**Is the idea that many humans carry some Neanderthal DNA correct?**

by Professor William Amos

*Bill Amos is Professor of Evolutionary Genetics in the Department of Zoology, Cambridge University. During his career, Bill has worked on many diverse topics including uncovering the breeding behaviour of marine mammals and the way heterozygosity impacts individual fitness in natural populations. He also became increasingly interested, not only in the use of DNA sequences as markers, but also in how these sequences evolve, leading to his idea that the rate of evolution scales with population size. It was this idea that provided an alternative explanation to introgression for why some humans are genetically closer to Neanderthals than others, a subject he has researched for about a decade now. The following a is a summary of numerous different analyses he has conducted that together put a huge question mark over the notion that 1-2% of non-African DNA is inherited from Neanderthals. Bill has published widely, with over 160 peer-reviewed publications including many in the highest ranked journals such as Nature, Science, Nature Genetics, PLoS, Current Biology and PNAS.*

***Introduction***

The idea that humans who left Africa to colonise the rest of the world mated with Neanderthals and, as a result, now carry a lasting genetic legacy has progressed rapidly from shocking revelation (Green et al., 2010) to widely accepted fact. There are now many tens if not hundreds of scientific papers all reporting or analysing the same pattern (Rinker et al., 2020; Skov et al., 2020). However, there are issues, including the basic way science was conducted. To some extent, this appears to be a story that is too interesting or too attractive to challenge and, as such, became written in stone almost as soon as it appeared in a high-profile publication. Since then, any form of counter-evidence has been extremely difficult to publish and has been met with fierce rejection by the community who have by now invested more than a decade of work on the assumption the story is true. Here I will try to outline a number of problems with the story and encourage readers not to take the accepted narrative at face value. I hope that some readers will try replicating and or extending what I have done to see for themselves how the accepted narrative crumbles under closer scrutiny. I have tried to write this piece in a way that will be reasonably accessible to non-geneticists, though I appreciate that this is quite a challenge, and in this light have included a brief glossary of key terms at the end. Constructive feedback is welcome!

***Inter-breeding versus a lasting genetic legacy***

It is useful to begin by making a distinction between inter-breeding and a lasting genetic legacy. Hybridisation between species does occur in nature but it is not common for good reason. In many cases, no viable progeny are produced but, even if they are produced, they are often sickly and / or infertile. In social species like primates, there is also the question of whether hybrid offspring would themselves be attractive and find mates. Thus, while human – archaic hybrid skeletons have been reported (Fu et al., 2015; Slon et al., 2018), there is no direct evidence that they left appreciable numbers of progeny. Indeed, such finds tend to feature either one or a pair of skeletons found in a cave. These caves appear not to be general burial sites, because there are no other individuals, and the caves are unlikely to have been used for living, not least because most societies would not share their living space with the dead. This raises the possibility that these skeletons were subject to some form of ostracism, an attempt to hide or remove from view an individual who was seen as suspiciously different. Interestingly, since caves seem to provide one of few environments that allow skeletons from this period to survive, if unusual individuals were hidden in this way, the assumption that these individuals are representative of their populations could be badly misleading.

***How did hybrid offspring survive?***

Apart from the general observation that inter-breeding between species is rarely successful, there is also some level of direct evidence. Most of our genes are inherited as two copies, one from the mother and one from the father. There are two exceptions. First, the Y chromosome is only carried by males, so is always passed from father to son. Second, mitochondrial DNA is passed only between a mother and her offspring, regardless of whether these are sons or daughters. Strangely, even though it is estimated that the Neanderthal legacy is around 2% of DNA in non-African humans, no one has yet found either a Neanderthal Y chromosome or a Neanderthal mitochondrial sequence in modern humans, despite the huge numbers from across the world who have had their DNA tested. By implication, the Neanderthal versions of both the Y chromosome and mitochondrial DNA must, to some extent, be toxic in humans. Every time any inter-breeding event occurred, any individuals born with either of these elements must have either died or at least been appreciably less likely to reproduce. This may sound fairly trivial, but it raises considerably the number of hybrid offspring that would have been needed to create any given size of modern legacy.

***Now the Real Problem: There are two hypotheses to test, not just one!***

Putting aside questions relating to offspring fertility and viability, it is worth taking a moment to appreciate that the basic experimental design used to infer introgression was, and continues to be, flawed. In general, if we measure the genetic distance between one species and two different populations of a second, closely-related species, we expect the two distances to be the same. However, this null expectation is found to be wrong in the case of Hominins, where Neanderthals are clearly genetically closer to non-Africans than they are to Africans. There are two possible explanations for this pattern. One option is that the Neanderthal to non-African distance is unexpectedly small, perhaps because non-Africans have inherited some Neanderthal DNA via inter-breeding. The other option is that the Neanderthal to African distance is unexpectedly large, for example because Africans have a higher mutation rate that has driven faster divergence. Prior to 2010, both these options would have been considered highly unlikely, yet (at least) one must be true.

When faced with a null hypothesis and two possible alternative hypotheses, the correct scientific approach is to treat the two alternative hypotheses as being equally likely and to try to find and test scenarios where the two alternative hypotheses give contrasting predictions. In practice, the original paper claiming introgression states explicitly that mutation rate is assumed to be constant (Green et al., 2010). This is remarkable because, by making this assumption, the authors completely dismiss one of the two possible hypotheses without any form of testing. I would argue that this is indefensibly poor science particularly, as I go on to discuss below, because there are several simple ways to tell the two apart. Furthermore, having ignored one of the two hypotheses, a second statistical *faux pas* follows quickly because the authors then equate rejection of the null hypothesis, i.e. that all humans are equally distant from Neanderthals, with proof that the only alternative hypotheses they consider, i.e. inter-breeding, is correct.

In fact, even if we put aside the issue that only one of two possible alternative hypotheses are tested, there are still many ways to check the validity of the conclusion that inter-breeding occurred. For example, if introgression occurred around 50,000 years ago and if most introgressed fragments that have been retained are approximately selectively neutral, the expected frequency distribution is easy to determine (very few introgressed fragments should exceed 10% in frequency). Perhaps surprisingly, I struggle to find instances in the literature where a serious attempt has been made to explore how this and other predictions from the inter-breeding hypothesis fit with real data. It remains unclear whether closer scrutiny was thought not to be needed or whether the results undermined the desired story and were tucked neatly under a nearby carpet.

