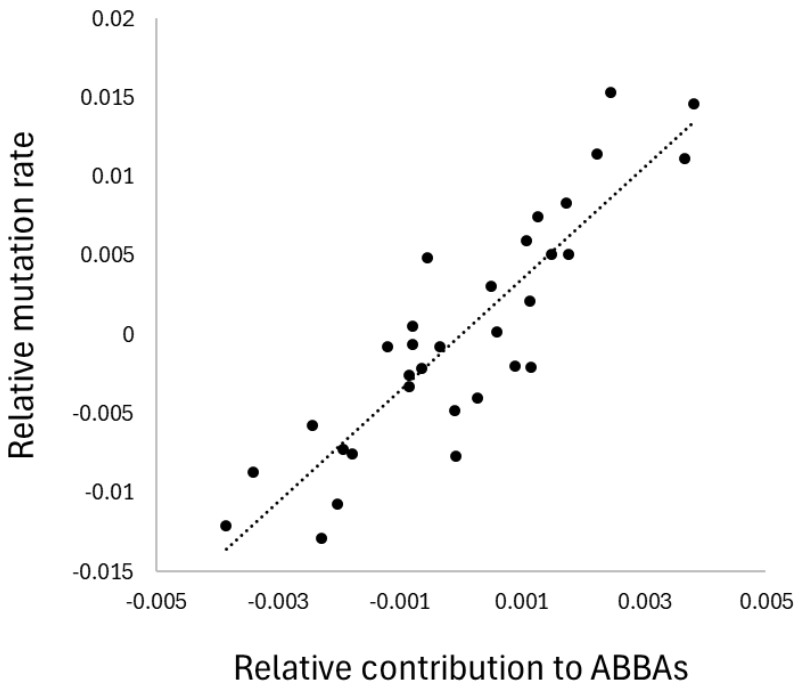

Figure 8. The relationship between relative triplet mutability and relative contribution to the signal captured by the introgression statistic D. Each axis represents the difference in the proportion that each of the 32 possible triplets contributes to a pair of closely related scenarios. For contribution to D, we take the difference in the contribution of each triplet to ABBAs (the excess state) minus the proportional contribution to BABAs. Positive scores indicate a higher representation in ABBAs relative to BABAs, the state currently interpreted as indicative of introgression. For mutation rate, we use the difference in proportion between sites where only humans are polymorphic (PAAAA), where only a single mutation has occurred, and sites where humans are polymorphic and the chimpanzee also differs from all three archaics (PBBBA), sites likely to have experienced two mutations. Each data point represents a specific triplet type, with a strong positive correlation indicating that triplets that are more likely to carry two rather than one mutation also tend to contribute disproportionately to positive D. The data presented are for transversions. Transitions exhibit a similar correlation for most triplets, but the XCG triplets are major outliers.

Non-zero D values are driven by recent mutations and reflect a balance between two completely opposing signals.

The two hypotheses give entirely opposing predictions as to which alleles should drive non-zero D. Under introgression, the key alleles are rare Neanderthal B alleles that entered non-Africans via inter-breeding. Under the MRVH, the key alleles are rare A alleles generated in Africa by recurrent mutations, mostly at the common BBBA state. To explore which hypothesis fits best, we calculated D(African, European, Neanderthal, Chimpanzee) twice, based on whether the chimpanzee A or the Neanderthal B allele is the minor allele in humans. We also partitioned the data according to the mutation spectrum and by transitions and transversion. For a semi-independent but less extreme comparison, we repeated the calculations for D(South Asia, East Asia, Neanderthal, Chimpanzee).

Remarkably, although the overall D value are as expected (overall D = 0.051 and 0.023 for the Africa-Europe and South Asia – East Asia comparisons respectively), splitting sites according to which allele is the major allele reveals two much larger but opposing component signals (Figure 9). For D(Africa, Europe, Altai, chimpanzee), large positive values (D ~ 0.3) occur when the major allele in humans is the Neanderthal allele B. This is consistent with the MRVH, where we expect to find the common BBBA mutating to recent, rare A alleles more often in Africa, but is the exact opposite of what is expected under introgression, where the Neanderthal B is expected to be rare and hence *must* be the minor allele. At sites where the Neanderthal B allele is the human minor allele, D is strongly negative (D ~ -0.25). Under the introgression hypothesis, this would have to be interpreted as widespread, extensive inter-breeding between Neanderthals and Africans, not Europeans!

The negative D values obtained when the B allele is the minor allele in humans are also consistent with the MRVH. When the B allele is the minor allele in humans, the most likely scenario for such sites is that humans were ancestrally A and have since had an A->B mutation. With the chimpanzee = A and Neanderthal = B, the typical sites where this occurs will appear as AABA. Here, a higher rate of recurrent mutations in Africa will generate more B alleles in Africa than outside, producing negative D.

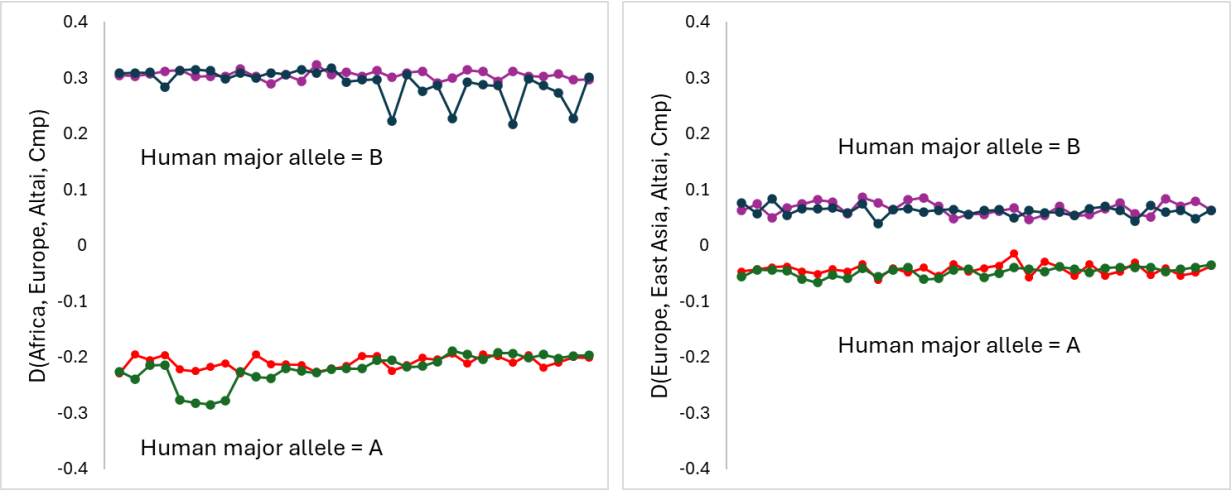

Figure 9. Comparative D values across the mutation spectrum, comparing cases where the chimpanzee A allele is either the major allele or the minor allele. For each triplet (x-axis), D is calculated four times, corresponding to transitions and transversions at sites where the chimpanzee A

allele is the minor allele in humans (TV = purple, TS = blue) and where the Neanderthal B allele is the minor allele (TV = red, TS = green). Two scenarios are illustrated: D(Africa, Europe, Altai, chimpanzee) and D(Europe, East Asia, Altai, chimpanzee). Triplet order is as in Figure 6.

Exploring which alleles tend to be common or rare among sites that contribute most to positive D

As shown, D-asymmetric sites are driven mainly by sites where A is rare in Africa rather than sites with a rare B in non-Africans. To examine this finding further, we sought to break D down into components that can be attributed to particular site classes.

Positive D values are dominated by sites where the Neanderthal B allele is fixed outside of Africa.

