Evidence for archaic adaptive introgression in humans

Fernando Racimo[1], Sriram Sankararaman[2], Rasmus Nielsen[1,3] and Emilia Huerta-Sánchez[4]

Abstract | As modern and ancient DNA sequence data from diverse human populations accumulate, evidence is increasing in support of the existence of beneficial variants acquired from archaic humans that may have accelerated adaptation and improved survival in new environments — a process known as adaptive introgression. Within the past few years, a series of studies have identified genomic regions that show strong evidence for archaic adaptive introgression. Here, we provide an overview of the statistical methods developed to identify archaic introgressed fragments in the genome sequences of modern humans and to determine whether positive selection has acted on these fragments. We review recently reported examples of adaptive introgression, grouped by selection pressure, and consider the level of supporting evidence for each. Finally, we discuss challenges and recommendations for inferring selection on introgressed regions.

Modern humans
Present-day humans and their recent ancestors, up to the time at which they diverged from their most closely related archaic human groups, the Neanderthals and Denisovans.

[1]*Department of Integrative Biology, University of California, Berkeley, Berkeley, California 97420, USA.*
[2]*Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA.*
[3]*Department of Statistics, University of California, Berkeley, Berkeley, California 97420, USA.*
[4]*Molecular and Cell Biology Unit, School of Natural Sciences, University of California, Merced, Merced, California 95343, USA.*
Correspondence to E.H.S.
e-mail: ehuerta-sanchez@ucmerced.edu
doi:10.1038/nrg3936
Published online 12 May 2015

The relationship between modern humans and other, now extinct, archaic hominin groups has been a subject of controversy since the 1970s. Two competing hypotheses were originally proposed: the multiregional model[1] posited that modern humans evolved in parallel throughout Africa and Eurasia from different archaic groups while exchanging migrants, whereas the out-of-Africa model proposed that all present-day humans had a recent origin in the African continent, from which they expanded across the world[2]. However, over the past 30 years, these two hypotheses were increasingly seen as an over-simplification. Other intermediate models that emerged involved a recent origin in Africa with limited amounts of admixture from Eurasian archaic groups[3], or considerable assimilation of Neanderthals into the modern human gene pool during modern human expansion into Europe[4].

Until recently, analyses of whole-genome sequences from modern human populations seemed to support the out-of-Africa model, although certain studies observed genomic patterns that might suggest local gene flow between modern and archaic populations[5,6]. Archaeological evidence also suggests Neanderthal presence up to 40,000 years ago in Europe and western Asia[7], which means that they could have coexisted with modern human populations for a period of at least 2,600 years. However, in the past five years, whole-genome sequences from two archaic human groups, Neanderthals[8–10] and Denisovans[11,12], have provided direct insights into the extent of gene flow between archaic humans and modern humans.

Although it is now widely accepted that admixture occurred between archaic and modern human groups, little is known about the adaptive contribution of the introgressed segments. Cases of introgression enabling adaptation in animals and plants have been extensively documented (reviewed in REFS 13–15), although surprisingly little attention has been devoted to adaptive introgression in humans[16,17].

Here, we review recent human genetic studies that have identified several examples of archaic adaptive introgression of modern humans with Neanderthals or Denisovans. First, we provide a brief introduction to the statistical methods that have been used to detect adaptive introgression based on the whole-genome sequences of modern and archaic humans. We then review the evidence in support of particular proposed instances of archaic adaptive introgression. Finally, we discuss several unanswered questions in the field and propose possible avenues of future research, such as the development of methods to jointly model selection and introgression.

Archaic gene flow

Although the majority of non-African human ancestry is shared with Africans, non-Africans also possess a small amount of DNA (1.5–2.1%) from Neanderthals[8].

© 2015 Macmillan Publishers Limited. All rights reserved

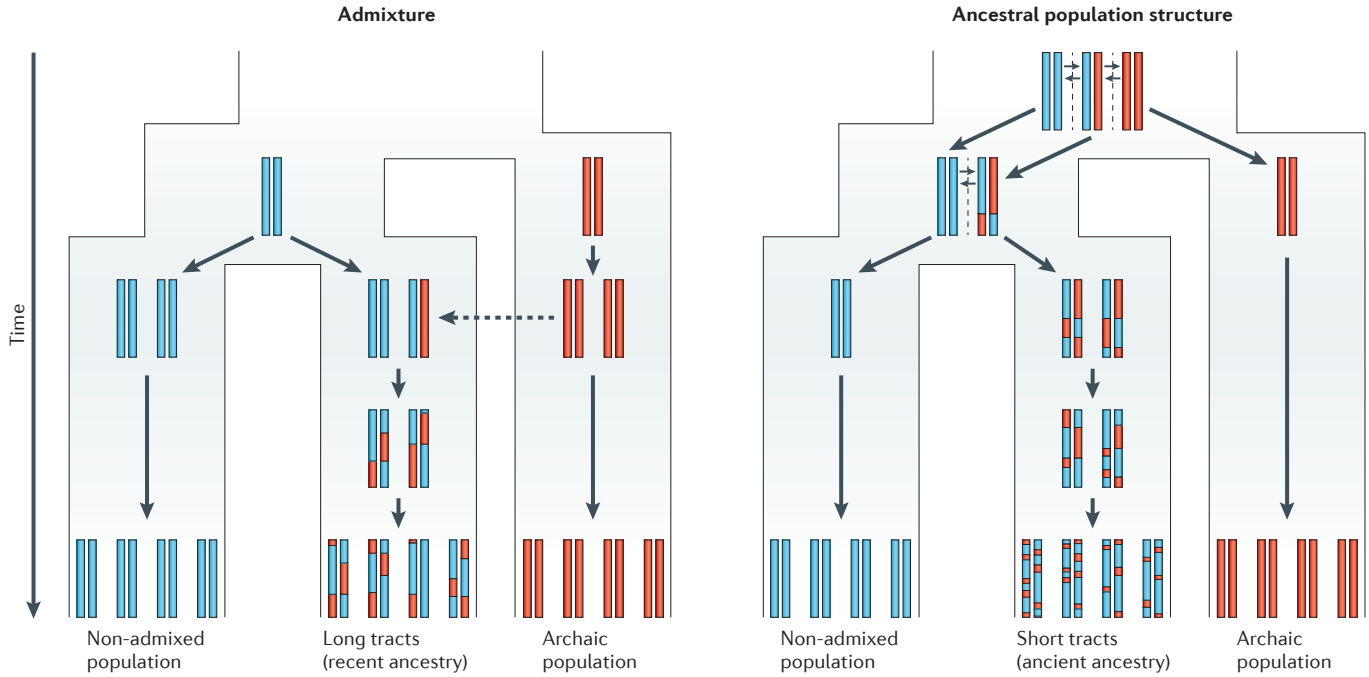


Figure 1 | **Expected length of archaic tracts under admixture and ancestral population structure scenarios.** Expected lengths of archaic-like tracts for a scenario of admixture (left panel) and a scenario of ancestral population structure (right panel) are shown. Because there is less time for recombination to break down the migrant tracts (red) in the admixed population, the expected tract length in this case will be longer than in the case of ancestral population structure.

Out-of-Africa model
A model of recent human evolution positing that all present-day humans had a recent origin in Africa and then expanded across the world, replacing other archaic groups.

Admixture
Genetic exchange between individuals from two populations that were isolated in the past.

Archaic humans
A broad category of human populations that diverged from present-day humans 550–765 thousand years ago (kya) (assuming a mutation rate of 0.5×10^{-9} per base pair per year) before present-day human populations started diverging from each other 86–130 kya (assuming the same mutation rate) and that are now extinct. This includes the Neanderthal and Denisovan populations.

Ancestral population structure
A demographic scenario in which an ancestral population is not homogenously mixing. For example, some subpopulations might exchange more migrants with certain other subpopulations than with the rest because of geography or mate choice.

The level of Neanderthal admixture varies in non-Africans; for example, it was recently found that the whole-genome sequences of Asian individuals show a larger proportion of Neanderthal ancestry segments than sequences of Europeans[12,18,19]. In addition, a small portion of Melanesian, Papuan and Australian ancestry (3–6%) is derived from Denisovans[11,12,20], and lower amounts of Denisovan ancestry (0.2%) are also found in East Asia[9,21]. In addition, recent analyses of whole-genome sequences of individuals from different populations in Africa suggest that some African populations may have also exchanged genetic material with as yet undetermined archaic human groups[5,22,23]. Recent demographic inferences from analyses of modern and archaic human genome sequences are reviewed in REF. 24.

The presence of Neanderthal and Denisovan DNA in the whole-genome sequences of present-day non-African individuals is generally accepted to be a consequence of admixture, probably as a result of limited interbreeding between modern and archaic humans[25]. An alternative explanation is ancestral population structure. Under this model, population subdivision in the ancestral population of archaic and modern humans may have resulted in some modern human groups, such as the ancestors of Eurasians, being more closely related to Neanderthals than other groups that stayed in Africa[26] (FIG. 1). However, the date of the last Neanderthal gene exchange into present-day Eurasians supports the admixture scenario. This event can be dated with some accuracy by examining the distribution of the length of tracts of introgressed DNA. Each round of recombination breaks down haplotypes into shorter fragments. Therefore, DNA from recent introgression events should fall into longer contiguous tracts than DNA resulting from old introgression events[27–29] or from ancestral population structure (FIG. 1). Measurements of introgression tract lengths indicate that the last gene flow event occurred 50–60 thousand years ago (kya)[30–32]. This timeframe is too recent to support a scenario of ancestral population structure, as the divergence time between humans and Neanderthals is thought to have occurred 550–765 kya (assuming a human mutation rate of 0.5×10^{-9} per base pair per year) or 275–383 kya (assuming a mutation rate of 10^{-9} per base pair per year)[9]. Furthermore, it also largely post-dates the African–Eurasian population split, which is estimated to have occurred 100–160 kya (under the slow mutation rate) or 50–80 kya (under the fast rate)[33–35]. The hypothesis of post-split admixture is also supported by additional analyses based on the distribution of single-nucleotide polymorphism (SNP) allele frequencies in Neanderthals and present-day humans and of allelic configurations in short genomic blocks[36,37]. Finally, all of these studies provided evidence for archaic human gene flow into modern humans[8,9,12] but not in the reverse direction. The reason for this may be that all archaic humans for whom genome sequences are available so far are likely to pre-date the time of contact with modern humans.