***How specific is the signal interpreted as evidence for introgression?***

The first clear distinction between the two alternative hypotheses concerns where the signal comes from. In the inter-breeding hypothesis, the signal comes from *outside* humans in the form of genetic material that entered humans from Neanderthals and is therefore specific to Neanderthals. Conversely, under the mutation rate difference hypothesis, the signal comes from *within* humans, reflecting a difference in mutation rate between Africans and non-Africans. To test which hypothesis fits better, all that needs to be done is to repeat the analysis, substituting the Neanderthal sequences with equivalent sequences from other species like a great ape, where successful interbreeding could not have occurred. If the reported signal really is due to humans carrying Neanderthal DNA, then replacing the Neanderthal with another species should remove the signal. However, if the signal is due to mutation rate variation *within* humans, then the Neanderthal is acting merely as a passive outgroup and substituting in a great ape will have rather little impact. In fact, when I tried replacing the Neanderthal with, variously, the bonobo, chimpanzee and gorilla, the signal persisted. Although this has yet to be published, this analysis is easy to replicate (please do!) and provides powerful evidence that the signal is by no means unambiguously due to a Neanderthal legacy. Certainly, while one can argue the toss whether humans could have inter-bred successfully with Neanderthals, I would hope that few would try to push a narrative where inter-breeding took place between humans and several different species of apes!

***Are putative Neanderthal alleles in humans generally heterozygous?***

It is estimated that non-Africans carry about 2% Neanderthal DNA. It is also thought that this legacy, if it exists, is spread across much of the genome (Sankararaman et al., 2014, 2016), suggesting that a typical Neanderthal fragment will occur at a frequency of around 2-4%. This agrees well with simulated data and theoretical expectations of the average frequency of a fragment that entered humans around 50,000 years ago. Now, the probability that a person inherits the same fragment from both their mother and their father is the square of that fragment’s frequency. Thus, if a fragment has a frequency of 5%, the probability of finding a person who carries two copies = 0.05 x 0.05 = 0.0025. In other words, if introgression has occurred, more than 99% of Neanderthal DNA fragments in humans should be found as heterozygotes, with the person having one Neanderthal copy and one human copy. Almost no one will inherit the same Neanderthal fragment from both their parents. The result is the strong prediction that a true signal of introgression will be carried almost entirely by regions of the genome that are heterozygous.

To test this prediction, I reanalysed a large dataset three times. In the first run I included all informative sites across the genome, thereby repeating the original study. In the second run I did the same except I excluded all DNA bases if they were heterozygous in Africans. Finally, for the third run I repeated but this time excluding all sites that are heterozygous in non-Africans. Under the introgression hypothesis, run 3 should give almost no signal because the bases expected to carry this signal have all been excluded. Conversely, under the mutation rate variation hypothesis, the signal should be driven primarily by new mutations in Africans acting to increase the rate of evolution. In exactly the same way that introgressed fragments tend to be rare, so too will these new mutation in Africans. Consequently, under the mutation rate variation hypothesis, it is run 2 where the signal will be killed. In practice, the signal is unaffected by excluding heterozygous sites in non-Africans, but it is killed almost completely when heterozygous sites in Africans are excluded (Amos, 2020). This analysis provides strong support for the mutation rate variation hypothesis and at the same time suggests that any signal due to introgression, if present, is very weak or even negligible.

***What are the characteristics of sites that contribute to the signal?***

The signal interpreted by many as evidence for introgression is often expressed as the statistic D (Durand et al., 2011; Green et al., 2010). Consider aligned DNA sequences from, in order, an African human, a non-African human, the Neanderthal and the chimpanzee. We focus only on sites where there are only two alleles, which we can call A and B, and where both the humans differ and also the Neanderthal and chimpanzee bases differ. These requirements create two types of site, ABBA and BABA, depending of which of the humans matches the Neanderthal and which matches the chimpanzee. The introgression statistic, D, is calculated as the difference in number of these two types of site divided by the total number of sites found, i.e. (ABBA – BABA)/(ABBA + BABA). As such, the statistic D varies between -1 and 1 with, by convention, positive values indicating that more Neanderthal bases matching the non-African than match the African.

The mathematics behind D have been derived (Durand et al., 2011), meaning that we understand how it should behave depending on the time humans and Neanderthals split or the amount of introgression that occurred and when this took place. However, all this theory depends on the validity of the key underlying assumption that mutation rate is constant and that recurrent mutations occur at a negligible rate. Consequently, if mutation rate is not constant or if recurrent mutations are not so rare, then the ability to translate any given value of D into ranges of values for, for example, the timing and amount of introgression, is lost. Nonetheless, D still acts as a simple and useful measure of the extent to which Neanderthals are closer to one human population than a second, even if the interpretation of that difference remains open to question. As such, it is worth looking deeper to explore the origins of why D varies and by how much.

We have already discussed the fact that introgressed fragments are mostly expected to be rare. More generally, the two hypotheses give contrasting expectations for frequencies of alleles that contribute to D. Under introgression, the Neanderthal B allele is expected to be rare outside Africa and absent inside Africa. Equally, under mutation rate variation we expect D to be driven mainly by sites where the common BBBA state (one mutation between Hominins and the chimpanzee) mutates to create an ABBA through a B->A recurrent / back mutation in Africa. In terms of allele frequencies, this means that the chimpanzee A allele is rare in Africa and absent outside Africa. The two hypotheses therefore give entirely converse predictions. To determine how these predictions fit with real data, I systematically classified each informative site across the genome into one of five different categories:

(1) all non-African alleles are B, Africans carry A and B

(2) all African alleles are B, non-Africans carry both A and B

(3) all non-African alleles are A, Africans carry A and B

(4) all African alleles are A, non-Africans carry both A and B

(5) both Africans and non-Africans carry A and B.