To explore further the characteristics of sites that contribute most to driving non-zero D, we sought to decompose D. D is the normalised difference in ABBA and BABA counts, (ABBA-BABA)/(ABBA+BABA) but normalisation masks the contribution of different classes of site. Consequently, we focused on the unnormalised quantity ABBA minus BABA, which we refer to as D*. We used a standard taxon combination: Yoruba, French, archaic, and chimpanzee and classified sites according to whether each variant represented a transition or a transversion and which allele was fixed in which human population group, using P to indicate both A and B are present (Table 3). All D* values are expressed as a percentage of the total signal (all classes of sites with TS ansd TV combined). The scenario expected to dominate under introgression, where the chimpanzee A allele is fixed in Africa, and the Neanderthal B is rare outside Africa (highlighted in blue), consistently contributes somewhat less than 30% overall, with transitions contributing almost exactly twice as much as transversions. In contrast, the scenario expected to dominate under mutation rate variation, where the Neanderthal B allele is fixed outside Africa and Africans carry a rare A allele (highlighted in pink) contributes in excess of 100%, with transitions contributing approximately three times as many counts as transversions (it is possible to exceed 100% because some components are by force negative).

Several additional trends should be noted. First, the TS/TV ratios exhibit interesting and informative patterns. For A/P and P/A, the ratios are consistently just above two for all three archaics (A/P = 2.12, 2.13, 2.22; P/A = 2.26, 2.25, 2.25 for Vindija, Altai and Denisovan respective). For B/P, the ratios are substantially higher at around five (5.15, 5.13, 4.61). Such sites are rare and are likely the product of recurrent B->A mutations outside Africa on the common BBBA background. This pattern is entirely at odds with introgression but might be explicable under the MRVH if a lower general mutation rate outside Africa means that an excess of recurrent mutations occurs only at the most mutable sites such as XCG transitions. Next, P/B has ratios around three (2.9, 2.9, 2.93), significantly above the ratios seen for A/P ($t = 19.86$, $df = 66$, $p < 2 \times 10^{-16}$, based on dividing the genome into 34 non-overlapping subsets of 1,000,000 informative sites). This is strongly consistent with the MRVH, where contributing sites are expected to be enriched for transitions with their higher mutation rates. Finally, when both population groups are polymorphic, D* is small and variable, being sometimes negative and sometimes positive, leading to a such high variance in the ratios that the average ratio is near meaningless. We note that the A/P and P/A scenarios give similar ratios despite having positive and negative D* values, respectively. Also, TS:TV ratios appreciably above four are possible because triplets vary greatly in their ratios (Figure S2), meaning that strong enrichment for recurrent mutations will enrich for triplets with the highest ratios and hence drive ratios above the four expected based on the sequence average ratio of about two.

Sites that are polymorphic in both populations generally contribute little to overall D*, meaning that the size of reported D depends on a balance between the negative and positive elements of the other

four classes. There is no formal theoretical framework for interpreting which negative signals cancel which positive signals, not least because this will depend ultimately on the underlying mechanism: introgression, MRV or a mixture of the two.

	Altai		Vindija		Denisovan	
YOR/FRC	**TV**	**TS**	**TV**	**TS**	**TV**	**TS**
A/P	17.2 (0.83)	36.4 (1.9)	16.9 (0.81)	36 (1.9)	19.3 (1)	42.9 (2.1)
P/A	-21.2 (0.5)	-47.8 (1.1)	-22.8 (0.5)	-51.3 (1.2)	-77 (1.4)	-173 (3.1)
B/P	-3.3 (0.23)	-16.8 (0.65)	-3.5 (0.24)	-18.1 (0.7)	-13 (0.8)	-60.2 (2.2)
P/B	35.2 (0.65)	102.1 (1.4)	37.6 (0.7)	109 (1.6)	105 (1.7)	307 (3.9)
P/P	3.4 (1.4)	-5.2 (3)	3 (1.6)	-6.9 (3.4)	-1.4 (4.19)	-49 (7.7)
Total	31.3%	68.7%	31.3%	68.7%	32.6%	67.4%

Table 3. Partitioning the signal captured by D according to which allele is fixed in which population group. A 'standard' population comparison D(Yoruba, French, archaic, chimpanzee) was used to generate ABBA and BABA counts for each of the three archaic hominins. Each site was classified according to whether the inferred mutation was a TV or a TS. Counts were also partitioned according to whether the chimpanzee A allele or the archaic B allele was fixed in one or other population (column YOR/FRC), coded such that A/P indicates a site where the A allele is fixed in Yoruba and the French carry both A and B. The scenarios expected to dominate under the introgression hypothesis and the MRVH are highlighted in blue and pink, respectively. All counts are expressed as a percentage of the total D* used to compute D, i.e., summed across all scenarios and across transitions and transversions. Confidence is indicated by one standard error of the mean (in parentheses) obtained by calculating separate counts for non-overlapping contiguous blocks of 1 million informative sites (*n* = 34 blocks). Note that when one population is fixed for one allele, either ABBAs or BABAs are impossible, and so the quantity D* is by force always either negative (P/A and B/P) or positive (A/P and P/B). Also, since the values in the Table are expressed as a percent of the total D*, the small overall difference for Denisovans creates much larger percent values despite rather similar raw counts.

Importantly, subtracting all negative D* values from the scenario expected under the MRVH still leaves a substantial signal, accounting for 48.2% of the total. However, if the negative D* values are subtracted from the A/P counts expected under introgression, the remaining signal becomes negative (-35.5%). Additionally, while D(Yor,Frc,Den,Cmp) is much smaller than D(Yor,Frc,Nea,Cmp), and the same applies to their corresponding D* values, the raw D* counts for the components are relatively similar. Consequently, when expressed as a percentage of the overall signal, the percentage values are approximately three times larger for Denisovans than Neanderthals. Once again, these data are consistent with a model where the probability of recurrent mutations varies between different classes of sites but is challenging to reconcile with introgression, where all sites are generated by single mutations. In the absence of some previously undiscovered process, elevated TS/TV ratios should only occur where an appreciable proportion of sites involve recurrent mutations, so they should not be found if D is driven mainly by variants created by a single mutation that then entered humans by introgression.

Finally, we extended the analysis of D* by further partitioning the data according to the mutating triplet involved. Results are summarised in five plots (Figure 10) for transversions (left-hand side) and five plots for transitions (right-hand side), arranged in vertical order as in Table 3. We note that the only scenario that contributes majorly to the overall profile involves transition mutations where the B

allele is fixed outside Africa. As expected (Table 3), the scenario expected under introgression (A/P) contributes only a very small proportion of the overall signal and largely fails to reflect the major peaks and troughs seen overall. These findings confirm that positive D is primarily driven by sites that are unlikely to have been created by introgression.

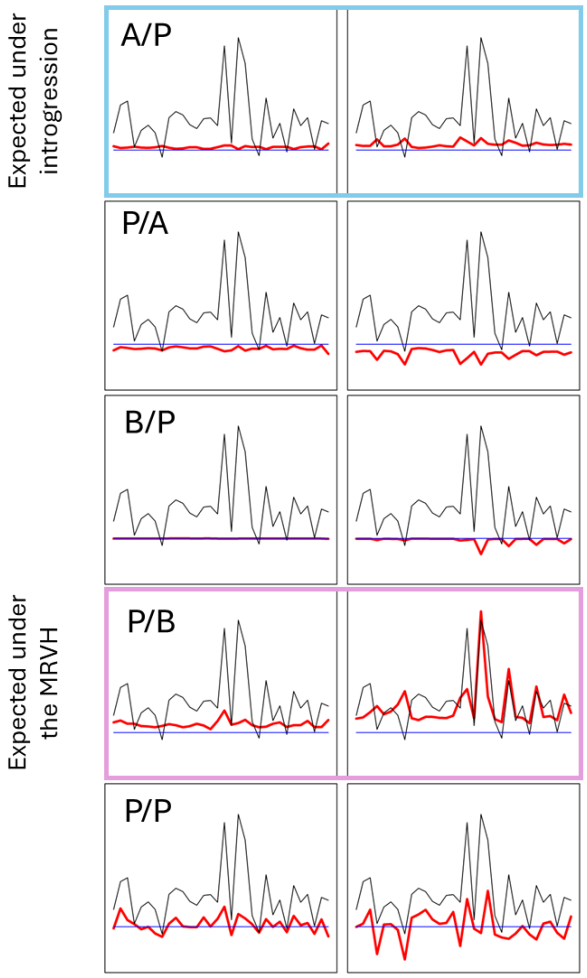

Figure 10. Partitioning the introgression statistic D across the mutation spectrum. D*, the total ABBA minus BABA (y-axis) was calculated over all sites for the taxon combination (Africans, Europeans, Altai, chimpanzee). Counts were partitioned according to triplet (x-axis) and whether the inferred mutation was a transversion (left-hand panels) or a transition (right-hand panels). The data were further partitioned according to which allele (chimpanzee A or Neanderthal B) was fixed in each human population group with labels and colors as in Table 3. Within each panel, the thick red lines represent the D* profiles for that particular scenario. For reference, every panel also carries the same thin black line representing the overall profile, with D* calculated for each triplet over all five scenarios and over both transitions and transversions. The horizontal blue line is D* = 0.

