Article

Evolutionary Genetic Signatures of Selection on Bone-Related Variation within Human and Chimpanzee Populations

Daryn A. Stover [1,2], Genevieve Housman [3], Anne C. Stone [4], Michael S. Rosenberg [5] and Brian C. Verrelli [1,5,*]

1   School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA; dastover@asu.edu
2   Arizona State University at Lake Havasu, Lake Havasu, AZ 86403, USA
3   Section of Genetic Medicine, University of Chicago, Chicago, IL 60637, USA; ghousman@uchicago.edu
4   School of Human Evolution and Social Change, Arizona State University, Tempe, AZ 85287, USA; acstone@asu.edu
5   Center for Biological Data Science, Virginia Commonwealth University, Richmond, VA 23284, USA; msrosenberg@vcu.edu
*   Correspondence: bverrelli@vcu.edu

Abstract: Bone strength and the incidence and severity of skeletal disorders vary significantly among human populations, due in part to underlying genetic differentiation. While clinical models predict that this variation is largely deleterious, natural population variation unrelated to disease can go unnoticed, altering our perception of how natural selection has shaped bone morphologies over deep and recent time periods. Here, we conduct the first comparative population-based genetic analysis of the main bone structural protein gene, collagen type I α 1 (*COL1A1*), in clinical and 1000 Genomes Project datasets in humans, and in natural populations of chimpanzees. Contrary to predictions from clinical studies, we reveal abundant *COL1A1* amino acid variation, predicted to have little association with disease in the natural population. We also find signatures of positive selection associated with intron haplotype structure, linkage disequilibrium, and population differentiation in regions of known gene expression regulation in humans and chimpanzees. These results recall how recent and deep evolutionary regimes can be linked, in that bone morphology differences that developed among vertebrates over 450 million years of evolution are the result of positive selection on subtle type I collagen functional variation segregating within populations over time.

Keywords: type I collagen; *COL1A1*; bone disease; BMD; osteoporosis; exon duplication; adaptation

1. Introduction

Bone-related disorders impact more than 200 million people globally [1–3]. In addition to geographic variation in disorders such as osteoporosis, bone strength in general—measured as bone mineral density (BMD)—also varies among populations, with individuals of African ancestry having better overall bone quality [4,5]. Although variation in bone strength is due in part to environmental differences [6,7], BMD is estimated to be as much as 85% heritable [8–10], with hundreds of loci and tens of thousands of variants identified [11–14]. While our understanding of bone-related variation linked to osteoporotic-related disorders and fractures is incredibly robust due to high-powered GWAS in case–control cohorts [15,16], we still have a limited understanding of natural population variation unrelated to disease, and what evolutionary significance it may have [17].

We expect that bone-related phenotypes are subject to strong purifying selection, yet variation linked to bone loss is not only common, but has likely been segregating in the population for the past 10 Ky [18], with variation attributed to ethnicity, age, and sex [19–21]. In fact, the observation of alleles related to increased BMD being significantly more common in sub-Saharan African populations has been attributed to positive selection driving ethnic differences in bone strength [22]. From a deeper evolutionary perspective, phenotypic data for non-human primates, though limited, suggest that genetic variation linked to bone

strength also varies within other species [23–25]. Chimpanzees, our closest-living relatives with whom we share a common ancestor ~5 Mya [26], show patterns of bone loss and microfractures with age [27–29]. With thousands of variants linked to BMD, a top priority is to link functional population variation to therapeutic treatment [16,30]. As such, an overlooked approach is the use of evolutionary genetics to test natural populations for signatures of adaptive functional variation in bone-related genes.

As an evolutionary genetic model, we focused on the collagen type I α 1 (*COL1A1*) gene, which encodes the most abundant protein in mammals and is the main structural protein of bone, teeth, and tendons [31]. *COL1A1* alone possesses more than 600 skeletal and connective tissue disease-associated mutations (DAMs), primarily linked to osteoporosis, osteogenesis imperfecta types I–IV, and Ehlers–Danlos syndromes [32,33]. *COL1A1* is also among the candidate genes commonly linked to natural phenotypic variation in bone strength across ethnic groups [34–36]. These factors make this locus a prime candidate for investigating evolutionary factors that shape bone-related phenotypic variation within and among populations and species over time [37].