I then calculated five versions of D, each time excluding sites belonging to a different category. I find that excluding category (4) sites only reduces D a little (by about 15%), even though this category should include the overwhelming majority of all sites generated through introgression. In contrast, excluding category (1) sites reduces D to, and indeed a little below, zero. Category (1) sites are the type expected to drive D under the mutation rate variation hypothesis. This work is currently being written up.

***Global variation in the inferred Neanderthal legacy***

One of the many remarkable aspects of the introgression story is the fact that all non-Africans are found to carry rather similar legacies, regardless of where in the world they live now or where their longer-term ancestry is rooted. By implication, all modern humans are descended from a rather homogeneous population of humans who inter-bred with Neanderthals, with no dispersal prior to the hybridisation event(s). Not only this, but the evenness of the size of the legacy across the world is at odds with likely very variable levels of overlap between Neanderthals and humans. This creates a strange contradiction. On the one hand, introgression statistics paint a picture where humans interbred with archaics almost wherever and whenever they encountered each other. However, if this really is the case, we would expect much more variation in the sizes of the legacies inferred, with humans descended from populations that coexisted with Neanderthals for longest having much larger legacies than those who overlapped less.

Archaeological evidence indicates that Neanderthal populations were spread mainly across central southern Europe (Banks et al., 2008; Serangeli & Bolus, 2008), implying that the highest legacies should be found in western Eurasia. However, the reality is very different, with inferred legacies increasing from West to East (Wall et al., 2013). Indeed, the more general pattern seems to be one where legacy size increases with distance from Africa, peaking in areas such as Oceania, north and eastern Siberia and even America, where there is no evidence that Neanderthals ever lived (Amos, 2021). This pattern is intuitively strange, given where most evidence for Neanderthals has been found, and is particularly puzzling in light of simulated data. I simulated an inter-breeding event soon after humans left Africa and then allowed the population to spread out across the world. What I found was that the inherited legacy is extraordinarily stable in size, meaning that the legacy essentially did not vary across the world. This makes intuitive sense. Any legacy will be spread across 22 chromosomes in the first generation and each of these will be continually broken down by recombination, yielding hundreds of different fragments within a few generations and thousands before any significant dispersal occurred. The legacy therefore becomes engrained in a way that is, by sheer weight of numbers, highly resistant to change.

The observed pattern of gently increasing inferred legacy with distance from Africa is, therefore, completely at odds with a simple model of introgression. Interestingly, this pattern does shadow the well-known way heterozygosity declines highly predictably with increasing distance from Africa (Prugnolle et al., 2005). As such, the trend fits well with one explanation I have put forward as to why mutation rate might vary. I have been working for a long time on the idea that mutations are more likely in the vicinity of heterozygous sites (Amos, 2011, 2013, 2016). If so, then then the large loss of heterozygosity that occurred during the out of Africa event would have caused a reduction in mutation rate and hence a reduction in the rate of divergence between non-Africans and related species like Neanderthals (see below).

***Do introgressed haplotypes prove introgression?***

One of the most frequent criticisms I encounter is that the evidence I present must be wrong because introgression has been proved through the identification of introgressed haplotypes. In fact, many of the methods used to infer introgressed haplotypes make the same unreliable assumption that underpins D, namely that mutation rate is constant. Often, the methods are said to have been ‘verified’ by simulations, but this represents dangerous wording. Simulations do not form some magic link to reality, they merely play out the scenario that is coded in the algorithm. All that is ‘verified’ is that, under one particular cartoon model of molecular evolution and introgression, the described methods reliably capture the signal of introgression generated. The simulations in no way prove that any signal detected in real data is genuine. In view of these important uncertainties, I have spent time examining individual high-profile papers where evidence for an introgressed haplotype in humans is presented. Here I present two examples and hope to show that these do not stand up to closer scrutiny.

Before considering the case studies, it is worth pointing out two important difficulties faced by attempts to identify introgressed fragments. First, the amount of useful information is much less than many would assume. For example, in round numbers, the original evidence for introgression was based on about 95,000 BABAs and 103,000 ABBAs, a difference of 8,000 across some 200,000 informative sites (Green et al., 2010). This study considered about 1Gb of aligned sequences, implying an average of one informative site every 125,000 bases. With a typical introgressed fragment expected to be around 50,000 bases long, this implies an average of less than half of an informative site per fragment: not very useful for estimating its presence, let alone length! Furthermore, for each of the 8,000 informative sites, there are around 190,000 sites that are ostensibly identical and therefore muddy the waters.

The second issue is a phenomenon called incomplete lineage sorting or ILS. When speciation occurs, the gene trees from one species are split into two. Immediately following speciation there will be many branches in one species that by chance join to a branch in the other. As time passes, new lineages are born while others go extinct. Consequently, the number of branches that root in the ‘wrong’ species decreases progressively until all branches in a species join to other branches within the same species. On an evolutionary timescale, humans and Neanderthals separated relatively recently, so many human branches still root on the Neanderthal tree. This phenomenon is called incomplete lineage sorting and has the unfortunate impact of mimicking introgression. A measure of the extent of ILS in humans is given by the ABBA/BABA counts. If we ignore recurrent mutations, all ABBAs and BABAs must be the result either of introgression or of ILS. So, taking the data of Green et al. (Green et al., 2010) at face value, the 190,000 ABBAs and BABAs mentioned in the previous paragraph would have been generated by ILS.

Given how common ILS is, the key question now becomes how to tell apart genuinely introgressed fragments and fragments created by ILS. Individual bases are indistinguishable, so clues have to be gleaned from the way contiguous polymorphic bases define larger fragments. There are two main sources of information that may be leveraged to try to distinguish. First would be to infer an origin via introgression by showing that a fragment is too similar to the archaic to have diverged more than the 50,000 years or so since it could have entered humans. This option is arguably too difficult to be of use because any measured level of genetic similarity only make sense in light of the local mutation rate, and this is both highly variable along a chromosome and extremely difficult to measure. Moreover, unless the sequenced Neanderthal was very closely related to the one who mated with a human, the nominal 50,000 year separation would be larger, possibly much larger, blurring the difference between introgression and ILS.