Testing for a possible mechanism capable of generating the observed patterns

Thus far, we have established that D is based on false assumptions and that its positive component is dominated by sites that have higher mutation rates and have usually mutated twice rather than once, with the second mutation being more recent. We also demonstrated that D is dominated by sites where the Neanderthal B allele is fixed outside Africa, while the chimpanzee A allele is rare inside Africa. Taken together, these strong trends lend unambiguous support to the idea that D captures a signal that is not driven mainly by introgression but instead arises from a higher mutation rate in Africa, causing an excess of recurrent mutations converting the common BBBA state to ABBA. This raises obvious questions about the possible mechanism that might be involved in generating mutations.

It has been suggested that variation in mutation rate might be linked to as yet unknown variants in enzymes involved in DNA replication (Liu and Samee 2023). However, striking changes in mutation spectra reported by Harris and colleagues (2017) suggest another possible mechanism. When the sequence context of each polymorphism is examined, some triplets increase in mutability when flanking heterozygosity is higher, while others become less mutable (Amos 2019). Given that non-Africans lost some 25% of their heterozygosity during and after the out-of-Africa event, creating a strong linear decline in heterozygosity with distance from Africa (Prugnolle, Manica, and Balloux 2005), the influence of heterozygosity on triplet mutation rate could conceivably account for both an overall change in mutation rate and a shift in mutation spectrum. Consequently, we next tested for a relationship between heterozygosity loss and D.

Heterozygosity and D are negatively correlated

Across the genome, natural selection has modulated the rate at which heterozygosity was lost out of Africa, with positive selection outside Africa sometimes acting to accelerate loss and balancing selection sometimes acting to retard loss or even, rarely, promote an increase in heterozygosity (Amos and Bryant 2011). Consequently, for any given genomic region, the slope of the relationship between heterozygosity and distance from Africa provides an excellent proxy for the impact of the out-of-Africa event on heterozygosity. To test the hypothesis that changes in mutation rate linked to loss of heterozygosity might be involved in driving positive D, we analyzed each non-overlapping 50kb window across the genome separately and within each Eurasian population. We calculated both D(Yoruba, French, Altai, chimpanzee) and heterozygosity in every population. To calculate the slope of the relationship between heterozygosity and distance for East Africa, we used all Eurasian populations, including the Middle East, but excluding all European populations. This conservative approach aims as far as possible to decouple the global process we are testing, the loss of diversity out of Africa, from the specifics of heterozygosity in a particular population, here the French, which might bias D. In other words, we circumvent the possible issue that if French people carry an unusually high or low legacy, this in turn driving higher or lower heterozygosity, by using a measure of expected heterozygosity based on all non-European populations. The resulting trend is depicted in Figure 11.

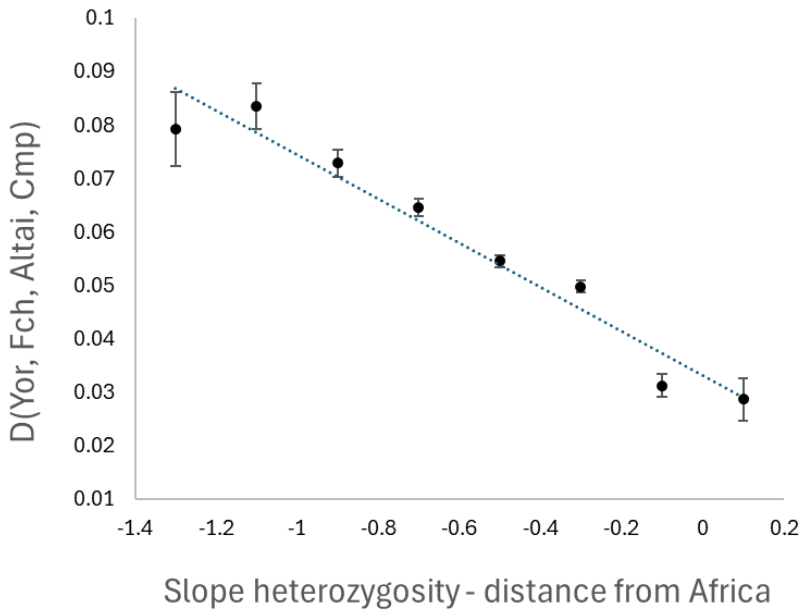

Figure 11. The relationships between heterozygosity and D. Data within every non-overlapping 50kb window across the genome were used to generate both ABBA and BABA counts for D(Yoruba,

French, Altai, chimpanzee) and an estimate of the slope relating heterozygosity to distance from East Africa. Slopes were binned (x-axis, slope = slope $*10^6$) and, within each bin, mean D +/- 1 s.e.m. plotted (y-axis). The strong, highly significant negative slope (Spearman rank correlation on the unbinned data, $S = 2.76 \times 10^{15}$, N = 258,750, $P << 2.2 \times 10^{-16}$) indicates that the greater the heterozygosity lost out of Africa (i.e., more strongly negative slope), the higher the value of D for that genomic region.

Our results indicated a highly significant negative correlation between D and the slope of the way heterozygosity declines with distance from Africa (Spearman's rho = -0.028, p-value = 2×10^{-8}), indicating that larger D values are strongly associated with greater loss of heterozygosity in populations farther from Africa. This finding supports the hypothesis that loss of heterozygosity is a driving force behind D patterns. Critically, while introgression might also impact heterozygosity, the resulting effect would likely be small, based on a legacy of about 2%, and in the 'wrong' direction: where sufficient introgressed material is present, if anything, it would tend to increase heterozygosity outside Africa, whereas the observed trend is for D to be highest where heterozygosity is unusually low due to more than average having been lost.

Discussion

Early anthropologists, such as Franz Weidenreich (1943) and Carleton S. Coon (1962), relied largely on physical traits and similarities between Neanderthal and human remains when they proposed that Neanderthals and anatomically modern humans might have coexisted, shared a recent common ancestor and possibly interacted. It was not until 2010, with the publication of the Neanderthal draft genome, that the quantification of possible introgression became feasible and claims were made that Eurasians carried an estimated 1–4% of Neanderthal-derived ancestry (Green et al. 2010). This tiny fraction will carry proportionately few functional variants likely to cause measurable phenotypic change (Graur et al. 2013). Despite this, a remarkable number of functional alleles and traits have been identified as being introgressed, all of which are in polygenic traits: Simonti et al. (2016) reported ~135,000 "high-confidence" Neanderthal single SNPs among the introgressed haplotypes and eight traits including overweight and depression; Dannemann and Kelso (2017) report associations for 11 phenotypes, including ease of skin tanning and being a morning or evening person; Wei et al. (2023) identified 235,592 "Neanderthal informative mutations" and 47 Neanderthal-associated phenotypes, including waist circumference and standing height. It stands to reason that the more genomic data and the more traits that are analyzed, the more links between modern traits and ancient origins will be found, as was the case with the COVID-19 risk alleles (Zeberg and Pääbo 2020). However, none of these studies considered alternative hypotheses. Interestingly, deviations from the predictions of introgression have been noted but were interpreted as a second introgression involving unknown 'ghost' hominin lineages (Hsieh et al. 2016; Durvasula and Sankararaman 2020) and even though some authors acknowledged the possibility of recurrent mutations, this idea was rejected without proper explanation (Durvasula and Sankararaman 2020).