Methods to infer introgressed segments

The challenge for methods seeking to identify introgression is to distinguish true introgression from shared ancestral genetic variation (FIG. 1). Any two populations

© 2015 Macmillan Publishers Limited. All rights reserved

will always share some segments of DNA that originated in their common ancestor because both populations descend from the same population and might therefore have inherited some of the same segments of DNA from the ancestral population. For this reason, two DNA segments sampled from different populations may share a most recent common ancestor (MRCA) more recently than two DNA sequences sampled from the same population. The same argument is true for species; in this Review, we do not, in general, distinguish between species and populations, in part to avoid entering into discussions of species concepts and definitions in hominins.

Patterson's D statistic and genome-wide data. There are several statistical methods available for identifying introgression from genome-wide data[8,38–40]. The most well known is Patterson's *D* statistic[8,38,39], which measures the excess sharing of derived alleles between each of two populations in a pair (ingroup populations) and an outgroup population (see Supplementary information S1 (box)). Patterson's *D* statistic takes advantage of a phylogenetic argument: if neither of the ingroup populations have had any gene flow from the outgroup population, then each of the two ingroup populations should share approximately the same number of derived alleles with the outgroup population. The significance of deviations from the symmetrical pattern expected in the absence of introgression is evaluated using a 'block-bootstrap' or 'jack-knife' method applied to genome-wide data. This statistic was used as one of the primary lines of evidence for identifying human–Neanderthal introgression by showing that populations outside Africa share more derived alleles with Neanderthals than do African populations[8].

The identification of specific genes or segments of the genome that are introgressed is even more challenging because the simple re-sampling methods for evaluating the significance of Patterson's *D* statistic are inapplicable to shorter genomic regions[40]. Inferences regarding specific regions must instead rely on demographic models that include assumptions about parameters such as divergence times, effective population sizes and recombination rates.

Phylogenetic information and sequence divergence. A number of statistics commonly used in genetic analyses may capture information regarding introgression. In combination with parametric simulations for evaluating significance, these statistics may be used to distinguish incomplete lineage sorting (ILS) from introgression. Although Patterson's *D* statistic (see above) captures phylogenetic information, an alternative is to use statistics based on sequence divergence. An introgressed haplotype should have low sequence divergence to the putative archaic source population but high sequence divergence to other present-day human individuals. One approach to identify an introgressed segment[41] is to calculate the most likely time of the MRCA (TMRCA) of the test haplotype and the archaic haplotype, as well as the TMRCA of the test haplotype and a second modern human haplotype[41]. A test human haplotype that has a very recent TMRCA with an archaic population, but a very ancient TMRCA with other human haplotypes, is likely to have been introgressed from the archaic population. We note that to formalize this into a test of introgression, it is necessary to make specific assumptions about divergence times and population sizes. Simulations can then be used to determine significance.

Tract length, linkage disequilibrium and the S* statistic. As mentioned above, the expected length of an introgressed tract depends on the time since introgression. In fact, under simple assumptions, the lengths of introgressed tracts should approximately follow an exponential distribution with mean length $[(1-m)r(t-1)]^{-1}$, where t is the number of generations since a proportion m of the population was replaced by migrants from another population, and r is the recombination rate (in units of Morgans) per base pair. Conditions under which this approximation breaks down are detailed in REF. 27. A defining feature of introgression is that it should, on average, generate longer tracts than ILS. Furthermore, as the length of the tracts only depends (to a first approximation) on r, m and t, and not on effective population sizes, using the length of the introgressed haplotype provides a more robust method for distinguishing between introgression and shared ancestral variation (BOX 1). The only caveat is that introgressed haplotypes are not directly observable but must be inferred from the data. Alternatively, statistics that summarize information related to haplotype length without directly inferring haplotypes can be used. In particular, the existence of a long introgressed haplotype should increase long-range linkage disequilibrium (LD). Therefore, examining patterns of LD provides an alternative method for identifying introgression. The *S** statistic (BOX 2) provides a commonly used method for extracting this information, although it also incorporates information regarding divergence[5,6,19,42]. *S** was originally derived to identify genome-wide evidence of introgression similarly to the *D* statistic, but without knowledge of the donor population. In subsequent studies, it was also used locally to identify highly divergent haplotypes harbouring variants in strong LD in order to search for regions introgressed from Neanderthals into non-Africans[19].

Probabilistic models: HMM and CRF. An alternative to using simulations to determine the significance of a summary statistic is to incorporate the parametric assumptions directly into a probabilistic framework. For example, both Prüfer *et al.*[9] and Seguin-Orlando *et al.*[31] developed a hidden Markov model (HMM) to detect fragments introduced by archaic introgression. In both studies, the authors used information from non-admixed Africans and an archaic (Denisovan or Neanderthal) genome sequence, as well as a test phased genome sequence with resolved haplotypes from a non-African population that may contain introgressed fragments. Under the HMM framework, the ancestry of each SNP in the genome is a hidden random variable with two states — archaic (that is, introgressed) or modern human — which are estimated from the genomic data.

Haplotypes
Sequences of contiguous alleles that are closely linked and that tend to be inherited together as a single unit.

Human mutation rate
The rate (per base pair) at which mutations appear in the genome sequence of an individual at each generation or year. Currently, the exact value of this rate in humans is a topic of debate, with most estimates ranging from a value of 0.5×10^{-9} per base pair per year to a value of 10^{-9} per base pair per year.

D statistic
A summary statistic based on differential sharing of derived alleles among different pairs of populations. When applied on a genome-wide scale, they can be used to detect significant deviations from a strict population tree with no admixture or migration.

Incomplete lineage sorting
(ILS). A phenomenon whereby two or more lineages from different populations or species share a common ancestor more recently than their respective most recent common ancestor within populations, causing discordance between the population tree and a gene tree.

Time of the MRCA
(TMRCA). Time in generations back into the past until two copies of an allele or two haplotypes shared a most recent common ancestor (MRCA). This is often an unknown parameter that can be estimated from genetic data.

Linkage disequilibrium
(LD). A nonrandom association of alleles in different loci along the same chromosome due to low recombination rate, population structure and/or selection.

*S** statistic
A summary statistic based on patterns of linkage disequilibrium that can be used to detect introgressed haplotypes.

© 2015 Macmillan Publishers Limited. All rights reserved

Box 1 | Using haplotype length to distinguish introgression from ancestral shared polymorphism

A high level of nucleotide similarity between sequences found in high linkage disequilibrium in a specific population and an outgroup population or species is not necessarily caused by introgression from the outgroup. An alternative explanation is ancestral shared polymorphism, perhaps caused by deep population structure. This model posits that a haplotype existed before the divergence from the outgroup and only survived in a particular population and the outgroup but not in other populations (FIG. 1). A way to distinguish ancestral shared haplotypes from introgressed haplotypes is by measuring their length[41,57,60].

Recombination breaks apart haplotypes over time; thus, long shared haplotypes between any two sets of lineages tend to be recent. In turn, short shared haplotypes tend to be older and are therefore more likely to be caused by ancestral shared polymorphism (FIG. 1). If the lengths of introgressed tracts are exponentially distributed, then the probability of observing a haplotype of length $\geq k$ that is shared between two populations approximately follows an exponential distribution with parameter k/L: $P(\text{block size} \geq k) = \exp(-k/L)$, where L is the expected length of a shared sequence[28] (but see REFS 27,29 for violations of this assumption). The parameter L is equal to $[(1-m)r(t-1)]^{-1}$, which can be approximated as $1/(rt)$, where t is the time separating the two populations (in generations), r is the recombination rate per base pair per generation, and m is the admixture fraction.

The above formula is valid for randomly chosen introgression tracts. However, for randomly chosen sites in the genome, conditionally on the site containing introgressed DNA, the length of the tract will be a sum of two exponential random variables (each one representing the sequence length to the left and the right from the chosen site). This model does not include selection, and it does not depend on the frequency of the haplotype.

For example, let there be a haplotype of length 40,000 bp that is shared between two populations split 8,000 generations in the past. This haplotype is in a region with a recombination rate of 2.5×10^{-8}. We can calculate that the expected length of a shared sequence (L) is approximately 5,000 bp, and the haplotype therefore has a probability of 0.0003 of having persisted as a result of ancestral shared polymorphism.

Hidden Markov model
(HMM). A statistical modelling method used to infer hidden states from observed data along an ordered sequence, in which each hidden variable is independent of all other hidden variables, conditional on knowing the state of the immediately previous hidden variable.

Archaic introgression
The introduction of genetic material into the ancestors of an extant population (for example, East Asians) from an archaic population that is currently extinct (for example, Neanderthals) via admixture.

Conditional random field
(CRF). A statistical modelling method that is similar to a hidden Markov model but that also allows contextual data (regional data not directly contiguous to a site in a sequence) to provide information about the state of a hidden variable.

Emission functions
Functions that relate the hidden variables to the observed data in the conditional random field framework.

Positive selection
Selection that favours a specific allele over others. The allele may consequently rise to high frequency or become fixed. Hitchhiking of neutral alleles tightly linked to the favoured allele leaves a known genetic footprint in the genome, sometimes allowing detection of positive selection at a particular locus.

Balancing selection
Selection that favours the maintenance of variability in a population, which can prevent any single allele from reaching fixation. Examples include frequency-dependent selection and heterozygous advantage (that is, overdominance).

Negative selection
Selection that acts to prune away deleterious variants from the genome.

For each site in the test genome, the most likely state (archaic ancestry or modern human ancestry) giving rise to the observed data is then inferred probabilistically with the posterior decoding of the HMM[43]. This approach allowed the authors of both studies to infer the probability of introgression at different regions of the genomes of non-African individuals. The primary difference between the two methods is that Prüfer *et al.*[9] used a priori chosen parameters, whereas Seguin-Orlando *et al.*[31] estimated parameters from a reference data set.