From the *COL1A1* DAMs previously identified to date, we might conclude that amino acid polymorphism is rare in frequency and associated with disease [32,38]. One bias, however, is that these characterizations come from screenings of clinical individuals with known bone-related disorders. For example, the triple-helix domain of type I collagen is comprised of two COL1A1 and one COL1A2 subunits that wind together, with protein length mutations of any size predicted to have deleterious effects on helix stability [39–41]. However, type I collagen genes likely originated hundreds of millions of years ago through a series of duplication events from an ancestral collagen with a single 54 bp exon [42], confirming that variation in length may lead to innovative function [43]. Secondly, the majority of previously identified *COL1A1* DAMs are found in the triple-helix domain, which is a repeating amino acid sequence with glycine in every third position, implying that change to this compact winding structure is deleterious [44–46]. However, our previous evolutionary analysis of *COL1A1* sequences from vertebrates spanning ~450 My of divergence showed that the C-terminal domain is more evolutionarily conserved than the triple-helix domain, and the latter exhibits spatial heterogeneity in selective constraint that distinguishes severe from osteoporotic-like phenotypes [37]. Thus, while clinical studies reveal severe *COL1A1* DAMs, evolutionary models show that *COL1A1* variation is historically consistent with both positive selection and disease [47], and predict that natural populations may harbor variation that has functional and adaptive potential.

An additional observation from our previous comparative vertebrate analysis of *COL1A1* was that intron structure and content may be evolutionarily conserved [37], consistent with the need for high gene expression given the importance of type I collagen in vertebrate development and wound repair [48,49]. Indeed, functional studies have shown that sites in the *COL1A1* promoter and first intron regulate gene expression [34,35,50], yet we know little about the evolutionary significance of variation across the > 11 kb of 50 introns. One example comes from a single nucleotide polymorphism (SNP) in an Sp1 transcription factor binding site in the first intron that increases *COL1A1* expression and, likely, accounts for reduced structural integrity and BMD [51,52]. However, this SNP reaches frequencies > 20% in individuals of Western European ancestry, and has also been linked to reduced soft tissue damage [53], suggesting that selective forces are complex. Our previous evolutionary analysis did not examine rate variation in introns, because our approach was limited by a protein-codon-based model [37]. However, in the 10 years since, the field has seen technological and statistical advances, resulting in hundreds of vertebrate genome sequences and evolutionary conservation scores that can be applied to any non-coding nucleotide site [54].

Only one previous study of *COL1A1* in a natural population exists, but with a focus on exons in admixed Americans [55], it does not reveal information on non-coding regions, nor does it reflect varying selective pressures among globally diverse populations. Here, we conduct the first comparative population genetic analyses of *COL1A1* in natural populations

of humans and chimpanzees to address the following questions: (1) Is *COL1A1* amino acid polymorphism in natural populations reflective of clinical samples, in that it is highly deleterious and rare or, alternatively, is cryptic protein polymorphism common, as predicted by deep-time evolutionary analyses? (2) Does *COL1A1* intron nucleotide diversity in natural populations reflect high functional constraint or, alternatively, do we find signatures of adaptive evolution suggesting positive selection for gene expression variation? From a larger perspective, comparative population and species analyses help us to understand how deep-time and recent evolutionary pressures are linked and shape adaptive phenotypes vs. disease-related phenotypes across human populations.

2. Materials and Methods

2.1. Human Population Datasets

We accessed VCFs from the 1000 Genomes Project (www.internationalgenome.org (accessed on 10 December 2021)) [56] for our natural human population sample (Table S1). These data include 2504 individuals (i.e., 5008 *COL1A1* allele copies) with no known bone abnormalities (i.e., a random sample with respect to phenotypic diversity), representing 26 geographically and ethnically diverse populations ("1000G", hereafter). This sample reflects human genetic diversity within and outside of sub-Saharan Africa, the latter typically having higher nucleotide and haplotype diversity owing to its older, larger, and ancestral effective population size compared to the more recent demographic history of expansion associated with non-African groups [57–60].