Here we take a closer look at a key tool used to infer introgression, the statistic D. D is founded on two key assumptions: mutation rate constancy and the idea that recurrent mutations are vanishingly rare. The assumption of mutation rate constancy is so widely used in population genetics that it is often treated as a well-supported fact rather than a convenient null hypothesis with no empirical basis. In reality, mutation rate varies along chromosomes, between species and among sequence motifs in different human populations (Kong et al. 2012; Harris and Pritchard 2017; Seoighe and Scally 2017). Wherever the mutation rate is higher in one population compared with a second, the population with the lower rate will diverge less from related taxa, creating a false signal that mimics introgression. Conversely, if the mutation rate does not vary between populations, D is expected to give the same

value regardless of the underlying mutation rate. In practice, D varies from slightly negative up to around 0.2, depending on which class of mutations is used. This pattern cannot be driven by introgression because it would require a biologically implausible process whereby predominantly neutral introgressed fragments increase or decrease in frequency depending on which specific variants they carry. We confirmed that genuine introgression does not drive variation in D by choosing pairs of Yoruba individuals and replacing 2% of one genome with Neanderthal DNA. As expected, and in contrast to the pattern in real data, D for spiked Yorubans is above zero but does not vary significantly between motifs. Since introgression can be ruled out as a mechanism, the variation in D must be due to variation in the mutation rate of different motifs between human populations, as reported by Harris and Pritchard (2017)

The second assumption is that recurrent mutations are vanishingly rare. This assumption is strange given both how easy it is to test and the fact that so many SNPs in dbSNP carry three or even four different alleles. We formalise this distribution by counting the numbers of SNPs with two, three and four alleles in non-overlapping windows across the genome, finding that around 40% of SNPs carry more than two alleles. Turning to the HGDP data, which is more representative of sample sizes in studies testing for introgression, we find that sites carrying more than two alleles are 28 times as common as D-asymmetric sites. Moreover, since triallelic sites are created by one transition and one transversion, for each of these, there will be two 'silent' sites where the same transitions have occurred twice. Thus, triallelic sites are actually 84 times as common as the sites used to infer introgression: hardly negligible to the point where they can safely be ignored! In this light, it is not surprising that many of the traits reportedly associated with introgressed Neanderthal DNA harbor tri-allelic SNPs. For example, rs4580704, rs6858749, and rs11022775 in the genes CLOCK (Molina-Montes et al. 2022), RGS16 (Dashti et al. 2015), and ARNTL, respectively, are all associated with circadian regulation and are all tri-allelic (Min et al. 2024).

That mutation rate varies between populations and recurrent mutations are common opens up the possibility that false signals suggesting introgression can develop. Next, we sought to distinguish introgression from the MRVH by examining the number of mutations that can be inferred from informative sites. Under the introgression hypothesis, all sites involve just a single mutation, while the MRVH requires two. Exploiting the fact that the probability of a single mutation scales linearly with mutation rate while the probability of a recurrent mutation scales with mutation rate squared, we studied transition to transversion ratios. We confirmed that sites generated by a single mutaiton have a ratio of around two while sites likely to carry two mutations have a ratio of around $2^2 = 4$. Across D-informative sites, the ratio is 2.3. This is significantly above two, meaning that introgression cannot be the only mechanism responsible. However, interpreting the observed ratio in terms of a mixture of single mutation sites due to ILS and sites with recurrent mutations, we find that 10% sites would have to carry recurrent mutations. Since D-symmetric sites account for around 90% of all D-informative sites, these numbers leave little, if any, room for a signal due to introgression. To further explore this idea, we analyzed the mutation spectrum, which includes highly mutable XCG triplets that should be exceptionally prone to recurrent changes. We show that XCG sites exhibit an even more pronounced impact on D-statistics, lending further support to the idea that recurrent mutations play a pivotal role in driving these signals.

The TS:TV ratio analysis also uncovers an inadvertent observation bias. As stated, sites where the chimpanzee differs from all archaics and are also polymorphic in humans show a TS:TV ratio of around four, indicative of recurrent mutations. However, D conditions on the archaic and chimpanzee alleles are different, so the idea of enriching sites carrying derived archaic alleles was introduced. Unfortunately, this conditioning also enriches for sites with higher mutability where recurrent mutations are more likely to generate spurious signals mimicking introgression. In terms of inferring

introgression, this would not matter if recurrent mutations were indeed rare enough to be ignored, but with approaching sites carrying 100 recurrent mutations for each site that would be generated by introgression, this conditioning may make introgression more difficult to infer rather than easier.

As a further test, we considered the very nature of the branch length asymmetry. Under the introgression hypothesis, the non-African branch is shorter than expected due to it carrying Neanderthal DNA. However, this shortening should only be apparent when humans are compared against Neanderthals or their closest relatives. If humans are compared with chimpanzees, all branch lengths should be approximately equal: the geological time separation between any base sampled in humans and its homologues sampled in chimpanzees is the same, regardless of whether part of the history was spent in the Neanderthal lineage. Conversely, the MRVH suggests that the non-African branch is not shorter, but rather, the African branch is longer due to a higher mutation rate. Since here, the branch asymmetry is due to processes entirely within humans, it no longer matters which outgroup is used, the branch length asymmetry will be the same. We find that across 50kb genomic windows, there is an extremely strong correlation between the asymmetry seen when Neanderthals are the outgroup and the asymmetry when the chimpanzee is used as the outgroup. Moreover, this correlation cannot be replicated simply by replacing one human's DNA with 2% Neanderthal DNA. This demonstration that the branch length asymmetry is endogenous to humans again largely denies introgression whilst supporting the MRVH.

Interestingly, not only are the two asymmetries correlated, but also, the mean branch length to Africans is substantially longer. A value of D'_C of 0.5 indicates that, on average, any given site is three times as likely to carry a derived allele in the HGDP sample of Africans that is not present in the HGDP Europeans compared with carrying a derived allele in Europeans not present in Africans. We formalised this by calculating D'_X values based on the probability that any *given* allele in an African is derived and not found in the sample of non-Africans (Figure S3). Branch length asymmetry increases with the extent to which heterozygosity is higher in Africa for all three measures (D'_N, D'_C, D'_{N+C}, where D'_{N+C} indicates sites where the chimpanzee and Altai carry the same base), being more pronounced in order $D'_C > D'_{N+C} > D'_N$, consistent with the branch length asymmetry constituting a greater proportion of the shorter human – Neanderthal distance. Based on single alleles, D'_X is expected to reflect the underlying mutation rates (Mallick 2016; Amos 2021). Remarkably, the large positive values we report contrast strikingly with the marginally higher mutation rate *outside* Africa that Mallick et al. (2016) reported because they considered all variants, a large majority of which reflect mutations that *predate* the out-of-Africa event, i.e., their reported difference is effectively between two African populations. By requiring informative derived variants to be present in only one population, we greatly enrich for variants that were generated *after* African and non-African populations became distinct.

Having established that the MRVH is both plauisble and fits the data better than introgression, we next sought an independent test to determine whether it fits the two hypotheses based on the frequencies of the critical alleles in humans. Introgression predicts a dominant role for rare Neanderthal B alleles in non-Africans. The MRVH makes the diametrically opposite prediction that the signal will be carried by rare chimpanzee A alleles generated by recurrent mutations more frequently in Africans. By calculating D separately for sites where the human major allele is the chimpanzee A *versus* the Neanderthal B, we uncovered a remarkable pattern where D is either very large and positive, ~ 0.3, or large and negative, ~ -0.25. This double pattern is profoundly telling. The large positive values result when the major human allele is B, the exact opposite of what should result from introgression where, unless strongly positively selected, the B allele will essentially always be the minor allele. When the B allele is the minor allele, D is strongly *negative*, suggesting that Africans rather than non-Africans carry a legacy, one that is five times larger than current thinking ascribes to non-Africans.