A similar method, called conditional random field (CRF), models the ancestry of a set of contiguous SNPs along windows of the genome[44]. Under this framework, the SNP ancestry is also included as a random variable with two states: archaic or modern. However, unlike the HMM methods, the CRF model contains one or more emission functions that incorporate information about LD, haplotype structure and allele configurations from nearby SNPs simultaneously from multiple individuals (BOX 3). The parameters of the method are calibrated using parametric simulations with specific demographic assumptions such as divergence times and effective population sizes.

Inferring selection on introgressed DNA

If a haplotype were introduced by introgression from archaic humans, the reason for its continued survival in present-day humans may be that it was affected by positive selection or balancing selection, or that it was not removed from the population by negative selection or genetic drift. Distinguishing between these hypotheses requires additional statistical analyses because when a genomic region contains introgressed DNA, the pattern of polymorphism may be different from that expected under a neutral model with no introgression. As mentioned above, in regions with introgressed DNA, the haplotype structure will change, there will be an increase in LD[45], and the distribution of allele frequencies will change from that expected without introgression[36]. These patterns are exactly the signals used by many standard methods for detecting selection. These include the integrated haplotype score (iHS)[46], extended haplotype homozygosity (EHH)[47], cross-population extended haplotype homozygosity (XP-EHH)[48], Tajima's D[49], Fay and Wu's H[50], composite of multiple signals (CMS)[51], F_{ST} (REFS 52,53) and variations on F_{ST} such as the locus-specific branch length (LSBL)[54] or the population branch statistic (PBS)[55]. Naive use of these methods can therefore easily lead to false inferences of selection, as the pattern generated by introgression alone may be incorrectly interpreted as evidence of selection.

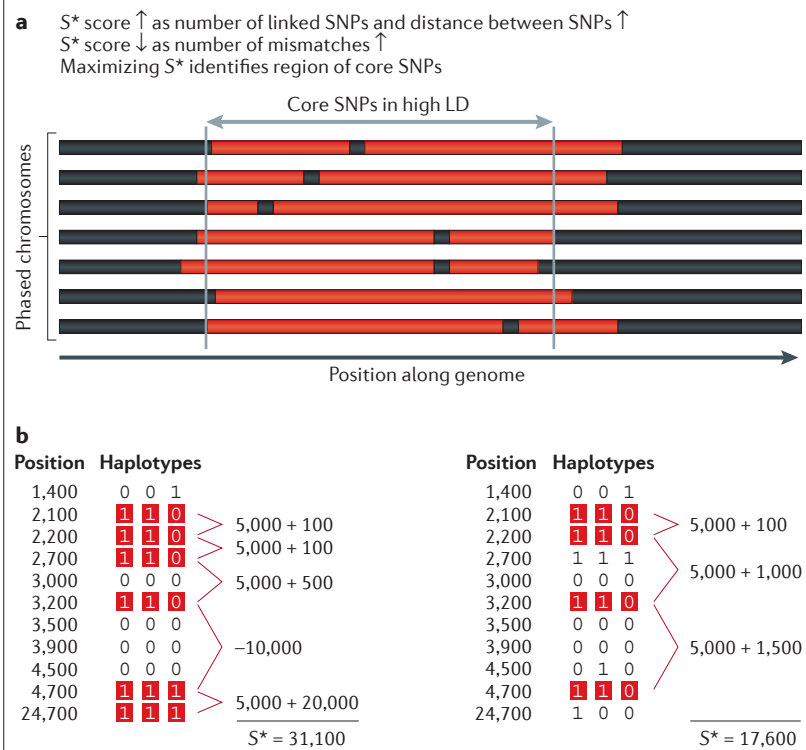
High archaic haplotype frequency. Positive selection produces an increase in selected allele frequencies over time. Therefore, a direct way of inferring selection is the detection of significantly high frequencies of an introgressed fragment in a specific population relative to other populations. This approach essentially obtains the information that F_{ST}-type methods rely on. These approaches may be less likely to be misled by introgression than many other methods for detecting selection. Introgression introduces new genetic variation, but as long as the level of introgression is low, a previously rare allele introduced by introgression will not usually segregate at high frequencies — genetic drift or selection must act on the allele to increase its frequency. Therefore, allele frequency differentiation methods for distinguishing between selection and pure genetic drift are mostly applicable. However, care must be taken to ensure that the strong LD characteristic of introgressed regions[45,56] does not create false positives. For example,

© 2015 Macmillan Publishers Limited. All rights reserved

Box 2 | **Description of the *S** statistic**

*S** is a both a measure of linkage disequilibrium (LD) in a set of phased chromosomes and a method to discover linked single-nucleotide polymorphisms (SNPs). Part **a** of the figure shows a region within a sample of phased chromosomes, and a shared haplotype is depicted in red. Mismatches within the haplotype are denoted as black segments internal to the red region, and regions where there is no longer a shared haplotype are depicted as black flanking regions. *S** is calculated by optimizing the sum of scores *S*(*i*,*j*) over all subsets of SNPs at a given locus, where *i* and *j* are two successive (not necessarily adjacent) SNP positions in a given subset being considered for inclusion in the final haplotype[5,6,19,42]. *S*(*i*,*j*) is a heuristic score, but it 'rewards' fully linked pairs of sites (d(*i*,*j*) = 0, where d(*i*,*j*) is the number of haplotypes that differ between position *i* and *j*), and the increase in reward is proportional to the distance between the positions. Therefore, it tends to discover subsets of SNPs within a window that are in high LD, and the *S** value increases with increasing LD. In the process of calculating *S** using a dynamic programming method, not only is the optimal score calculated, but the set of SNPs that yield the optimum score — denoted by the two vertical grey lines defining the boundary of the region determined by *S** — is also calculated. Although there is some sharing of haplotypes outside this region, the number of mismatches would make *S** suboptimal, and they are therefore excluded.

Part **b** of the figure shows two examples of *S** calculations (using haplotype data for simplicity). Eleven positions (coordinates are in base pairs) and three haplotypes are shown in each example. Sites labelled 0 are ancestral (chimpanzee-like) and sites labelled 1 are derived. The *S** calculation proceeds via a dynamic programming algorithm and thereby guarantees that the optimal solution (in red) gives the maximal value for the sum of *S*(*i*,*j*) values; that is, there is no other set of SNPs that will give a higher sum of *S*(*i*,*j*) in each of these examples. In the example on the left, position 2,700 has higher linkage to its neighbouring SNPs than in the example on the right, and this site is therefore included in the haplotype. Additionally, positions 4,700 and 24,700 are linked; their distance is rewarded by *S*(*i*,*j*) and outweighs the penalty incurred from position 3,200 to 4,700. The particular scoring function *S*(*i*,*j*) that we apply in the example, the same form as developed in REF. 19, is shown in the box beneath, with heuristic choices for the penalty term, –10,000, and reward term, 5,000.



$$S(i,j) = \begin{cases} -\infty \text{ if } j \text{ is a singleton} \\ -\infty \text{ if } |i-j| < 10 \text{ or } |i-j| > 50{,}000 \\ 5{,}000 + |i-j| \text{ if } d(i,j) = 0 \text{ (no mismatches)} \\ -10{,}000 \text{ if } 1 \le d(i,j) \le 5 \text{ (up to 5 mismatches)} \end{cases}$$

$d(i,j)$ = Number of haplotypes that do not match on SNPs i and j
$|i-j|$ = Distance in base pairs between SNPs i and j

methods that summarize allele frequency differences from many linked SNPs within a region using sliding-window approaches could be biased or have increased variance because the allele frequencies of the introgressed alleles are more correlated than expected in the absence of introgression.

However, a tell-tale signature of adaptive introgression is the presence of mutations in strong LD that exist at high frequencies in a particular population and that are only present in the archaic source population, while being absent or at very low frequencies in other present-day human populations (FIG. 2). For example, a set of five such mutations cluster tightly together in the endothelial PAS domain protein 1 (*EPAS1*) gene in Tibetans, suggesting that archaic adaptive introgression has occurred[57]. Other candidates for adaptive introgression also display similar patterns of archaic allele sharing, including those found in REFS 19,44.

Balancing selection. The identification of introgressed regions that persist in humans owing to balancing selection can be a complicated task. Local admixture and balancing selection both generate increased intra-population variability and deep coalescent genealogies, which make it difficult to decipher whether one or both of these events affected a particular region of the genome. In one study[58], the authors looked for adaptive introgression via balancing selection in two ways. First, they searched for long human leukocyte antigen (*HLA*) haplotypes that are deeply divergent (showing a large number of differences with other haplotypes in the same locus), that have high sequence homogeneity in individuals that possess them and that show extreme LD only in non-African populations. The authors argued that these are unusual features that are unlikely to be predicted exclusively by a model of balancing selection: even though this type of selection generally acts to preserve sequence diversity, it should not produce long homogenous haplotypes that are deeply divergent from other haplotypes, especially in the *HLA* region, which is known to undergo rapid diversification by recombination. One variant fulfilling the authors' criteria (*HLA-B*73:01*) is also in strong LD with variants that are present in the Denisovan genome, suggesting that it may have introgressed from an archaic source. Simulations rejected a scenario of ancestral polymorphism but assumed a model without ancestral structure and without gene flow, which may not be realistic given the complexities of human demographic history. Their second approach was to computationally infer modern *HLA* haplotypes that may have been present in the Neanderthal and Denisovan genomes, and then analyse the global distribution of these haplotypes in present-day humans. The authors could not determine directly whether the haplotypes were actually present in the archaic genomes, as it is currently impossible to phase these sequences. Many of the haplotypes they inferred to be present in Neanderthals and Denisovans were found at high frequencies in different populations of Eurasia but at low frequencies or absent in Africa, which again suggests adaptive introgression.