Arguably a more promising approach would be to show that the fragment is too long to be anything other than a recent addition to the human genome. Here, the idea is that recombination acts continually to break DNA sequences into shorter and shorter pieces. Like mutation rate, it is still necessary to pay attention to the fact that recombination rates vary greatly along a chromosome (Halldorsson et al., 2019a; McVean et al., 2004), with hot-spots and cold-spots that likely come and go over timescales comparable to human evolution (Ptak et al., 2005). However, more so than with mutation rates, reasonably accurate recombination maps are becoming available (Hinch et al., 2011). The following two examples illustrate how these issues play out in published claims for introgressed haplotypes in humans.

*a)* *A Neanderthal haplotype linked to the response to COVID.*

This story was published by Zeberg and Pääbo and relates to a ~50Kb fragment (Zeberg & Pääbo, 2020, 2021). The primary argument they use is that the haplotype they consider is too long to be anything but a recent addition to humans. However, they use as their reference recombination map one published as long ago as 2002 (Kong et al., 2002). By modern standards, this map is extremely crude. The haplotype itself is 70 times (!) shorter than the interval used to estimate the recombination rate, making this estimate only a very crude regional average. Using this estimate, the most likely age of the haplotype is calculated as being 50,000 years, approximately the time introgression may have occurred and far too young to be the product of ILS. However, a more recent, high-resolution map based on large numbers of Icelandic families sequenced to high coverage (Halldorsson et al., 2019b), gives an estimated recombination rate that is 10 times lower, pushing the likely age back to around 500,000 years, far too old to be consistent with introgression but eminently compatible with ILS. Indeed, the high-resolution map reveals that the supposed introgressed fragment actually lies in a recombination cold-spot, with hot-spots on either side, emphasising the importance both of having accurate estimates of recombination rate and of having a resolution well below the size of fragment being studied. This study seems to be a classic marriage between trendy topics, COVID and archaic introgression, that has been snapped up by a top journal more for its headline grabbing title than for its biological plausibility or scientific rigour.

*b) A Denisovan haplotype linked to altitude adaptation in Tibetans*

This is regarded by many as the flagship example of an introgressed haplotype being present in humans, functional and robustly identified. The story features a 37Kb fragment within the gene EPAS1 shared between the Denisovan (a second non-human Hominin) and many Tibetans, yet rare in other human populations and absent from Africa (Huerta-Sanchez et al., 2014). This haplotype is associated with adaptation to year-round living at high altitude, giving a narrative where introgression helped to ‘preadapt’ the Tibetan lineage. However, just as with introgression in general, there is a logical problem. High altitude year-round living only became possible around 5,000 years ago with the development of technologies such as agriculture that allow survival during the harshest months (Lu et al., 2016). Consequently, neither Denisovans nor the ancestors of Tibetans would have benefited from this haplotype. Moreover, Denisovan legacies outside Oceania are thought to be small or negligible (Jacobs et al., 2019; Meyer et al., 2012), raising questions about why a non-beneficial haplotype was retained for tens of thousands of years before the move to high altitude.

Putting these issues aside, as in other examples thought to be due to introgression, the key question is whether this haplotype is introgressed rather than being the product of incomplete lineage sorting. The authors reject ILS largely on the basis of highly significant introgressions statistics, D and S\* (Huerta-Sanchez et al., 2014). However, the significance of D is calculated on the assumption that all informative sites are completely independent, an assumption that could not be more strongly violated in the case of a single haplotype! Similarly, neither S\* nor D are designed to be used on regions experiencing strong selection, as the EPAS1 gene clearly is. Moreover, while it is highly questionable whether introgression statistics provide the strength of evidence suggested in the paper, such discussion is largely a moot point because it is quite easy to test for ILS directly.

Consider a polymorphism that is found both within humans and in archaic Hominins. Such a scenario can arise in one of only three different ways. First, the polymorphism might predate the origin of humans and be found in both groups through incomplete lineage sorting. Second, introgression may introduce into humans an allele that they do not otherwise carry. Third, the two polymorphisms may be the products of two different mutations, each creating the same allele. Of these, double or recurrent mutations are assumed to be extremely rare, though they appear not to be as rare as a naive calculation based on the average, genome-wide mutation rate would indicate (see below). Nonetheless, if there are numerous informative sites in a short region, it seems reasonable to assume that only ILS and introgression offer plausible explanations. However, archaic introgression into Africa is thought to be minimal (Green et al., 2010; Sankararaman et al., 2016) (though see (Chen et al., 2020)), so if the human version of the polymorphism occurs in Africans, introgression is unlikely to be the mechanism. In practice, almost all the variants that define the supposedly Tibetan haplotype are also found in Africans, leaving ILS as the only plausible mechanism.

Interestingly, although the Neanderthal is largely ignored in the original paper and is not included as a control species in several critical analyses, when we added the Neanderthal to the mix an interesting pattern emerged. We find that Neanderthal sequences exhibit about the same number of differences from their most similar human equivalents as the Denisovan does to Tibetans. This is very much the pattern expected under ILS but is not what we would expect in the introgression model, where the Denisovan and Tibetan sequences should be much more similar to each other than the Neanderthal sequences are to their most similar humans.

***A mechanism to explain the patterns we observe?***

A long time ago I notice some trends in data from whales that seemed explicable only if mutation rate was higher in larger populations. But how can a piece of DNA ‘know’ the size of the population it is in? As far as I can see, there is only one possible mechanism, by detecting heterozygosity. All else being equal, heterozygosity increases with population size. Consequently, if mutations are more likely to occur at or near existing heterozygous sites, mutation rate will increase if a population expands and decrease if it contracts. I have since published a number of papers presenting evidence for this effect (Amos, 2010a, 2010b, 2013, 2016, 2019) but population geneticists in particular treat the idea as heretical. Amusingly, I once had a paper categorically rejected by two different Referees. The first was likely a population geneticist who angrily said this is so obviously wrong it must not even be considered. The second was presumably a biochemist who was aware of the extensive work in yeast, and they said ‘this is so obviously correct it cannot be considered novel’! In fact, the yeast work (reviewed in (Amos, 2010b)) shows clearly a process in which heterozygous sites attract ‘repair’ during meiosis, invoking an extra round of DNA replication that will afford extra opportunities for mutations over and above the background rate.