Given that most opportunities for Neanderthals and humans to interact were in Eurasia, a huge African legacy is unlikely. A more parsimonious explanation is that non-zero D is driven primarily by recurrent mutations. Since differences between Africans and non-Africans have accrued over less than 100,000 years, these will be dominated by minor alleles. Consequently, partitioning the data according to which allele is the major allele has the effect of splitting the overall signal into a positive and a negative component. This is entirely consistent with the MRVH, where a single process acting on A versus B alleles would be expected to drive opposing trends. However, the strongly opposing signals are difficult to reconcile with introgression, where the expectation would be that the two components would comprise a non-zero D when the B allele is the minor allele and a value of D=0 when the A allele is the minor allele.

Having discovered that the overall typical reported value of D(Africa, Europe, Neanderthal, chimpanzee) = 0.05 reflects a balance between two much larger but opposing signals, we next sought a more detailed view of these components. For this, we omitted the normalisation step of D and focused just on D* the differences in numbers of ABBAs and BABAs, allowing us to visualise the relative contributions of different classes of site both in terms of major and minor alleles and across the mutation spectrum. We confirmed that sites that should dominate the signal under the introgression hypothesis actually contribute minimally to the overall signal. Conversely, sites expected to dominate under the MRVH can explain almost the entire observed signal. This reinforces a consistent and compelling narrative in which the data align far better with a model where Africans are less similar to Neanderthals due to recurrent mutations occurring at a higher rate in Africans, than a model where non-Africans are closer to Neanderthals due to introgression. These results are consistent with earlier findings showing that the D signal is primarily driven by heterozygous sites in African populations rather than heterozygous sites outside Africa (Amos 2020).

Previous studies have often sought to verify the effectiveness of statistics like D by testing on simulated data (e.g., Green et al. 2010; Durand et al. 2011). Unfortunately, the simulations themselves incorporate the same assumptions that we challenge. Here, we try a more direct approach by simply spiking the genome of an individual thought to lack introgressed material, for example, a Yoruban, with an appropriate amount of genuine Neanderthal sequences. This method should accurately capture features such as differences between the species in base composition or the relative mutability of motifs across the mutation spectrum without relying on the assumptions that we have shown to be false. We used this spiked genome approach to test a further key difference between introgression and the MRVH. Under introgression, the shorter non-African to Neanderthal branch length should only be seen when the Neanderthal (or close relative) is used as the test taxon. Conversely, the MRVH depends on a process that is endogenous to humans, so it should have little or no dependence on the test taxon. We show that the branch length asymmetry in humans does *not* depend on the test taxon but is instead similar and highly correlated across the genome when the Neanderthal is substituted for the chimpanzee. In spiked data simulating genuine introgression, the branch length asymmetry is radically different: being absent for negative values; steeper for positive values, reflecting the much stronger impact of Neanderthal-human distances, and, crucially, extending over a much narrower range of values. These results show that the branch length asymmetry captured by D reflects a difference in evolutionary rate between the African and non-African lineages and that branch-shortening through introgression plays a minimal role.

If the introgression hypothesis is not supported by the data, and D is instead driven by a higher mutation rate in Africans, an important question arises as to what mechanism might be responsible. Harris and Pritchard (2017) have already shown that the mutation rates of various triplet motifs differ significantly across different populations, providing a proximate process capable of driving non-zero D wherever the underlying mutation rate is sufficiently high. Building on this, we have shown how the

probability of a given triplet mutating is influenced by the level of heterozygosity in its flanking sequences, with some triplets becoming more mutable and other triplets less so as flanking heterozygosity increases (Amos 2019). Together, these processes offer a plausible mechanism that might be capable of driving D, where the loss of around 25% of heterozygosity out of Africa caused significant shifts in the relative mutability of different triplets Harris and Pritchard (2017). Since it is difficult to imagine how the genome could 'track' changes in mutability across the entire mutation spectrum and somehow modulate these so that the overall mutation rate remains constant, it seems certain that changes across the mutation spectrum will result in changes to the overall mutation rate, predicting a correlation between the relative amount of heterozygosity lost and D. Testing this, we find that, across the genome, the steeper the slope of the negative relationship between heterozygosity and distance from Africa, the larger the corresponding D value. This correlation cannot be due to introgression because introgressed material would tend to *increase* heterozygosity, the exact opposite of what is observed.

Overall, we present several lines of evidence that non-zero D in humans is driven by a higher mutation rate in Africa rather than by introgression, summarised in Table 1. It may be argued that D has never been viewed as a perfect measure and that evidence from other approaches, such as the inference of introgressed haplotypes, is far more reliable. We are not convinced by this view for two important reasons. First, the assumption that recurrent mutations are negligibly rare not only underpins D but is also central to methods for inferring haplotypes. Recurrent mutations blur the distinction between identity by descent rather than by state. Even if non-Africans do carry a 2% Neanderthal legacy, for each informative introgressed allele, seen as a D-asymmetric site, we estimate there will be approaching 100 recurrent mutations that are merely identical by state. Together with the rarity of D-asymmetric sites, only ~0.3 in a typical introgressed fragment that is 33kb long, this massive ambiguity more or less precludes the identification of all but exceptionally distinctive fragments.

Second, our analyses do not merely show that D does not work, in the sense that (in humans) its results cannot be interpreted purely in terms of measuring introgression, but rather that the signal captured by D is generated by a mechanism that is *not* introgression. If a Neanderthal legacy was present, there is no reason to believe that it would not be reflected in D, whether as a pure signal or as part of a mixed signal. We find no evidence that a signal compatible with introgression exists. Any Neanderthal alleles that entered humans around 50,000 years or 2,000 generations ago are unlikely to have reached a frequency above 10% without the aid of positive selection. Consequently, almost all introgressed B alleles will manifest as minor alleles rather than major alleles in humans. However, when we analyse such alleles as a separate class of sites, instead of greatly emphasising the signal, D is strongly *negative*, suggesting an implausibly large legacy inside rather than outside Africa. Alternative methods used to infer introgression must both account for this unexpected trend and show that they are robust to the mechanism that generates it.

Putting aside the specific problems with D, we show further that if a legacy does exist, it is far too dispersed across the genome. A 2% legacy implies that 98% of the genome will carry no legacy, yet we show that more than 11% of 50kb windows in individual genomes give non-zero D (after removing D-symmetric sites). As elsewhere, the standard model of introgression appears inadequate unless the humans carry five times more Neanderthal DNA than is widely reported. The importance of the way any legacy is distributed has not been given the attention is warrants. At the extremes, a 2% legacy could either be entirely fixed, leaving 98% of the human genome entirely free of Neanderthal material, while at the other, or all fragments occur at the same average frequency of 2%. The former is contradicted by maps of introgressed material, which suggest that 60-70% of the genome carries some legacy (Sankararaman et al. 2014) while the latter denies positive selection, raising some fragments to high frequency.

The last point we wish to emphasise is the difference between interbreeding and introgression. There appears to be good evidence that interbreeding occurred in the form of the discovery of hybrid skeletons (Fu et al. 2015; Slon et al. 2018). However, the fact that viable progeny were formed is a far cry from showing that a detectable downstream legacy resulted. These hybrid individuals may well have been infertile or socially rejected, and even if they did manage to reproduce, no one has yet found either Neanderthal Y chromosomes or Neanderthal mitochondrial DNA in humans. By implication, there was a strong selection against these components that would have reduced fitness and created a significant barrier to Neanderthal DNA increasing in frequency. The recent sequencing of the Neanderthal Thorin from Grotte Mandrin in Mediterranean France (Slimak et al. 2024) supports this position. Thorin belonged to a population with a small group size that showed no genetic introgression with other known late European Neanderthals. Compared to other west Eurasian Neanderthals, Thorin also does not show excess allele sharing with modern humans, and the possibility of recent interbreeding with early modern humans at Grotte Mandrin Cave1 was ruled out. The authors commented that the genetic and cultural evidence suggests the possibility of the presence of multiple genetically isolated Neanderthal lineages over millennia, suggesting that the absence of interbreeding between Neanderthals and humans should not be surprising as interbreeding even between different lineages within Neanderthals was uncommon.