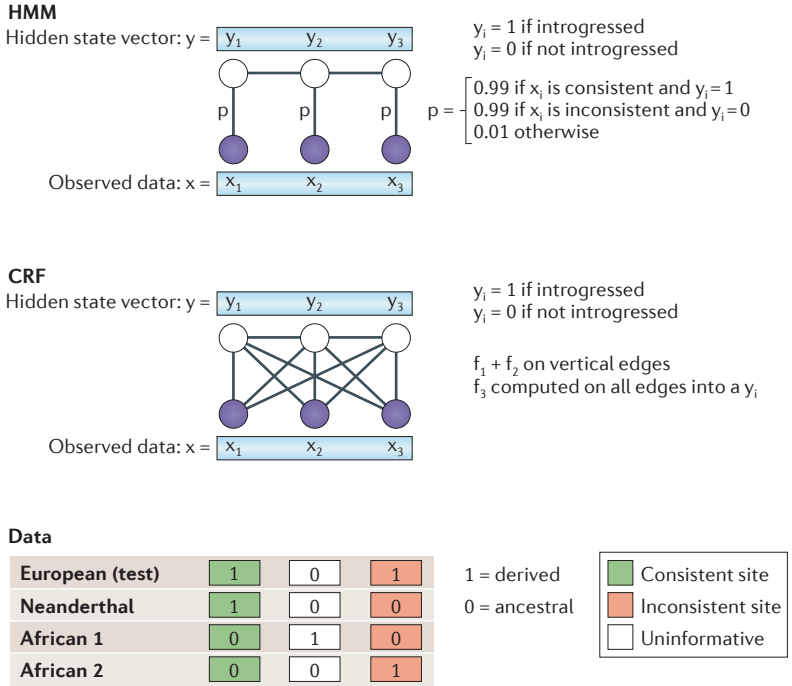
© 2015 Macmillan Publishers Limited. All rights reserved

Coalescent
Pertaining to coalescence: an event in the past at which two genetic lineages sampled in the present shared their most recent common ancestor at a specific locus in the genome.

Genotype–phenotype associations. If an introgressed variant is associated with a phenotype known to confer an advantage on a particular population, the variant may have undergone selection in that population. For example, the introgressed *EPAS1* gene in Tibetans contains SNPs associated with significantly reduced haemoglobin levels, which was possibly an adaptation to high-altitude hypoxia[59]. Similarly, a relationship between geographical location and genotype frequencies can be suggestive of environmental selection, as with haplotype frequency in the ultraviolet-B response gene *HYAL2* (hyaluronoglucosaminidase 2) and latitude[60]. However, a genotype–phenotype association by itself does not constitute enough support for adaptation[61], and additional tests must be performed to assess whether the genotype was actually driven to high frequencies due to selection.

Box 3 | **Hidden Markov model and conditional random field frameworks**

These are two related probabilistic models for estimating the ancestry (y_i) of a single-nucleotide polymorphism (SNP) *i* in a genome sequence. Each y_i takes two possible states: 1 for introgressed (that is, archaic ancestry) and 0 for not introgressed (that is, modern human ancestry). The observed data (x_i) are a matrix of haplotypes which, in the example, consist of whole-genome sequences of individuals from a European test population, an African population (with 2 haplotypes) and Neanderthals, each examined at 3 SNP positions (see the figure). x_1 is a site that is consistent with introgression, as the derived allele is seen in the test and Neanderthal sequences but not in the African sequences. Similarly, x_3 is an inconsistent site because the derived allele seems to be of modern human origin, whereas the x_2 site is uninformative. The ancestry states and observed data are connected through the emission probabilities (p) for the hidden Markov model (HMM) or emission functions (f_i) for the conditional random field (CRF); in the figure, these are denoted by the edges connecting the x_i and the y_i. The CRF can have more general relationships as denoted by the diagonal edges. Sankararaman *et al.*[44] used the sum of three emission functions f_1, f_2 and f_3, where f_1 scores consistent sites 1, f_2 scores inconsistent sites 1, and f_3 evaluates to 1 if the entire test haplotype is closer to the Neanderthal sequences than to the African haplotype. By contrast, the HMM used by Prüfer *et al.*[9] have fixed emission probabilities (p) for observed states that are either consistent or inconsistent. Edges between the y_i and y_{i+1} states represent transition probabilities in HMM or more general transition functions in CRF. The transition probabilities and functions model linkage between ancestral states along the genome, and the transition parameters depend on the recombination distance between sites, the admixture proportion and admixture time. Crucially, both of these frameworks have efficient algorithms for inferring the most likely sequence of ancestral states.



Candidates for adaptive introgression

Below, we review the evidence for adaptive introgression to date, in the context of the methods described in the two sections above. These recent studies either test single loci, or characterize adaptive introgression genome-wide, based on analyses of the available archaic human genomes (TABLE 1). We omit studies that had proposed adaptive introgression for some human haplotypes before the whole-genome sequences of Neanderthals and Denisovans[62,63] became available, as many of the early candidates were subsequently found to be absent in the genome sequences[8,64] (see John Hawk's blog post on Denisova microcephalin status).

The divergence times estimated from genome sequence data in several of the studies required use of a parameter for the human mutation rate. The precise estimate of the human mutation rate has been controversial and remains under study, and a range of accepted values are routinely used in current studies[65]. Therefore, whenever possible, we state the human mutation rate assumed in each paper when reporting divergence times and note that divergence times obtained from the more commonly used faster rate (10^{-9} per base pair per year) can be converted to those that would be obtained from the slower rate (0.5×10^{-9} per base pair per year) by simply doubling the original time.

Genome-wide admixture maps. Two studies[19,44] have produced Neanderthal genome-wide admixture maps of European and Asian genomes based on the whole-genome sequence of a female Neanderthal from Siberia[9] and present-day human data from the 1000 Genomes Project[66]. In one study, the authors used a CRF model[44] (BOX 3), whereas in the in the other study the authors relied on the *S** statistic to identify introgressed segments[19] (BOX 2). Although these two studies find notable signatures of negative selection pushing introgressed variants to low frequencies and a depletion of introgressed tracts in functional genomic regions, possibly owing to hybrid sterility, they both found considerable evidence of adaptive introgression. Several examples are outlined below. Additionally, one of the studies performed a functional enrichment analysis and found that genes involved in keratin filament, sugar metabolism, muscle contraction and oocyte meiosis have been involved in archaic human adaptive introgression[44].

A different genome-wide study (reviewed below) used the *D* statistic locally to detect Neanderthal-like regions at specific genes[67]. The authors found that lipid catabolism genes tended to have significantly higher *D* values in Europeans than the rest of the genome, supporting a tree clustering Neanderthals with Europeans to the exclusion of Africans. Additionally, signals of recent positive selection were reported in the same regions with high *D* values in modern Europeans[67].

© 2015 Macmillan Publishers Limited. All rights reserved

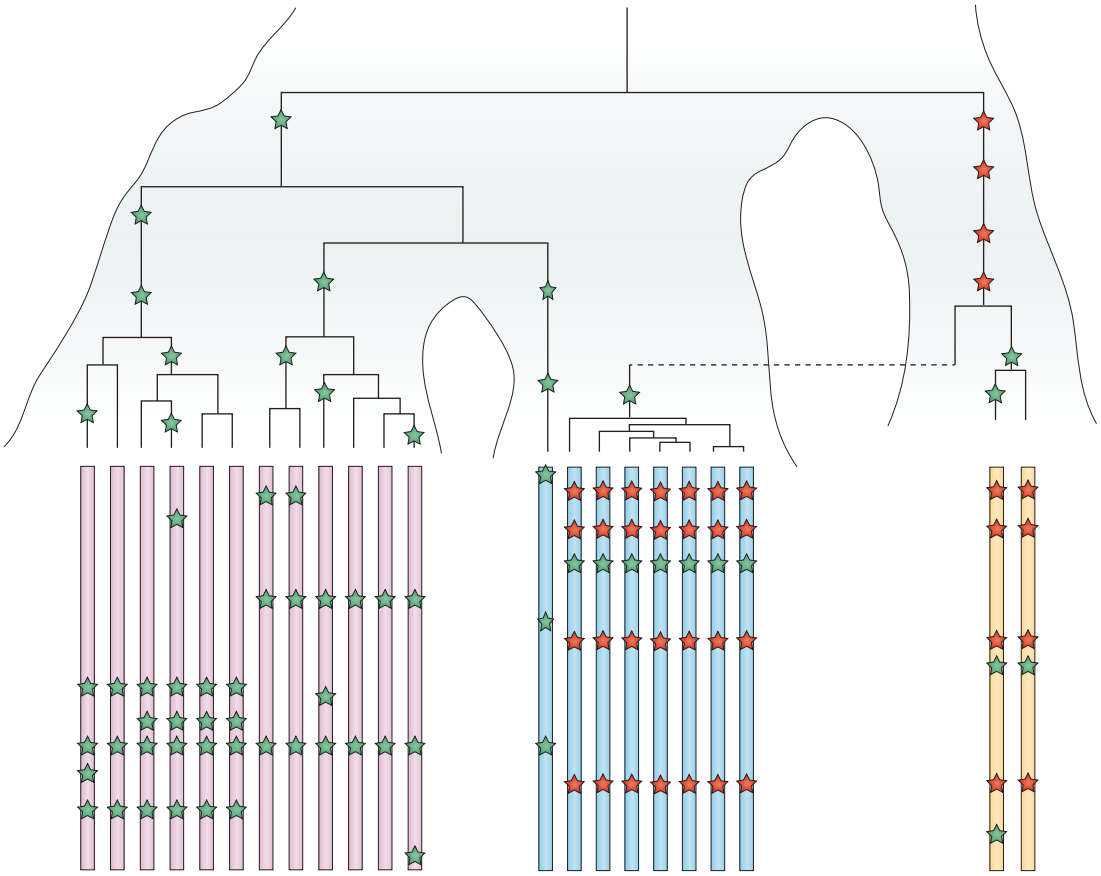


Figure 2 | **Example coalescent genealogy of uniquely shared mutations.** Several DNA fragments from two modern populations (pink and blue chromosomes) are sequenced. A diploid sequence is also obtained from an extinct archaic population (yellow chromosomes) that split from the population tree more anciently than the two modern populations split from each other. Uniquely shared mutations (red stars) occur in the archaic population but are passed on to the ancestors of the blue modern population via admixture (dashed line). These are then swept to high frequency by selection, producing a shallow local coalescent genealogy. This process results in sites with high-frequency derived alleles in the blue samples that are present in the archaic sample but not in the pink samples from the other modern population. Mutations in the genealogy that are not uniquely shared are shown as green stars.