I have spent two decades working on and off on the idea that heterozygosity is mutagenic. Humans offer an excellent test-ground because they lost 25-30% of their heterozygosity during the out of Africa event, predicting a parallel drop in mutation rate. I was thinking about this when I attended a talk by David Reich, who explained that Neanderthals were closer to non-Africans due to introgression. After the talk I emailed to say that non-Africans would also be closer to Neanderthals (and other extinct Hominins) if they had suffered a reduction in mutation rate due to the out of Africa bottleneck. He batted this idea away, saying that the introgression analysis controlled for variation in mutation rate. I only found out later, through careful reading of pages and pages of Supplementary material, that there is no such control and that there is instead simply an explicit assumption that mutation rate does not vary!

There are many attractive features of the idea that mutations are directed towards existing polymorphisms. Most obviously, genes vary hugely in the extent to which a random mutation will be beneficial rather deleterious: many genes are more or less perfectly adapted and almost any change will degrade function while the immune genes are in constant need for novelty to fight an ever-changing array of threats. The heterozygosity effect would act to reduce disturbance to conserved, well-adapted genes and increase the mutation rate at immune genes where novelty is beneficial. Beautiful!

***Introgression statistics are remarkably predictable***

If a tendency for mutation rate and population size to be correlated represents the key driving force behind the unexpected similarity between humans and Neanderthals, we would expect to find a much more general pattern. Specifically, we would expect to find a correlation between heterozygosity difference between human populations and the introgression statistic D. This pattern should be found across the world, regardless of whether introgression was likely or even possible. I tested this prediction using data from the high coverage HGDP genomes, which cover 54 global populations. I find striking correlations both between populations from major geographic regions but also between populations *within* each region, including within Africa. These correlations are always in the direction of the population with lower heterozygosity having (according to general consensus) experienced Neanderthal introgression, exactly the opposite of what would be expected if introgression had actually occurred. If introgression was the driver, the introgressed material would tend to *increase* heterozygosity, creating the exact opposite trend. The generality of these correlations and the way they are similar in all regions of the world, including inside Africa, offers further powerful evidence that, in humans, D is not driven by introgression but is instead driven primarily by something that correlates with heterozygosity. In my opinion, this has to be mutation rate.

***The problem of Supplementary Materials***

Many of the key findings have been published in an astonishing series of papers in Nature and Science, where the condensed format of the main paper forces the narrative into a series of dogmatic assertions supported by references to supplementary materials (SM). These SMs are often vast, running into well over 100 pages of highly technical text, far too much for proper scrutiny during the review process. This acts as a form of double whammy. Not only is there too much to review thoroughly by the journal but the sheer volume also makes it difficult for the authors themselves to check. As a result, the SMs are littered with errors ranging from the trivial, such as mislabelled figures and typos, right up to fundamental issues that undermine the validity of the entire paper. To illustrate the problem, I will give one example based on the crucial data table from the original block-buster paper (Green et al., 2010).



Above is Table S51 which can be found on page 139 of the SM (!). I wonder how many people have ever looked at this, let alone given it critical scrutiny? This table gives counts of all possible base combinations found in an alignment of a non-African, an African, the Neanderthal and the chimpanzee in that order. The vast majority of bases are the same in all individuals (AAAA). We have already met the asymmetric BABA (103,612) and ABBA (95,347) counts from which introgression was inferred. However, there are several other interesting features.

First, look at the numbers of BAAA (756,324) and ABAA (689,594). These represent respectively instances where the African and non-Africa bases have mutated and have nothing to do with introgression because the Neanderthal and chimpanzee are both A. Remarkably, this second pair show almost exactly the same asymmetry as do ABBA and BABA: D(ABBA/BABA) = 0.042, D(BAAA/ABAA) = 0.046. The authors dismiss this problem as likely due to sequencing errors, despite the fact that it involves seven times as many sites as ABBA and BABA! If these data were presented like this in the main paper, I struggle to believe that the introgression story would have ever been published! Of course, since D(BAAA/ABAA) quantifies the excess chance that Africans rather than non-Africans carry a derived variant this it effectively shows directly that the African mutation rate is higher and, as a pattern, explains very neatly why ABBA and BABA show the asymmetry used by many to infer introgression.

Second, consider the state BBBA. Here, a single mutation separates the chimpanzee from Hominins. As such this reflects that length of the evolutionary branch stretching back some 6,000,000 years from the chimpanzee to our common ancestor and then forward again back up to us, a total of around 12,000,000 years. We can compare this number with the number of AABA sites. AABA sites reflect single mutations on the branch leading to Neanderthals, a length that should be around 500,000 years. So, on a nice, well-behaved tree we would expect to find approximately 24X as many BBBA sites as AABA sites. The actual numbers are 8,156,936 and 5,827,247, a ratio of under 1.5! I have no idea what has happened here, though the most direct explanation is that the Neanderthal sequences have a vastly higher sequencing error rate than the other species.

Finally, apart from a constant mutation rate, a second important assumption that underpins many analyses, particularly some that infer introgressed haplotypes, is that recurrent or back-mutations are vanishingly rare. This would be a valid assumption if all bases mutate at the same genome-wide average rate of around 10-8 per site per generation. However, the genome comprises a patchwork of mutation hot-spots and cold-spots, and since the probability of a recurrent mutation scales with the square of the mutation rate, recurrent mutations are disproportionately likely to occur in mutation hot-spots. Moreover, the Table S51 shows just how common they are. All sites with three alleles (i.e. base combinations that contain a C in the Table S51) require two mutations to explain them and these total over 50,000, far too many to ignore as negligible. Moreover, this is a sizeable under-estimate of the true number of recurrent mutations because not all mutations are equal: transition mutations (C<->T and A<->G) occur at twice the rate of transversion mutations (all other changes). Consequently, for every site where three different alleles are found (by force, one transition and one transversion) there will be two other ‘silent’ sites where only two alleles are seen because both mutations were transitions. Note, the three allele sites add further weight to idea that the Neanderthal sequence contains lots of errors because by far the largest number, 32,607 for BBCA, is where the Neanderthal differs from both humans and the chimpanzee.