Recently, Sümer et al. (2024) sequenced six ancient humans from ~45,000-year-old remains from Germany and calculated their interbreeding with Neanderthals using IBD. During this time (60,000-43,000), Neanderthals were the predominant human species in Europe (Richter 2016), so it was expected that this group would have experienced multiple interbreeding events, as the authors noted. Surprisingly, the authors calculated the "standard" 2.9% Neanderthal ancestry (union of 95% CIs: 2.8-3.0%) for high- and low-coverage genomes with similar estimates for three other low-coverage genomes (union of 95% CIs: 2.7-3.3%). Even more surprisingly, when estimating the number of interbreeding events, the data best fitted a single event, presumably taking place 56-98 generations earlier. Even the authors did not deem this finding realistic and proposed a more realistic model (unsupported by their data) based on continued interbreeding over multiple generations. We agree with the authors that their results do not support the expectation under introgression. However, rather than proposing a model that fits received wisdom rather than the data, we would interpret the findings as further evidence for the MVRH, which explains both the low and temporally stable "Neanderthal ancestry" and the fitting of a "single interbreeding event." Interestingly, Tournebize and Chikhi (2024) also showed that commonly used interbreeding detection approaches identified long Neanderthal fragments and an admixture event that never occurred. Overall, our findings show that MRVH resolves conflicting findings in the literature regardless of the methods used to infer inbreeding.

Finally, it is worth noting the challenge that the current narrative poses to classical theory, where successful interbreeding appears to have occurred repeatedly, almost whenever and wherever species of hominin met, including the need to invoke yet-to-be-discovered species. Not only does this suggest minimal genetic barriers, which is at odds with the strong selection that would have been needed to remove both introgressed mitochondrial and Y-chromosome sequences, but there is also a contradiction in terms of speciation. Speciation is generally thought to be difficult in the face of gene flow, yet hominin species diversity is high. How did speciation occur, often sympatrically, if the groups involved interbred so freely and successfully whenever they met, and what prevented such a successful process from carrying on to erase species differences? In this light, the MRVH offers a generic explanation in which variation in heterozygosity can drive signals suggestive of introgression wherever they occur, whether the variation in heterozygosity is genome-wide and driven by demography, the result of strong selection acting on specific genomic regions.

In conclusion, our study challenges the prevailing assumption that the genetic signal captured by D-statistics provides robust evidence for Neanderthal introgression into modern human genomes. While interbreeding may have occurred, the absence of compelling genetic evidence, such as Neanderthal Y chromosomes or mitochondrial DNA in modern humans, underscores the difficulty of establishing a detectable legacy of such events. This lack of evidence, coupled with our findings that genetic patterns attributed to introgression are much better explained by mutation rate variation, calls for a reassessment of the introgression hypothesis. The idea that mutation rate variation linked to changes in heterozygosity can drive detectable variation in branch lengths between species could help explain the implausible ubiquity with which interbreeding between hominins is inferred, apparently regardless of where and when the species lived. The most parsimonious evolutionary scenario that emerges from our study is that humans and Neanderthals diverged separately and at different times from their common ancestor, possibly *Homo heidelbergensis* or a related hominin, who lived in Europe and Africa some 600,000–200,000 years ago or a related species like *Homo* antecessor (Bermúdez de Castro et al. 2017), with the European branch giving rise to Neanderthals quite early on and the African branch giving rise to humans some 200,000 years ago, according to the ages of the most ancient Y (Elhaik et al. 2014) and mtDNA (Cyran and Kimmel 2010) haplogroups. Humans and Neanderthals met in Europe and experienced limited interbreeding, but introgression, if it occurred, was negligible.

We share the scientific community's enthusiasm for uncovering connections between modern humans and other hominins. The question of whether Neanderthals represent a separate species or a divergent branch of the same species remains a profound and complex issue. However, the weight of genetic evidence increasingly supports the view that Neanderthals were a distinct species, reproductively isolated from *Homo sapiens*, and that the evolutionary tale of humans should be sought with other hominins. Our findings not only refine our understanding of human evolutionary history and resolve conflicts in the literature but also lend credence to the broader notion of Neanderthal distinctiveness.

Methods

Humans and Neanderthals are sufficiently closely related that most of the sites contributing to D are generated by incomplete lineage sorting (ILS) (Durand et al. 2011). To compare introgression with the MRVH, we test the assumptions on which D is based. We then test a series of scenarios where the two hypotheses give opposing predictions and test these using the HGDP deep-resolution dataset (Bergström et al. 2020), which includes the chimpanzee and three archaic hominins (Altai Neanderthal, Vindija Neanderthal, and Denisovan), unless specified otherwise. We consider the distribution of any legacy across the genome, how many mutations have occurred at D-asymmetric sites, which alleles are common and rare, and we compare real data with introgression simulated by ‘spiking’ African genomes with Neanderthal bases.

To study the distribution of multi-allelic SNPs, we downloaded the data from dbSNP (build 156) (https://ftp.ncbi.nih.gov/snp/redesign/latest_release/VCF/ last accessed 9/9/2024) and counted the numbers of SNPs with one, two and three alternate alleles in 50kb windows across the autosomes using Python 3.8.0 scripts. Conservatively, we analyzed only common SNPs, defined as SNPs with at least one 1000 Genomes Population with a minor allele frequency ≥ 1% for which two or more founders contribute to that minor allele frequency. Consequently, we repeated this analysis on the high-coverage HGDP data, which may be considered a typical dataset used to test for introgression. To replicate the same filtering criteria with the much smaller HGDP populations, we retained only positions where at least one population of French, Yoruba, Chinese, and Japanese had minor alleles of frequency >= 1%.

Data and script availability

All our data and scripts that can be used to replicate our results and figures are available via GitHub (https://github.com/eelhaik/UnexpectedDtourAhead).

Acknowledgment

EE was partially supported by the Swedish Research Council (2020-03485) and Erik Philip-Sörensen Foundation (G2020-011). The computations were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at Lund, partially funded by the Swedish Research Council through grant agreement no. 2018-05973.

Conflict of Interests

EE is a co-owner of Anath Genomic Consultants AB.

Contributions

WA conceived the study. Both authors carried out the analyses and wrote the paper.

References

Amos W. 2013. Variation in Heterozygosity Predicts Variation in Human Substitution Rates between Populations, Individuals and Genomic Regions. PLOS ONE. 8:e63048.

Amos W. 2019. Flanking heterozygosity influences the relative probability of different base substitutions in humans. R. Soc. Open Sci. 6:191018.

Amos W. 2020. Signals interpreted as archaic introgression appear to be driven primarily by faster evolution in Africa. R. Soc. Open Sci. 7:191900.

Amos W. 2021. Correlated and geographically predictable Neanderthal and Denisovan legacies are difficult to reconcile with a simple model based on inter-breeding. R. Soc. Open Sci. 8:201229.

Amos W, Bryant C. 2011. Using human demographic history to infer natural selection reveals contrasting patterns on different families of immune genes. Proceedings of the Royal Society B: Biological Sciences. 278:1587-1594.

Bergström A, et al. 2020. Insights into human genetic variation and population history from 929 diverse genomes. Science. 367:eaay5012.

Bermúdez de Castro JM, et al. 2017. Twentieth anniversary of Homo antecessor (1997-2017): a review. Evolutionary Anthropology: Issues, News, and Reviews. 26:157-171.