Finally, present-day human genome data were used in HMMs to identify putatively introgressed haplotypes in two other studies[9,31]. These methods identified present-day human haplotypes that are phylogenetically consistent with forming a clade with an archaic individual (Neanderthal or Denisovan). The genes contained in the identified regions were not characterized, and tests of selection were not performed in these cases. However, the divergence between the introgressing Neanderthal in Eurasians and the sequenced Neanderthal genome (77–114 kya assuming a mutation rate of 0.5×10^{-9} per base pair year or 38–57 kya assuming a rate of 10^{-9} per base pair per year) was much younger than the divergence between the introgressing Denisovan in the Sahul people (Oceanians, including native Papuans and Australians) and the sequenced Denisovan genome (276–403 kya assuming the slower mutation rate or 138–202 kya assuming the faster rate)[9]. This may indicate that the Denisovan population was more diverse and/or more structured than the Neanderthal population.

Examining the candidate genes identified from both the genome-wide studies and the single-loci studies, we can speculate on the plausible selection pressures and mechanisms underlying examples of adaptive introgression. We organize our discussion of candidate genes below by broad functional categories.

Defence against pathogens. Signal transducer and activator of transcription 2 (*STAT2*) is an innate immune gene that is involved in interferon response after viral infection. A recent study[41] found a long (130–260 kb) haplotype called 'N' that overlaps this gene and that is broadly distributed across Eurasia. The haplotype is absent in sub-Saharan African populations, has a very ancient TMRCA with other present-day human haplotypes (609 kya assuming the faster mutation rate) and also has a very recent TMRCA with the Neanderthal genome (78 kya), suggesting that it introgressed from this archaic hominin group. Although found throughout Eurasia at a frequency of ~5%, it is present at substantially higher frequencies in Papuans (~54%). The authors

Ancestral polymorphism
A present-day polymorphism that exists at a site or haplotype because more than one allele existed in the ancestor of the two populations before they diverged from each other.

Hybrid sterility
Reduced viability or fertility of offspring from a mating between individuals from populations or species that diverged a long time ago; it is often due to incompatible mutations that occurred in each daughter population after they separated from each other.

© 2015 Macmillan Publishers Limited. All rights reserved

used a neutrality test that controlled for demography to show that this difference in frequency was likely to be a result of positive selection acting on the region in Papuans. Although the study focused on *STAT2* because of the greater availability of sequence data in this gene, the introgressed N haplotype overlaps two other genes that could also have been the targets of selection: *ERBB3* (encoding a receptor tyrosine kinase involved in cell growth and apoptosis)[68] and *ESYT1* (encoding extended synaptotagmin-like protein 1, a transmembrane protein with a role in fibroblast differentation)[69,70].

The highly polymorphic *HLA* region on chromosome 6 also seems to show signatures of adaptive introgression, although in this case by balancing selection. The fact that the *HLA* region is highly prone to *trans*-specific polymorphisms makes it difficult to distinguish adaptive

Table 1 | **Candidates for adaptive introgression**

Introgressed haplotype	Closest putative archaic source population	Modern populations showing evidence of introgression	Evidence for positive selection and selection tests performed	Most likely population in which selection occurred
HLA-A, *HLA-B* and *HLA-C*	Neanderthal and Denisovan	Europeans, East Asians and Melanesians	Extreme allelic and haplotypic diversity in the *HLA* region, which is indicative of balancing selection[58]	Europeans, East Asians and Melanesians
HLA-DPB1	Neanderthal (?), but this was questioned in REF. 73	Europeans (?)	No formal test of neutrality performed[72], but a phylogenetic analysis suggests that the haplotype is not introgressed[73]	–
STAT2 (haplotype N)	Neanderthal	Non-Africans	One-tailed test for elevated frequency of diagnostic SNP in introgressed haplotype based on empirical distribution of SNP frequencies[41]	Melanesians
STAT2 (haplotype D)	Denisovan (?)	Melanesians	No formal test of neutrality performed, but haplotype is only present in Papuans at a low frequency[41]	–
OAS1	Denisovan, but extremely ancient coalescence with human reference suggests that the direct source was a different archaic group	Melanesians	• No evidence for positive selection in REF. 75 • Evidence for positive selection in REF. 44 • Suggestive signal of balancing selection based on genetic differentiation of haplotypes across continents[74]	–
OAS gene cluster	Neanderthal	Non-Africans		
HYAL2 (3p21.31)[60]	Neanderthal	East Asians	iHS[46], EHH[47] and CMS[51]	East Asians
MC1R[78]	Neanderthal	Non-Africans	Tajima's D[49], Fu and Li's test[82], and iHS[46]	Taiwanese (?)
SLC16A11 and *SLC16A13*	Neanderthal	Native Americans	Genotype–phenotype association[90]	Native Americans (?)
DMD	Neanderthal	Non-Africans	No tests performed[91,92]	–
EPAS1 (REF. 57)	Denisovan	East Asians or Tibetans only	• High population differentiation[55] • Genotype–phenotype association[55] • High archaic haplotype frequency[57]	Tibetans
Various regions identified via S* that contain genes involved in the integumentary system	Neanderthal	Europeans and East Asians	• High F_{ST} between Europeans and Asians in variants identified to be introgressed (local adaptation)[19] • Introgressed haplotypes with frequencies in both Europeans and Asians that are higher than expected under neutrality (shared adaptation)[19]	Europeans and East Asians
Various regions identified via CRF that contain genes involved in keratin filament, sugar metabolism, muscle contraction and oocyte meiosis	Neanderthal	Europeans and East Asians	Windows in which the population frequency of the Neanderthal genetic material is too high to be explained by neutral drift[44]	Europeans and East Asians
Various regions identified via HMM	Neanderthal and Denisovan	Europeans, East Asians and Melanesians	No claims about selection and no formal tests of neutrality performed[9,31]	–

CMS, composite of multiple signals; CRF, conditional random field; *DMD*, dystrophin; EHH, extended haplotype homozygosity; *EPAS1*, endothelial PAS domain protein 1; *HLA*, human leukocyte antigen; HMM, hidden Markov model; *HYAL2*, hyaluronoglucosaminidase 2; iHS; integrated haplotype score; *MC1R*, melanocortin 1 receptor; *OAS*, 2′-5′-oligoadenylate synthetase; *SLC*, solute carrier; SNP, single-nucleotide polymorphism; *STAT2*, signal transducer and activator of transcription 2. The table provides a summary of recent studies reporting adaptive introgression with a comparison of the evidence provided in support of positive selection acting on the introgressed regions. A question mark denotes that the source population, the receiving population or the population in which selection occurred remains unresolved and/or has conflicting reports from different studies.

© 2015 Macmillan Publishers Limited. All rights reserved

introgression from balancing selection for different variants persisting over long time periods[71]. Various *HLA* haplotypes have been identified in Eurasians and Melanesians carrying functional variants that are likely to have been introgressed from archaic human groups[58]. For example, a deeply divergent haplotype (*B*73*) with strong sequence homogeneity among its carriers is present at high frequencies in populations in West Asia but absent or infrequent in the rest of the world's populations. Simulations show that the global distribution of this haplotype is best explained by a model of introgression from an archaic source in West Asia[58]. The Denisovan genome[11] carries two variants (*C*12:02* and *C*15*) that are in strong LD with *B*73*, although the actual *B*73* allele is absent in the Denisovan genome. These variants are also only found at high frequencies in West Asia. A separate analysis revealed a substantial Neanderthal *HLA-A* ancestry in modern non-African humans that has a wider geographical distribution than *B*73* (REF. 58). In summary, unusual levels of LD and high archaic haplotype frequencies at the *HLA* locus in present-day humans seem to support the hypothesis of adaptive introgression.

Another *HLA* study examined an important amino acid motif in the family of antigen receptors encoded by genes in the *HLA* region (*HLA-DPB*; allele *DPB1*0401*) that is required for allowing unmatched DRα and DPβ subunits to form a functional complex. The underlying DNA sequence closely matches the Neanderthal genome at the same locus[72]. However, a phylogenetic analysis of different variants of this motif in modern humans found that the putatively introgressed haplotype coalesced with other common modern human haplotypes before coalescing with the Neanderthal genome[73]. Furthermore, although the motif variant is present at a frequency of 68% across Europeans, it is also present in sub-Saharan Africans at a frequency of 11%. Finally, the divergence time between the haplotype of interest and the Neanderthal genome was estimated to be 2.2 million years ago (assuming a mutation rate of 10^{-9} per base pair per year), which largely predated the modern human–Neanderthal population split time. Thus, it is more likely that the presence of this haplotype in Europeans is a consequence of ancient population structure in Africa[73].

Other regions of the human genome show characteristic signatures of introgression, but the evidence for positive selection is weak or absent. For example, the 2′-5′-oligoadenylate synthetase (*OAS*) gene cluster, which encodes proteins that help to inhibit viral replication as part of the innate immune response, has been subjected to two separate archaic introgression events: one into non-Africans from Neanderthals (although the haplotype is absent in Papuans)[74] and another one into Papuans from an extremely deep lineage (which has a TMRCA with modern humans of 3.7 million years ago, assuming a mutation rate of 10^{-9} per base pair per year)[75]. The haplotype for the latter matches best with the Denisovan sequence across 90 kb of high-LD sequence and has much higher sequence diversity than that observed in other populations[75]. Although no evidence for positive selection in the region has been found[74,75], the genetic distribution of the different haplotypes across continents is perhaps a signal of balancing selection[74]. The deep lineage could belong to another archaic hominin group that may have admixed with both Papuans and Denisovans, or with a common ancestor of the two populations[75]. In the same direction, but in a different study, Prüfer *et al.*[9] recently found a genome-wide signal of 'super-archaic' introgression in the Denisovan genome, which is likely to be attributable to an unsampled hominin group that diverged from modern humans before Denisovans and Neanderthals, and that later interbred with Denisovans. However, the specific regions of the Denisovan genome that have super-archaic ancestry have yet to be identified, so it is unknown whether the *OAS* gene cluster is one of them. Finally, evidence for adaptive introgression into Europeans from Neanderthals in *OAS2* was more recently confirmed using the CRF model[44].