***The problem of false positives***

In biology, a trend is usually considered ‘significant’ if the odds that it arose by chance are less than one in twenty (P < 0.05). This means that we have to be very careful wherever we conduct lots of different, independent tests. If 20 scientists all test the same phenomenon, on average one will find a ‘significant’ result by the 1 in 20 criterion. In medical genetics, this problem has been much debated because it is common to search the entire genome to discover which genes are linked to a particular disease. With the power of modern techniques, as many as a million or more independent tests may be conducted, meaning that, experiment-wide, true statistical significance is only achieved when the odds of the result occurring by chance drop below around 1 in 20,000,000! Papers that fail to control for this effect are laughed out of town and would never be accepted for publication.

For studies of introgression, a similar problem exists. We already know that ILS creates large numbers of genomic regions that are very difficult to tell apart from introgressed fragments. Any rigorous search for introgressed fragments must therefore tread the same path as medical association studies and adjust the statistical bar for significance to reflect the number of different regions that were tested. Unfortunately, this does not seem to be what is generally done. I have yet to find any formal statement in a paper reporting introgression about any correction for the number of tests conducted, even though it seems obvious that the result being presented is just one result selected from many other, unreported tests.

To illustrate, let us revisit the EPAS1 story. The results are presented as if only a single hypothesis was tested. However, the work reflects a whole series of choices where multiple options were available. Thus: (1) the Denisovan is tested and not the Neanderthal, even though the ancestors of Tibetans are believed to have had a far greater Neanderthal legacy; (2) it is never explained why only a single selected trait was analysed out of many possible options for traits under selection in humans; (3) even if only a single trait was tested, why choose adaptation to year-round living at high altitude, a trait that had no relevance either to Denisovans or to Tibetan ancestors; (4) why choose EPAS1 for testing rather than any of the other genes linked to altitude adaptation (Huerta-Sanchez et al., 2013; Jorgensen et al., 2023; Yi et al., 2010); (5) within EPAS1, why choose just one, intronic region rather than, for example, coding sequences; (6) why select Tibetans rather than other population groups. Only the authors know how many tests they conducted and whether their failure to report and correct for other, non-significant tests was an oversight. However, I think it is interesting to turn the problem around and ask what did their funding application looked like? Was it really to conduct one single, highly specific test, end of story, or was it to conduct a systematic search for candidate regions? If the latter is the case, where is the correction for multiple testing? Personally, I find it inconceivable that here, as elsewhere, wider searches were not made and that the result that got published is simply the most convincing case found out of many that were set aside.

**Conclusion**

It is undeniably romantic to think that the enigmatic Neanderthals live on as a genetic legacy within us, perhaps as long as this legacy does us no harm! I hope that reading this piece you will see that there are considerable grounds for scepticism. The dominant issue is that the almost universal assumption of a constant mutation rate means that any rejection of the null hypothesis is then taken as evidence for introgression. At the same time, almost no attention has been given to features such as the allele frequency distributions to see how well these fit the introgression narrative. All this is made much worse by what I feel is likely to be a widespread failure to appreciate the statistical problem of searching for patterns and then publishing only those found that are significant, without adjusting for the number of tests conducted. Conversely, if BOTH alternate hypotheses are treated equally, everything I have done points resoundingly towards a lower mutation rate in non-Africans as being the primer driver behind their greater similarity to Neanderthals. More generally, instead of Hominins apparently mating successfully with almost anything bipedal they encountered, a mutation rate that varies as populations expand and contract provides a simple explanation for why evidence of introgression is found almost wherever it is looked for.

**Glossary**

**Africans and non-Africans.** Unless otherwise stated, these terms refer to modern-day humans as sampled for large-scale human sequencing projects such as the 1000 genomes study or the Human Genetic Diversity Panel.

**Archaic.** When used as a noun, this is shorthand for archaic Hominin, usually meaning Neanderthal but able to include related species such as the Denisovan and potentially other, as yet undescribed species.

**Denisovan.** One of two archaic Hominins for which a high-quality genome has been obtained, the other being the Neanderthal. Very little is known about the Denisovans but they appear a little more related to Neanderthals than either is to humans.

**Haplotype.** A DNA sequence on a single chromosome that has not been broken down by recombination or degraded by mutation such that the same sequence can be found in multiple copies sampled from the same or different populations.

**Heterozygosity.** Humans carry two copies of almost every gene, one copy inherited from each parent. When these copies are identical, we say the individual is homozygous and when they differ, we say the individual is heterozygous. These terms and general and can refer to any segment of DNA from a single nucleotide base up to an entire gene or more.

**The signal.** I use this as shorthand for the evidence used to infer archaic introgression. In most cases, this refers to the introgression statistic, D, where, *if mutation rate is constant*, positive values tend to indicate introgression into non-Africans.

**Links to four key open-access papers I have published that relate directly to the question of Neanderthal introgression and how this may be explained.**

Correlated and geographically predictable Neanderthal and Denisovan legacies are difficult to reconcile with a simple model based on inter-breeding.

[**https://doi.org/10.1098/rsos.201229**](https://doi.org/10.1098/rsos.201229)

Signals interpreted as archaic introgression appear to be driven primarily by faster evolution in Africa.

[**https://doi.org/10.1098/rsos.191900**](https://doi.org/10.1098/rsos.191900)

Flanking heterozygosity influences the relative probability of different base substitutions in humans.

[**https://doi.org/10.1098/rsos.191018**](https://doi.org/10.1098/rsos.191018)

Variation in Heterozygosity Predicts Variation in Human Substitution Rates between Populations, Individuals and Genomic Regions.

[**https://doi.org/10.1371/journal.pone.0063048**](https://doi.org/10.1371/journal.pone.0063048)

**Bibliography**

*This is not intended to be anywhere near exhaustive but does include key papers that play a major role in the narrative.*

Amos, W. (2010a). Even small SNP clusters are non-randomly distributed: is this evidence of mutational non-independence? *Proc. R. Soc. B*, *277*(1686), 1443–1449.