Bonfante B, et al. 2021. A GWAS in Latin Americans identifies novel face shape loci, implicating VPS13B and a Denisovan introgressed region in facial variation. Science Advances. 7:eabc6160.

Brown WM, et al. 1982. Mitochondrial DNA sequences of primates: Tempo and mode of evolution. J. Mol. Evol. 18:225-239.

Chintalapati M, Moorjani P. 2020. Evolution of the mutation rate across primates. Curr. Opin. Genet. Dev. 62:58-64.

Coon CS. 1962. The origin of races. New York: Alfred A Knopf.

Cyran KA, Kimmel M. 2010. Alternatives to the Wright–Fisher model: The robustness of mitochondrial Eve dating. Theor. Popul. Biol. 78:165-172.

Dannemann M, Kelso J. 2017. The Contribution of Neanderthals to Phenotypic Variation in Modern Humans. The American Journal of Human Genetics. 101:578-589.

Dashti HS, et al. 2015. Habitual sleep duration is associated with BMI and macronutrient intake and may be modified by CLOCK genetic variants234. The American Journal of Clinical Nutrition. 101:135-143.

Durand EY, et al. 2011. Testing for ancient admixture between closely related populations. Mol. Biol. Evol.

Durvasula A, Sankararaman S. 2020. Recovering signals of ghost archaic introgression in African populations. Science Advances. 6:eaax5097.

Elhaik E, et al. 2014. The ‘extremely ancient’ chromosome that isn’t: a forensic bioinformatic investigation of Albert Perry’s X-degenerate portion of the Y chromosome. Eur. J. Hum. Genet. 22:1111-1116.

Fu Q, et al. 2015. An early modern human from Romania with a recent Neanderthal ancestor. Nature.

Graur D, et al. 2013. On the immortality of television sets: "function" in the Human genome according to the evolution-free gospel of ENCODE. Genome Biol. Evol. 5:578-590.

Green RE, et al. 2010. A draft sequence of the Neandertal genome. Science. 328:710-722.

Halldorsson BV, et al. 2019. Characterizing mutagenic effects of recombination through a sequence-level genetic map. Science. 363.

Harris K. 2015. Evidence for recent, population-specific evolution of the human mutation rate. Proc. Natl. Acad. Sci. U.S.A. 112:3439-3444.

Harris K, Pritchard JK. 2017. Rapid evolution of the human mutation spectrum. eLife. 6:e24284.

Hsieh P, et al. 2016. Model-based analyses of whole-genome data reveal a complex evolutionary history involving archaic introgression in Central African Pygmies. Genome Res. 26:291-300.

Huerta-Sánchez E, et al. 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature. 512:194-197.

Jacobs GS, et al. 2019. Multiple Deeply Divergent Denisovan Ancestries in Papuans. Cell. 177:1010-1021.e1032.

Kong A, et al. 2012. Rate of de novo mutations and the importance of father's age to disease risk. Nature. 488:471-475.

Liu Z, Samee Md Abul H. 2023. Structural underpinnings of mutation rate variations in the human genome. Nucleic Acids Res. 51:7184-7197.
Mallick S. 2016. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature.
Meyer M, et al. 2012. A high-coverage genome sequence from an archaic Denisovan individual. Science. 338:222–226.
Min Q, et al. 2024. The correlation of shift work and CLOCK, BMAL1, and PER1 gene polymorphisms with hypertension in healthcare workers: A cross-sectional study. Medicine (Baltimore). 103.
Molina-Montes E, et al. 2022. Circadian clock gene variants and their link with chronotype, chrononutrition, sleeping patterns and obesity in the European prospective investigation into cancer and nutrition (EPIC) study. Clin. Nutr. 41:1977-1990.
Nesta AV, et al. 2021. Hotspots of Human Mutation. Trends Genet. 37:717-729.
Peter BM. 2016. Admixture, Population Structure, and F-Statistics. Genetics. 202:1485-1501.
Phillips C, et al. 2020. A compilation of tri-allelic SNPs from 1000 Genomes and use of the most polymorphic loci for a large-scale human identification panel. Forensic Sci. Int. Genet. 46.
Prufer K, et al. 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature. 505:43–49.
Prugnolle F, et al. 2005. Geography predicts neutral genetic diversity of human populations. Curr. Biol. 15:R159-160.
Richter J. 2016. Leave at the height of the party: A critical review of the Middle Paleolithic in Western Central Europe from its beginnings to its rapid decline. Quaternary International. 411:107-128.
Rogers AR, Bohlender RJ. 2014. Bias in Estimators of Archaic Admixture. arXiv preprint arXiv:1412.6691.
Rogozin IB, Pavlov YI. 2003. Theoretical analysis of mutation hotspots and their DNA sequence context specificity. Mutation Research/Reviews in Mutation Research. 544:65-85.
Sankararaman S, et al. 2014. The genomic landscape of Neanderthal ancestry in present-day humans. Nature. 507:354-357.
Sankararaman S, et al. 2016. The combined landscape of Denisovan and Neanderthal ancestry in present-day humans. Curr. Biol. 26:1241-1247.
Segurel L, et al. 2014. Determinants of mutation rate variation in the human germline. Annu. Rev. Genomics Hum. Genet. 15:47-70.
Seoighe C, Scally A. 2017. Inference of Candidate Germline Mutator Loci in Humans from Genome-Wide Haplotype Data. PLoS Genet. 13:e1006549.
Simonti CN, et al. 2016. The phenotypic legacy of admixture between modern humans and Neandertals. Science. 351:737-741.
Skov L, et al. 2020. The nature of Neanderthal introgression revealed by 27,566 Icelandic genomes. Nature.
Slimak L, et al. 2024. Long genetic and social isolation in Neanderthals before their extinction. Cell Genomics. 4.
Slon V, et al. 2018. The genome of the offspring of a Neanderthal mother and a Denisovan father. Nature. 561:113-116.
Sümer AP, et al. 2024. Earliest modern human genomes constrain timing of Neanderthal admixture. Nature.
Tournebize R, Chikhi L. 2024. Ignoring population structure in hominin evolutionary models can lead to the inference of spurious admixture events. Nat. Ecol. Evol.
Wei X, et al. 2023. The lingering effects of Neanderthal introgression on human complex traits. eLife. 12:e80757.
Weidenreich F. 1943. The "Neanderthal Man" and the ancestors of "Homo sapiens". Am. Anthropol. 45:39-48.
Zeberg H, Pääbo S. 2020. The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. Nature. 587:610-612.