Pigmentation. In region p21.31 of chromosome 3, there is a 200-kb haplotype of Neanderthal origin that has a high frequency (>49%) in the East Asians sequenced as part of the 1000 Genomes Project[60]. The introgressed region shows very high LD and significantly high values of the iHS statistic[46], which measures extended haplotype homozygosity and is a hallmark of a recent selective sweep. However, as mentioned above, it is unclear how the iHS score would be affected by admixture in the absence of selection. One of the most likely targets of selection is a nonsynonymous SNP in *HYAL2*, which is involved in the cellular response to ultraviolet radiation. This SNP is absent in other non-African populations, so it seems to have been lost in the ancestors of Eurasians after migrating out of Africa but regained in East Asians via admixture with Neanderthals. The authors performed a bootstrapped phylogenetic analysis to support the shared ancestry of the haplotype with the Neanderthal sequence and obtained a significant *P* value for the observed LD value compared to a null model without introgression. Its frequency distribution shows a weak latitudinal gradient, suggesting that it was involved in the adaptive response to ultraviolet radiation as modern humans expanded throughout Asia[60]. A putative signal of adaptive introgression in East Asians in *HYAL2* has also been identified using the CRF framework[44].

Basonuclin 2 (*BNC2*) seems to be a strong candidate for adaptive introgression, as shown in two genome-wide archaic ancestry analyses[19,44]. Sankararaman *et al.*[44] applied the CRF model to detect introgressed segments and then inferred selection based on departures from a null model of neutrally introgressed alleles. Vernot and Akey[19] also found the introgressed region in an analysis using the *S** statistic, confirmed its ancestry by matching it with the Neanderthal genome and finally inferred selection by observing that the region has high levels of differentiation between Europeans and Asians, as measured by F_{ST}. A SNP in *BNC2* is associated with skin pigmentation[76] and freckling in Europeans[77], and the archaic haplotype is present at 70% frequency in Europeans but is absent in Asians. Interestingly, both studies also found a strong adaptive introgression signal in a cluster of keratin genes on chromosome 12 in both Asians and Europeans[19,44].

© 2015 Macmillan Publishers Limited. All rights reserved

Two neighbouring genes (*POU2F3* and *TMEM136*) have significant evidence for adaptive introgression in East Asians only, and this is again based on the two genome-wide archaic ancestry analyses[19,44] (see above). *POU2F3* encodes a transcription factor that mediates keratinocyte differentiation and proliferation, and the archaic haplotype is at 66% frequency in East Asians but almost absent in Europeans. *TMEM136* encodes a transmembrane protein, but little information is available about its function[44].

Ding *et al.*[78] identified an introgressed haplotype of Neanderthal origin in Eurasians carrying a loss-of-function variant (Val92Met) in MC1R, a melanocyte-stimulating hormone receptor. The *MC1R* gene is known to affect hair colour in mice[79] and is associated with red hair, freckles and type I/II fair skin type in humans[80,81]. However, analyses using Tajima's D[49], Fu and Li's test[82], and iHS[46] found that this locus exhibits no significant departures from neutrality at the introgressed region in Europeans or East Asians, presumably because the frequency of the archaic haplotype only ranges from 5% to 22%. In addition, the loss-of-function variant (Val92Met) is not actually seen in the high-coverage Neanderthal genome[9], despite being almost exclusively observed within haplotypes inferred to be introgressed from Neanderthals in Eurasian populations. The variant is also present in three African HapMap samples[83], which weakens the argument for introgression into Eurasians, unless the variant was later introduced into Africans via admixture from Eurasians. Intriguingly, the same variant is found at very high frequencies in Taiwanese aborigines (60–70%), but lack of extensive sequence data at this locus has prevented formal rejection of neutrality at the putatively introgressed haplotype in these populations[78].

Altitude. *EPAS1* encodes a transcription factor that has a role in the response to hypoxia at high altitudes[57]. This gene was previously identified as being under positive selection in Tibetans in several studies[55,59,84–88], but it did not fit with a simple model of selection from *de novo* mutation or from standing variation[57]. A set of 5 *EPAS1* intronic mutations in a 32-kb window are present in the Denisovan genome and at high frequency (~80%) in Tibetans, whereas they are absent in all of the individuals sequenced as part of the 1000 Genomes Project[66] with the exception of 2 Han Chinese individuals[57]. A haplotype network of present-day human and Denisovan haplotypes revealed that the Denisovan individual contained the closest-matching sequence to the introgressed haplotype. Their statistical evidence for introgression is based on highly significant *D* statistics, *S** statistics (BOX 2; see Supplementary information S1 (box)) and haplotype length. The unusually long haplotype length also rules out ILS (BOX 1) as a source of similarity to the Denisovan haplotype. Positive selection was supported by highly significant local differentiation based on the PBS[55] statistic. Furthermore, the putatively selected Tibetan variants in *EPAS1* are significantly associated with haemoglobin concentrations, a phenotype that distinguishes Tibetans from lowland populations[54]. Denisovan ancestry throughout the genome of East Asian populations is very low[9] (~0.2%), which suggests that selection was important in maintaining this specific haplotype at such high frequencies in Tibetans.

Metabolism. A recent study[67] proposed that Neanderthal alleles in lipid catabolism genes have been targets of recent positive natural selection in Europeans. The *D* statistic was used locally to detect regions of the genome that support a gene tree in which a particular human population clustered with Neanderthals to the exclusion of other human populations[89]. Lipid catabolism genes were significantly enriched for high *D* values (compared to the genome-wide expectation) in all European populations tested but not in Asians or Africans. The authors then computed CMS scores[51] (which test for positive selection) along the genome, and these showed a significant positive correlation with the *D* value for genes within the lipid catabolism gene set. However, it is possible that other factors (such as local effects of admixture or population structure) could have influenced the CMS score. The authors used phenotypic information to corroborate the result: they observed significantly diverged levels of lipid concentrations (by mass spectrometry) and lipid metabolic enzyme gene expression (by RNA sequencing) in brain tissues of Europeans compared to Asian or African populations.

Genes encoding solute carriers (*SLC16A11* and *SLC16A13*) are also implicated in lipid metabolic processes and have been found to harbour an introgressed haplotype. A genome-wide association study for type 2 diabetes in 8,214 individuals from Mexico and other parts of Latin America identified a novel association with the disease in a region that spanned these genes[90]. The strongest association was seen in a five-SNP haplotype within *SLC16A11*, with four of the SNPs causing an amino acid change. The haplotype is at high frequency in Mexican populations, at lower frequency in Asians and very rare or absent in Europeans and Africans. Estimates of divergence time of the haplotype between Mexicans and Europeans of 799 kya (assuming a mutation rate of 10^{-9} per base pair per year) suggested that the haplotype may have archaic origins. They found that all of the five SNPs are present and homozygous in the high-coverage Neanderthal sequence[9]. They provided two further lines of evidence for ancient admixture. First, over an extended 73-kb region around the 5 SNPs, those individuals with the 5-SNP haplotype have a TMRCA of 250 kya to Neanderthals compared to a TMRCA of 677 kya to Neanderthals for the individuals who do not have the 5-SNP haplotype. Second, the length of the haplotype is significantly longer than that expected from a split with Neanderthals unless a more recent admixture had occurred. As a functional assessment, the authors carried out a combination of *in vitro* experiments with the introgressed *SLC16A11* sequence, and determined endogenous intracellular localization and tissue-specific expression of this solute carrier, all of which indicated a role for the protein in hepatic lipid metabolism. They proposed that selection may be acting on the locus owing to the high frequency of the haplotype but carried out no further analyses in this regard. We can speculate that the change in lipid metabolism caused by introgression may

© 2015 Macmillan Publishers Limited. All rights reserved

have conferred an evolutionary advantage on the recipient population, possibly in relation to an altered diet.

Uncharacterized function. For some adaptive introgression candidates, it is less clear what the selective pressure may have been. One example is the dystrophin (*DMD*) gene, which was identified by Zietkiewicz *et al.*[91] as having an unusually diverged 8-kb X-linked haplotype, B006, and confirmed by both Sankararaman *et al.*[44] and Yotova *et al.*[92] to have come from Neanderthals. The B006 haplotype is found in all non-African populations at low but considerable frequencies (7% in Middle Eastern individuals, 12.9% in Europeans, 4.1% in Asians and 17.6% in Native Americans) and is extremely rare in sub-Saharan African populations (0.7% frequency on average). Therefore, it may have undergone weak positive selection in non-Africans, but there is no formal evidence in favour of this hypothesis yet. The DMD protein is an important structural component of skeletal muscle, although the confirmed archaic SNP is non-coding (intronic), and its functional effect remains uncertain.

Future studies

Although many individual loci have been proposed to be adaptively introgressed, the total extent and relative contribution of adaptive introgression in human evolution remain unknown. Some of the recently reported candidate adaptive introgression regions have limited or no clear evidence for positive selection. Furthermore, in certain cases, ILS or ancestral structure has not been definitively ruled out. This requires more extensive analyses: for example, by simulating data under alternative scenarios that do not include selection and/or introgression to derive more-rigorous null distributions of the statistics being used to investigate hypotheses regarding adaptive introgression.

A further caveat is that the true donor archaic population of an introgressed haplotype may be confounded by ILS in the population ancestral to multiple archaic (but not modern) human groups and/or by gene flow between archaic populations. It may be difficult to distinguish between Neanderthal and Denisovan admixture in genomic regions where these populations are not highly differentiated.

Although we have focused here on admixture between modern and archaic humans, there is well-documented evidence for admixture between different groups of modern humans in both recent[93,94] and ancient times[38,95]. This type of genetic exchange between modern humans has already been proposed to have facilitated adaptation to harsh environments[96,97]. However, the extent of adaptive gene flow among modern human populations remains an open question. For example, a recent study of nearly 30,000 African Americans found no evidence for adaptive gene flow in this recently admixed population[98]. As most genetic variation is shared among modern human populations, there is less opportunity for gene flow to introduce new genetic variation that was previously absent in the recipient population.