Amos, W. (2010b). Heterozygosity and mutation rate: evidence for an interaction and its implications. *BioEssays*, *32*, 82–90.

Amos, W. (2011). Population-specific links between heterozygosity and the rate of human microsatellite evolution. *J. Mol. Evol.*, *72*(2), 215–221.

Amos, W. (2013). Variation in heterozygosity predicts variation in human substitution rates between populations, individuals and genomic regions. *PLoS ONE*, *8*(4), e63048.

Amos, W. (2016). Heterozygosity increases microsatellite mutation rate. *Biology Letters*, *12*(1). https://doi.org/10.1098/rsbl.2015.0929

Amos, W. (2019). Flanking heterozygosity influences the relative probability of different base substitutions in humans. *R. Soc. Open Sci.*, *6*, 191018. https://doi.org/dx.doi.org/10.1098/rsos.191018

Amos, W. (2020). Signals interpreted as archaic introgression appear to be driven primarily by accelerated evolution in Africa. *Royal Society Open Science*, *7*, 191900. https://doi.org/doi.org/10.1098/rsos.191900

Amos, W. (2021). Correlated and geographically predictable Neanderthal and Denisovan legacies are difficult to reconcile with a simple model based on inter-breeding. *Royal Society Open Science*, *8*(6). https://doi.org/10.1098/rsos.201229

Banks, W. E., d’Errico, F., Peterson, A. T., Kageyama, M., Sima, A., & Sánchez-Goñi, M.-F. (2008). Neanderthal extinction by competitive exclusion. *PLoS ONE*, *3*(12), e3972.

Chen, L., Wolf, A. B., Fu, W., Li, L., & Akey, J. M. (2020). Identifying and Interpreting Apparent Neanderthal Ancestry in African Individuals. *Cell*, *180*, 1–11. https://doi.org/https://doi.org/10.1016/j.cell.2020.01.012

Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for admixture between closely related populations. *Mol. Biol. Evol.*, *28*(8), 2239–2252.

Fu, Q., Hajdinjak, M., Moldovan, O. T., Constantin, S., Mallick, S., Skoglund, P., Patterson, N., Rohland, N., Lazaridis, I., Nickel, B., Viola, B., Prufer, K., Meyer, M., Kelso, J., Reich, D., & Paabo, S. (2015). An early modern human from Romania with a recent Neanderthal ancestor. *Nature*, *524*(7564), 216-+. https://doi.org/10.1038/nature14558

Green, R. E., Krause, J., NBriggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H.-Y., Hansen, N. F., Durand, E. Y., Malaspinas, A.-S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. A., … Pääbo, S. (2010). A draft sequence of the Neandertal genome. *Science*, *328*, 710.

Halldorsson, B. V, Palsson, G., Stefansson, O. A., Jonsson, H., Hardarson, M. T., Eggertsson, H. P., Gunnarsson, B., Oddsson, A., Halldorsson, G. H., Zink, F., Gudjonsson, S. A., Frigge, M. L., Thorleifsson, G., Sigurdsson, A., Stacey, S. N., Sulem, P., Masson, G., Helgason, A., Gudbjartsson, D. F., … Stefansson, K. (2019a). Characterizing mutagenic effects of recombination through a sequence-level genetic map. *Science*, *363*(6425), 364-+. https://doi.org/10.1126/science.aau1043

Halldorsson, B. V, Palsson, G., Stefansson, O. A., Jonsson, H., Hardarson, M. T., Eggertsson, H. P., Gunnarsson, B., Oddsson, A., Halldorsson, G. H., Zink, F., Gudjonsson, S. A., Frigge, M. L., Thorleifsson, G., Sigurdsson, A., Stacey, S. N., Sulem, P., Masson, G., Helgason, A., Gudbjartsson, D. F., … Stefansson, K. (2019b). Characterizing mutagenic effects of recombination through a sequence-level genetic map (vol 363, eaau1043, 2019). *Science*, *363*(6427). https://doi.org/10.1126/science.aaw8705

Hinch, A. G., Tandon, A., Patterson, N., Song, Y. L., Rohland, N., Palmer, C. D., Chen, G. K., Wang, K., Buxbaum, S. G., Akylbekova, E. L., Aldrich, M. C., Ambrosone, C. B., Amos, C., Bandera, E. V, Berndt, S. I., Bernstein, L., Blot, W. J., Bock, C. H., Boerwinkle, E., … Myers, S. R. (2011). The landscape of recombination in African Americans. *NATURE*, *476*(7359), 170-U67. https://doi.org/10.1038/nature10336

Huerta-Sanchez, E., DeGiorgio, M., Pagani, L., Tarekegn, A., Ekong, R., Antao, T., Cardona, A., Montgomery, H. E., Cavalleri, G. L., Robbins, P. A., Weale, M. E., Bradman, N., Bekele, E., Kivisild, T., Tyler-Smith, C., & Nielsen, R. (2013). Genetic Signatures Reveal High-Altitude Adaptation in a Set of Ethiopian Populations. *MOLECULAR BIOLOGY AND EVOLUTION*, *30*(8), 1877–1888. https://doi.org/10.1093/molbev/mst089

Huerta-Sanchez, E., Jin, X., Asan, Bianba, Z., Peter, B. M., Vinckenbosch, N., Liang, Y., Yi, X., He, M. Z., Somel, M., Ni, P. X., Wang, B., Ou, X. H., Huasang, Luosang, J. B., Cuo, Z. X. P., Li, K., Gao, G. Y., Yin, Y., … Nielsen, R. (2014). Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature*, *512*(7513), 194-+. https://doi.org/10.1038/nature13408

Jacobs, G. S., Hudjashov, G., Saag, L., Kusuma, P., Darusallam, C. C., Lawson, D. J., Mondal, M., Pagani, L., Ricaut, F. X., Stoneking, M., Metspalu, M., Sudoyo, H., Lansing, J. S., & Cox, M. P. (2019). Multiple Deeply Divergent Denisovan Ancestries in Papuans. *Cell*, *177*(4), 1010-+. https://doi.org/10.1016/j.cell.2019.02.035