Supplementary materials

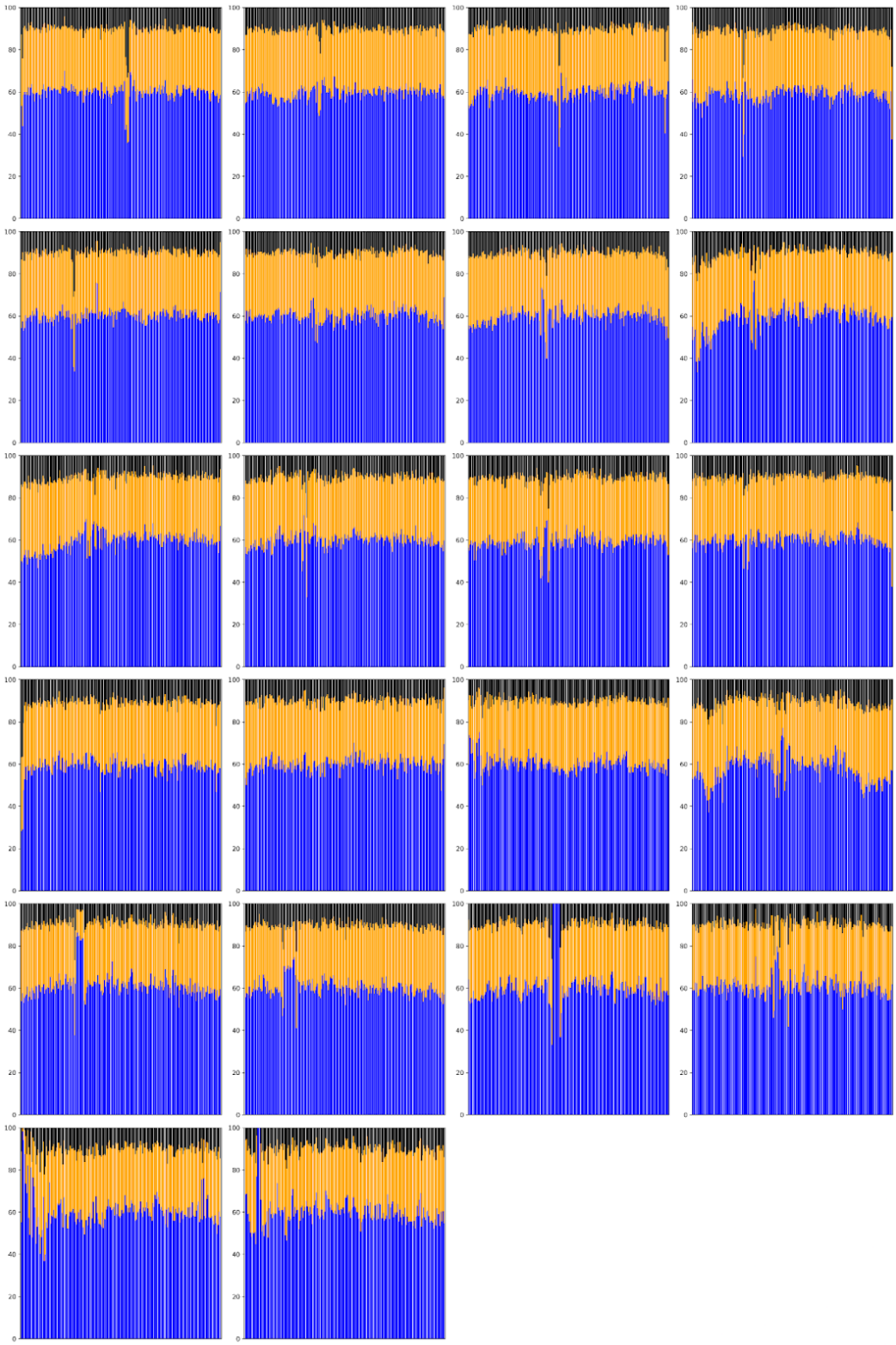

Figure S1. **Numbers of SNPs carrying two, three and four alleles in non-overlapping 50kb windows along human chromosome 1-22 (left to right, top to bottom).** The y-axis shows the percentage of variable SNPs carrying two (blue), three (orange), and four (black) different bases in 50kb non-overlapping windows (x-axis) along each chromosome 1.

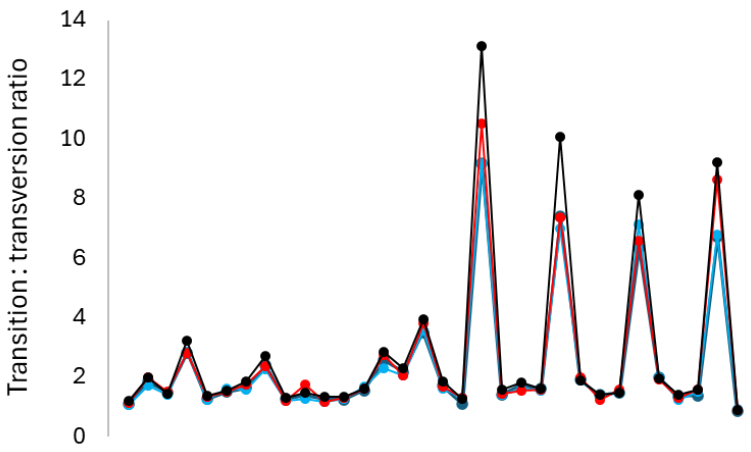

Figure S2. Variation in transversion : transition ratios across the mutation spectrum. The widely cited ratio of approximately two is the average across all sequences. We partitioned data according to the mutating triplet and for each calculated the TS:TV ratio for three scenarios where a single mutation is likely to have occurred: from Table 2, AABAA (red), AAABA (blue), PAAAA (black). As seen, the ratios at the XCG triplets are consistently extreme, peaking at nearly 14 for ACG at polymorphic sites in humans, while several triplets have ratios nearer to one. Triplet order is as in Figure 6.

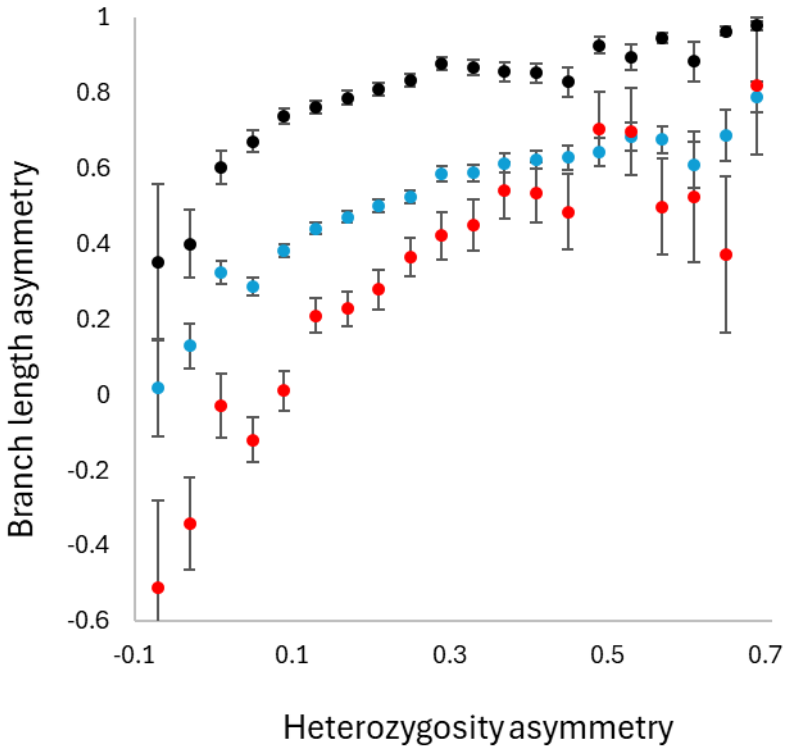

Figure S3. Relationship between African – non-African asymmetry in heterozygosity and the size of the branch length asymmetries involving the chimpanzee and Neanderthals. The autosomal genome was divided into non-overlapping 50kb windows. Within each, we calculated heterozygosity in both Yorubans and French and calculated the asymmetry in heterozygosity as $(Het_{YOR} – Het_{FRC})/(Het_{YOR} + HetF_{RC})$, where Het_{YOR} and Het_{FRC} are the heterozygosities in Yoruba and French respectively. To measure branch length asymmetries, we focussed on biallelic sites where either Yorubans or French were fixed for one allele, A, and where the other group was polymorphic, A/B. Within each window, the frequencies of the B alleles were summed at three classes of site: 1) Altai = chimpanzee = A (blue); 2) Altai = A, chimpanzee = B (black); 3) Altai = B, chimpanzee =A (red). Binning by heterozygosity asymmetry (x-axis), the mean branch length asymmetries are plotted on the y-axis +/- 1 s.e.m., asymmetries calculated as $(fB_{YOR} – fB_{FRC})/(fB_{YOR} – fB_{FRC})$, where fB_{YOR} is the sum of the frequencies of the B alleles in Yorubans. As seen, as the asymmetry in heterozygosity (almost always positive = higher heterozygosity in Yorubans) is greater, so too are the branch length asymmetries in favour of relatively longer African branches. The greatest branch length asymmetries are seen with respect to the Altai because, although the branch length differences are similar, these make up a relatively larger proportion of the shorter human–Neanderthal distance.