Interestingly, there is also genetic evidence for admixture between different archaic groups. The sequencing of the Neanderthal and Denisovan genomes at high coverage identified signals of admixture from eastern Neanderthals into Denisovans as well as into Denisovans from an unknown super-archaic group that diverged from present-day humans before the Neanderthals and Denisovans[9,99,100]. The introgressed regions for most of these events have yet to be identified but, once found, they may reveal important adaptations that were shared between archaic humans. For example, Neanderthal introgression in Denisovans has been identified at loci that include the CRISP cluster of genes, which are known to have a role in sperm maturation and egg fertilization, as well as the *HLA* region[9].

All instances of putative archaic adaptive introgression published so far have been found by detecting signatures of selection in regions previously known to be introgressed; detecting signatures of introgression in regions known to be selected; or observing archaic haplotypes in regions identified in genome-wide association studies. There is currently no method that can jointly infer the probability of admixture and selection in a particular region. One approach could be to develop informative summary statistics or combinations of summary statistics (such as D, S^*, the number of uniquely shared sites and F_{ST}-based statistics) under different models of introgression and adaptation. These could be incorporated into a likelihood framework for adaptive introgression to scan the genome in search for this signal. To be truly able to discern archaic introgression from other forces, this method should also require that the length of archaic segments be consistent with archaic introgression (FIG. 1) under a realistic demographic scenario. Such a model-oriented approach could potentially allow estimation of informative parameters, such as selection coefficients and times of introgression.

Conclusions

The biological processes that seem to have been affected by adaptive introgression in humans closely mirror the processes that are typically found to be targeted by positive selection in humans and in other organisms. It appears that adaptive introgression has provided a rich reservoir of new genetic variation that has allowed humans to adapt rapidly to a variety of new environmental conditions. The studies reviewed here suggest that adaptive introgression should be considered as an important mode of selection in human population genetics and can provide fascinating insights into the evolution of our species.

However, the rigorous identification of regions affected by adaptive introgression in humans is not trivial: care should be taken to distinguish ancestral polymorphism from true introgression and to use appropriate null models that include introgression when testing for selection. The advent of larger numbers and a broader range of human genome sequence data, both modern and archaic, in the near future may allow us to further disentangle the details of the process of adaptive introgression in humans, ultimately advancing our understanding of the genetic basis of phenotypic differences among human populations.

Uniquely shared sites
Sites containing high-frequency derived alleles in a particular population that are also present in a distantly related population but that are absent or at low frequencies in other populations more closely related to the first population. Such sites serve as necessary, but insufficient, evidence for adaptive introgression from the distantly related population.

© 2015 Macmillan Publishers Limited. All rights reserved

1. Wolpoff, M. H., Wu, X. & Thorne, A. G. in *The Origins of Modern Humans: A World Survey of the Fossil Evidence* (eds Smith, F. H. & Spencer, F.) 411–483 (Liss, 1984).
2. Stringer, C. B. & Andrews, P. Genetic and fossil evidence for the origin of modern humans. *Science* **239**, 1263–1268 (1988).
3. Bräuer, G. The Afro-European sapiens hypothesis and hominid evolution in East Asia during the late Middle and Upper Pleistocene. *Cour. Forschungsinst. Senckenb.* **69**, 145–165 (1984).
4. Smith, F. H., Jankovič, I. & Karavanič, I. The assimilation model, modern human origins in Europe, and the extinction of Neandertals. *Quaternary Int.* **137**, 7–19 (2005).
5. Plagnol, V. & Wall, J. D. Possible ancestral structure in human populations. *PLoS Genet.* **2**, e105 (2006).
This study develops a novel statistic algorithm, *S, to find candidate linked SNPs within haplotypes that may have been introduced by admixture with archaic humans.**
6. Wall, J. D., Lohmueller, K. E. & Plagnol, V. Detecting ancient admixture and estimating demographic parameters in multiple human populations. *Mol. Biol. Evol.* **26**, 1823–1827 (2009).
7. Higham, T. *et al.* The timing and spatiotemporal patterning of Neanderthal disappearance. *Nature* **512**, 306–309 (2014).
8. Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
This paper reports the draft sequence of a Neanderthal genome. Analysis of this sequence provided the first estimates of the proportion of archaic admixture in present-day human genomes.
9. Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).
This paper reports the first high-coverage sequence of a Neanderthal genome. Analysis of this sequence allowed the identification of Neanderthal segments in present-day humans.
10. Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053–1060 (2010).
11. Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012).
12. Castellano, S. *et al.* Patterns of coding variation in the complete exomes of three Neandertals. *Proc. Natl Acad. Sci. USA* **111**, 6666–6671 (2014).
13. Arnold, M. L. & Martin, N. H. Adaptation by introgression. *J. Biol.* **8**, 82 (2009).
14. Rieseberg, L. H. Evolution: replacing genes and traits through hybridization. *Curr. Biol.* **19**, R119–R122 (2009).
15. Hedrick, P. W. Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Mol. Ecol.* **22**, 4606–4618 (2013).
This is an introductory review of adaptive introgression in non-human species, providing a comparison with other types of selection and a list of many specific examples.
16. Hawks, J., Cochran, G., Harpending, H. C. & Lahn, B. T. A genetic legacy from archaic *Homo*. *Trends Genet.* **24**, 19–23 (2008).
17. Hawks, J. & Cochran, G. Dynamics of adaptive introgression from archaic to modern humans. *PaleoAnthropology* **2006**, 101–115 (2006).
This is a visionary review that proposed adaptive introgression as a mode of adaptation in humans before the availability of archaic genome sequences.
18. Wall, J. D. *et al.* Higher levels of Neanderthal ancestry in East Asians than in Europeans. *Genetics* **194**, 199–209 (2013).
19. Vernot, B. & Akey, J. M. Resurrecting surviving Neandertal lineages from modern human genomes. *Science* **343**, 1017–1021 (2014).
This key paper estimates the sum total of Neanderthal segments in a large sample of present-day human genomes using an LD-based statistic, *S, and identifies segments that may be targets of positive selection.**
20. Reich, D. *et al.* Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *Am. J. Hum. Genet.* **89**, 516–528 (2011).
21. Skoglund, P. & Jakobsson, M. Archaic human ancestry in East Asia. *Proc. Natl Acad. Sci. USA* **108**, 18301–18306 (2011).
22. Hammer, M. F., Woerner, A. E., Mendez, F. L., Watkins, J. C. & Wall, J. D. Genetic evidence for archaic admixture in Africa. *Proc. Natl Acad. Sci. USA* **108**, 15123–15128 (2011).
23. Lachance, J. *et al.* Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter-gatherers. *Cell* **150**, 457–469 (2012).
24. Veeramah, K. R. & Hammer, M. F. The impact of whole-genome sequencing on the reconstruction of human population history. *Nature Rev. Genet.* **15**, 149–162 (2014).
25. Currat, M. & Excoffier, L. Strong reproductive isolation between humans and Neanderthals inferred from observed patterns of introgression. *Proc. Natl Acad. Sci. USA* **108**, 15129–15134 (2011).
26. Eriksson, A. & Manica, A. Effect of ancient population structure on the degree of polymorphism shared between modern human populations and ancient hominins. *Proc. Natl Acad. Sci. USA* **109**, 13956–13960 (2012).
27. Liang, M. & Nielsen, R. The lengths of admixture tracts. *Genetics* **197**, 953–967 (2014).
This paper presents important theoretical results on the expected length of admixture tracts and how they can be used to identify particular models of admixture.
28. Pool, J. E. & Nielsen, R. Inference of historical changes in migration rate from the lengths of migrant tracts. *Genetics* **181**, 711–719 (2009).
29. Gravel, S. Population genetics models of local ancestry. *Genetics* **191**, 607–619 (2012).
30. Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from Western Siberia. *Nature* **514**, 445–449 (2014).
31. Seguin-Orlando, A. *et al.* Genomic structure in Europeans dating back at least 36,200 years. *Science* **346**, 1113–1118 (2014).
32. Sankararaman, S., Patterson, N., Li, H., Pääbo, S. & Reich, D. The date of interbreeding between Neandertals and modern humans. *PLoS Genet.* **8**, e1002947 (2012).
33. Gronau, I., Hubisz, M. J., Gulko, B., Danko, C. G. & Siepel, A. Bayesian inference of ancient human demography from individual genome sequences. *Nature Genet.* **43**, 1031–1034 (2011).
34. Gravel, S. *et al.* Demographic history and rare allele sharing among human populations. *Proc. Natl Acad. Sci. USA* **108**, 11983–11988 (2011).
35. Harris, K. & Nielsen, R. Inferring demographic history from a spectrum of shared haplotype lengths. *PLoS Genet.* **9**, e1003521 (2013).
36. Yang, M. A., Malaspinas, A. S., Durand, E. Y. & Slatkin, M. Ancient structure in Africa unlikely to explain Neanderthal and non-African genetic similarity. *Mol. Biol. Evol.* **29**, 2987–2995 (2012).
37. Lohse, K. & Frantz, L. A. Neandertal admixture in Eurasia confirmed by maximum-likelihood analysis of three genomes. *Genetics* **196**, 1241–1251 (2014).
38. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
39. Durand, E. Y., Patterson, N., Reich, D. & Slatkin, M. Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* **28**, 2239–2252 (2011).
This is a detailed analysis of Patterson's *D* statistic to test for admixture with archaic humans.
40. Martin, S. H., Davey, J. W. & Jiggins, C. D. Evaluating the use of ABBA–BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* **32**, 244–257 (2014).
41. Mendez, F. L., Watkins, J. C. & Hammer, M. F. A haplotype at *STAT2* introgressed from Neanderthals and serves as a candidate of positive selection in Papua New Guinea. *Am. J. Hum. Genet.* **91**, 265–274 (2012).
42. Wall, J. D. Detecting ancient admixture in humans using sequence polymorphism data. *Genetics* **154**, 1271–1279 (2000).
43. Viterbi, A. J. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans. Inform. Theory* **13**, 260–269 (1967).
44. Sankararaman, S. *et al.* The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507**, 354–357 (2014).
This paper presents a framework that combines different DNA sequence features to identify Neanderthal segments in present-day humans and determine which of these segments may have been positively selected.
45. Moorjani, P. *et al.* The history of African gene flow into Southern Europeans, Levantines, and Jews. *PLoS Genet.* **7**, e1001373 (2011).
46. Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A map of recent positive selection in the human genome. *PLoS Biol.* **4**, e72 (2006).
47. Sabeti, P. C. *et al.* Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832–837 (2002).
48. Sabeti, P. C. *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918 (2007).
49. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).
50. Fay, J. C. & Wu, C. I. Hitchhiking under positive Darwinian selection. *Genetics* **155**, 1405–1413 (2000).
51. Grossman, S. R. *et al.* A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science* **327**, 883–886 (2010).
52. Barreiro, L. B., Laval, G., Quach, H., Patin, E. & Quintana-Murci, L. Natural selection has driven population differentiation in modern humans. *Nature Genet.* **40**, 340–345 (2008).
53. Lewontin, R. C. & Krakauer, J. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* **74**, 175–195 (1973).
54. Shriver, M. D. *et al.* The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. *Hum. Genom.* **1**, 274–286 (2004).
55. Yi, X. *et al.* Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* **329**, 75–78 (2010).
56. Chakraborty, R. Gene admixture in human populations: models and predictions. *Am. J.Phys. Anthropol.* **29**, 1–43 (1986).
57. Huerta-Sánchez, E. *et al.* Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* **512**, 194–197 (2014).
58. Abi-Rached, L. *et al.* The shaping of modern human immune systems by multiregional admixture with archaic humans. *Science* **334**, 89–94 (2011).
59. Beall, C. M. *et al.* Natural selection on *EPAS1* (*HIF2α*) associated with low hemoglobin concentration in Tibetan highlanders. *Proc. Natl Acad. Sci. USA* **107**, 11459–11464 (2010).
60. Ding, Q., Hu, Y., Xu, S., Wang, J. & Jin, L. Neanderthal introgression at chromosome 3p21.31 was under positive natural selection in East Asians. *Mol. Biol. Evol.* **31**, 683–695 (2014).
61. Gould, S. J. & Lewontin, R. C. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc. R. Soc.* **205**, 581–598 (1979).
62. Hardy, J. *et al.* Evidence suggesting that *Homo neanderthalensis* contributed the H2 *MAPT* haplotype to *Homo sapiens*. *Biochem. Soc. Trans.* **33**, 582–585 (2005).
63. Evans, P. D., Mekel-Bobrov, N., Vallender, E. J., Hudson, R. R. & Lahn, B. T. Evidence that the adaptive allele of the brain size gene microcephalin introgressed into *Homo sapiens* from an archaic *Homo* lineage. *Proc. Natl Acad. Sci. USA* **103**, 18178–18183 (2006).
64. Setó-Salvia, N. *et al.* Using the neanderthal and denisova genetic data to understand the common *MAPT* 17q21 inversion in modern humans. *Hum. Biol.* **84**, 633–640 (2012).
65. Ségurel, L., Wyman, M. J. & Przeworski, M. Determinants of mutation rate variation in the human germline. *Annu. Rev. Genom. Hum. Genet.* **15**, 47–70 (2014).
66. Abecasis, G. R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
67. Khrameeva, E. E. *et al.* Neanderthal ancestry drives evolution of lipid catabolism in contemporary Europeans. *Nature Commun.* **5**, 3584 (2014).
68. Baselga, J. & Swain, S. M. Novel anticancer targets: revisiting ERBB2 and discovering ERBB3. *Nature Rev. Cancer* **9**, 463–475 (2009).
69. Lalioti, V. *et al.* The atypical kinase Cdk5 is activated by insulin, regulates the association between GLUT4 and E-Syt1, and modulates glucose transport in 3T3-L1 adipocytes. *Proc. Natl Acad. Sci. USA* **106**, 4249–4253 (2009).