Jorgensen, K., Song, D. S., Weinstein, J., Garcia, O. A., Pearson, L. N., Inclan, M., Rivera-Chira, M., Leon-Velarde, F., Kiyamu, M., Brutsaert, T. D., Bigham, A. W., & Lee, F. S. (2023). High-Altitude Andean H194R HIF2A Allele Is a Hypomorphic Allele. *MOLECULAR BIOLOGY AND EVOLUTION*, *40*(7). https://doi.org/10.1093/molbev/msad162

Kong, A., Gudbjartsson, D. F., Sainz, J., Jonsdottir, G. M., Gudjonsson, S. A., Richardsson, B., Sigurdottir, S., Barnard, J., Hallbeck, B., Masson, G., Shlien, A., Palsson, S. T., Frigge, M. L., Thorgeirsson, T. E., Gulcher, J. R., & Stefansson, K. (2002). A high-resoluton recombination map of the human genome. *Nat. Genet.*, *31*, 241–247.

Lu, D. S., Lou, H. Y., Yuan, K., Wang, X. J., Wang, Y. C., Zhang, C., Lu, Y., Yang, X., Deng, L. A., Zhou, Y., Feng, Q. D., Hu, Y., Ding, Q. L., Yang, Y. J., Li, S. L., Jin, L., Guan, Y. Q., Su, B., Kang, L. L., & Xu, S. H. (2016). Ancestral Origins and Genetic History of Tibetan Highlanders. *American Journal of Human Genetics*, *99*(3), 580–594. https://doi.org/10.1016/j.ajhg.2016.07.002

McVean, G., Myers, S. R., Hunt, S., Deloukas, P., Bentley, D. R., & Donnelly, P. (2004). The fine-scale structure of recombination rate variation in the human genome. *Science*, *304*(581–584).

Meyer, M., Kircher, M., Gansauge, M. T., Li, H., Racimo, F., Mallick, S., Schraiber, J. G., Jay, F., Prufer, K., de Filippo, C., Sudmant, P. H., Alkan, C., Fu, Q. M., Do, R., Rohland, N., Tandon, A., Siebauer, M., Green, R. E., Bryc, K., … Paabo, S. (2012). A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science*, *338*(6104), 222–226. https://doi.org/10.1126/science.1224344

Prugnolle, F., Manica, A., & Balloux, F. (2005). Geography predicts neutral genetic diversity of human populations. *Curr. Biol.*, *15*(5), R159–R160.

Ptak, S., Hinds, D. A., Koehler, K., Nickel, B., Patil, N., Ballinger, D. G., Przeworski, M., Frazer, K. A., & Pääbo, S. (2005). Fine-scale recombination patterns differ between chimpanzees and humans. *Nat. Genet.*, *37*, 429–434.

Rinker, D. C., Simonti, C. N., McArthur, E., Shaw, D., Hodges, E., & Capra, J. A. (2020). Neanderthal introgression reintroduced functional ancestral alleles lost in Eurasian populations. *Nature Ecology & Evolution*, *4*(10), 1332-+. https://doi.org/10.1038/s41559-020-1261-z

Sankararaman, S., Mallick, S., Dannemann, M., Prufer, K., Kelso, J., Paabo, S., Patterson, N., & Reich, D. (2014). The genomic landscape of Neanderthal ancestry in present-day humans. *Nature*, *507*(7492), 354-+. https://doi.org/10.1038/nature12961

Sankararaman, S., Mallick, S., Patterson, N., & Reich, D. (2016). The Combined Landscape of Denisovan and Neanderthal Ancestry in Present-Day Humans. *Current Biology*, *26*(9), 1241–1247. https://doi.org/10.1016/j.cub.2016.03.037

Serangeli, J., & Bolus, M. (2008). Out of Europe - The dispersal of a successful European hominin form. *Quartär*, *55*, 83–98.

Skov, L., Macia, M. C., Sveinbjornsson, G., Mafessoni, F., Lucotte, E. A., Einarsdottir, M. S., Jonsson, H., Halldorsson, B., Gudbjartsson, D. F., Helgason, A., Schierup, M. H., & Stefansson, K. (2020). The nature of Neanderthal introgression revealed by 27,566 Icelandic genomes. *Nature*, *582*(7810), 78-+. https://doi.org/10.1038/s41586-020-2225-9

Slon, V., Mafessoni, F., Vernot, B., de Filippo, C., Grote, S., Viola, B., Hajdinjak, M., Peyregne, S., Nagel, S., Brown, S., Douka, K., Higham, T., Kozlikin, M. B., Shunkov, M. V, Derevianko, A. P., Kelso, J., Meyer, M., Prufer, K., & Paabo, S. (2018). The genome of the offspring of a Neanderthal mother and a Denisovan father. *Nature*, *561*(7721), 113-+. https://doi.org/10.1038/s41586-018-0455-x

Wall, J. D., Yang, M. A., Jay, F., Kim, S. K., Durand, E. Y., Stevison, L. S., Gignoux, C., Woerner, A., Hammer, M. F., & Slatkin, M. (2013). Higher Levels of Neanderthal Ancestry in East Asians than in Europeans. *Genetics*, *194*(1), 199-+. https://doi.org/10.1534/genetics.112.148213

Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z. X. P., Pool, J. E., Xu, X., Jiang, H., Vinckenbosch, N., Korneliussen, T. S., Zheng, H. C., Liu, T., He, W. M., Li, K., Luo, R. B., Nie, X. F., Wu, H. L., Zhao, M. R., Cao, H. Z., … Wang, J. A. (2010). Sequencing of 50 Human Exomes Reveals Adaptation to High Altitude. *SCIENCE*, *329*(5987), 75–78. https://doi.org/10.1126/science.1190371

Zeberg, H., & Paabo, S. (2020). The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. *Nature*, *587*(7835), 610-+. https://doi.org/10.1038/s41586-020-2818-3

Zeberg, H., & Paabo, S. (2021). A genomic region associated with protection against severe COVID-19 is inherited from Neandertals. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(9). https://doi.org/10.1073/pnas.2026309118