© 2015 Macmillan Publishers Limited. All rights reserved

70. Min, S. W., Chang, W. P. & Südhof, T. C. E-Syts, a family of membranous Ca^{2+}-sensor proteins with multiple C2 domains. *Proc. Natl Acad. Sci. USA* **104**, 3823–3828 (2007).
71. DeGiorgio, M., Lohmueller, K. E. & Nielsen, R. A model-based approach for identifying signatures of ancient balancing selection in genetic data. *PLoS Genet.* **10**, e1004561 (2014).
72. Temme, S. *et al.* A novel family of human leukocyte antigen class II receptors may have its origin in archaic human species. *J. Biol. Chem.* **289**, 639–653 (2014).
73. Ding, Q., Hu, Y. & Jin, L. Non-Neanderthal origin of the *HLA-DPB1*0401*. *J. Biol. Chem.* **289**, 10252 (2014).
74. Mendez, F. L., Watkins, J. C. & Hammer, M. F. Neandertal origin of genetic variation at the cluster of *OAS* immunity genes. *Mol. Biol. Evol.* **30**, 798–801 (2013).
75. Mendez, F. L., Watkins, J. C. & Hammer, M. F. Global genetic variation at *OAS1* provides evidence of archaic admixture in Melanesian populations. *Mol. Biol. Evol.* **29**, 1513–1520 (2012).
76. Jacobs, L. C. *et al.* Comprehensive candidate gene study highlights *UGT1A* and *BNC2* as new genes determining continuous skin color variation in Europeans. *Hum. Genet.* **132**, 147–158 (2013).
77. Eriksson, N. *et al.* Web-based, participant-driven studies yield novel genetic associations for common traits. *PLoS Genet.* **6**, e1000993 (2010).
78. Ding, Q. *et al.* Neanderthal origin of the haplotypes carrying the functional variant Val92Met in the *MC1R* in modern humans. *Mol. Biol. Evol.* **31**, 1994–2003 (2014).
79. Nachman, M. W., Hoekstra, H. E. & D'Agostino, S. L. The genetic basis of adaptive melanism in pocket mice. *Proc. Natl Acad. Sci. USA* **100**, 5268–5273 (2003).
80. Valverde, P. *et al.* The Asp84Glu variant of the melanocortin 1 receptor (*MC1R*) is associated with melanoma. *Hum. Mol. Genet.* **5**, 1663–1666 (1996).
81. Valverde, P., Healy, E., Jackson, I., Rees, J. L. & Thody, A. J. Variants of the melanocyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans. *Nature Genet.* **11**, 328–330 (1995).
82. Fu, Y. X. & Li, W. H. Statistical tests of neutrality of mutations. *Genetics* **133**, 693–709 (1993).
83. Altshuler, D. M. *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
84. Bigham, A. *et al.* Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. *PLoS Genet.* **6**, e1001116 (2010).
85. Simonson, T. S. *et al.* Genetic evidence for high-altitude adaptation in Tibet. *Science* **329**, 72–75 (2010).
86. Peng, Y. *et al.* Genetic variations in Tibetan populations and high-altitude adaptation at the Himalayas. *Mol. Biol. Evol.* **28**, 1075–1081 (2011).
87. Xu, S. *et al.* A genome-wide search for signals of high-altitude adaptation in Tibetans. *Mol. Biol. Evol.* **28**, 1003–1011 (2011).
88. Wang, B. *et al.* On the origin of Tibetans and their genetic basis in adapting high-altitude environments. *PLoS ONE* **6**, e17002 (2011).
89. Abecasis, G. R. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
90. Williams, A. L. *et al.* Sequence variants in *SLC16A11* are a common risk factor for type 2 diabetes in Mexico. *Nature* **506**, 97–101 (2014).
91. Zietkiewicz, E. *et al.* Haplotypes in the dystrophin DNA segment point to a mosaic origin of modern human diversity. *Am. J. Hum. Genet.* **73**, 994–1015 (2003).
92. Yotova, V. *et al.* An X-linked haplotype of Neandertal origin is present among all non-African populations. *Mol. Biol. Evol.* **28**, 1957–1962 (2011).
93. Moorjani, P. *et al.* Genetic evidence for recent population mixture in India. *Am. J. Hum. Genet.* **93**, 422–438 (2013).
94. Hellenthal, G. *et al.* A genetic atlas of human admixture history. *Science* **343**, 747–751 (2014).
95. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014).
96. Huerta-Sánchez, E. *et al.* Genetic signatures reveal high-altitude adaptation in a set of Ethiopian populations. *Mol. Biol. Evol.* **30**, 1877–1888 (2013).
97. Jeong, C. *et al.* Admixture facilitates genetic adaptations to high altitude in Tibet. *Nature Commun.* **5**, 3281 (2014).
98. Bhatia, G. *et al.* Genome-wide scan of 29,141 African Americans finds no evidence of directional selection since admixture. *Am. J. Hum. Genet.* **95**, 437–444 (2014).
99. Waddell, P. J. Happy New Year *Homo erectus*? More evidence for interbreeding with archaics predating the modern human/Neanderthal split. *arXiv* [online], http://arxiv.org/abs/1312.7749 (2013).
100. Waddell, P. J. & Tan, X. New g%AIC, g%AICc, g%BIC, and power divergence fit statistics expose mating between modern humans, Neanderthals and other archaics. *arXiv* [online], http://arxiv.org/abs/1212.6820 (2012).

Acknowledgements
E.H.-S. is supported by start-up funds from the University of California, Merced, USA. R.N and S.S are supported by the US National Institutes of Health (grants R01HG003229-09 and K99 GM111744). F.R. is supported by the US National Institutes of Health grant to M. Slatkin (R01-GM40282). The authors thank F. Casey for discussions and help with Box 2, as well as E. Durand for help with Supplementary information S1 (box).

Competing interests statement
The authors declare no competing interests.

FURTHER INFORMATION
John Hawk's blog post on Denisova microcephalin status: http://johnhawks.net/weblog/reviews/denisova/denisova-mcph1-2011.html

SUPPLEMENTARY INFORMATION
See online article: S1 (box)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

© 2015 Macmillan Publishers Limited. All rights reserved