The Leiden Open Variation Database [38] (previously the Osteogenesis Imperfecta and Ehlers–Danlos Syndrome Variant Databases) [32] was used to access hundreds of clinically relevant *COL1A1* disease-associated mutations ("DAMs", hereafter). The majority of *COL1A1* mutations are associated with osteogenesis imperfecta (OI), and are categorized according to clinical severity of disease symptoms [61]. Similar to our previous analysis [37], we filtered these OI mutations into severity categories of 1–4 (Table S2), from "category 1" OI—reflecting mild bone weakening similar to that caused by osteoporosis, and considered the least severe—to "category 4" OI, reflecting lethality, as the most severe.

2.2. Non-Human Primate Population Datasets

We accessed chimpanzee and bonobo raw data and VCFs from the publicly available population genomics project of the *Pan* genus ("*Pan* genomes", hereafter) [62,63]. This dataset includes individuals from the central (*Pan troglodytes troglodytes*), eastern (*P. t. schweinfurthii*), western (*P. t. verus*), and Nigeria–Cameroon (*P. t. ellioti*) chimpanzee and bonobo (*P. paniscus*) groups. Chimpanzees and bonobos have an estimated divergence time of ~1.6–2.1 Mya, with chimpanzee subspecies groups splitting ~139–633 Kya; however, a complex evolutionary history of gene flow and admixture has been found among groups [63]. Central chimpanzees show the highest nucleotide diversity and lowest long-range linkage disequilibrium (LD), owing to a potentially larger historical population size, whereas the western chimpanzee has consistently shown the lowest nucleotide diversity, no evidence of population structure, and greater separation from other subspecies [62–65].

From the *Pan* genomes dataset, we sampled 59 individuals from the 4 chimpanzee subspecies (Table S3). This sample comes after multiple filters were applied to the data, removing individuals and nucleotide sites that exhibited unusual patterns (i.e., deviations from Hardy–Weinberg equilibrium). The western subspecies represents the most appropriate contrast with humans, because of similar levels of nuclear population genetic diversity [66–70]. As such, we generated high-coverage nucleotide sequence data from our samples of 20 wild-born, unrelated western chimpanzees (40 allele copies) to complement the *Pan* genomes data. These data include the *COL1A1* locus (Table S4), as well as 20 targeted intergenic regions totaling > 30 Kb spread over 100 Kb in each of the upstream and downstream directions of *COL1A1* (Table S5) to examine long-range LD patterns.

We also sampled bonobo data from the *Pan* genomes data to align with the four chimpanzee subspecies from the same project. We generated high-coverage nucleotide

sequence data from 13 wild-born, unrelated bonobos (26 allele copies), to provide alignment for the homologous chimpanzee targeted intergenic regions upstream and downstream of *COL1A1* noted above (Table S6). All western chimpanzee and bonobo sequence data generation followed our previous protocol for targeted PCR, sequencing, and short-read assembly required with the complex repetitive nature of *COL1A1* [37]. Finally, we also used bonobo, orangutan (*Pongo abelii*), and macaque (*Macaca mulatta*) alignments from the UCSC Genome Browser (www.genome.ucsc.edu) [71] to make inferences about derived versus ancestral states of polymorphisms within human and chimpanzee lineages.

2.3. Statistical Analyses of COL1A1 Amino Acid Variation

For estimates of evolutionary conservation, we used the vertebrate track for phyloP scores in the UCSC Genome Browser (hg19 assembly), as type I collagen is one of the most abundantly expressed proteins found across all vertebrates. The phyloP statistic identifies rate variation as acceleration (faster) and conservation (slower), and assigns negative and positive scores, respectively, compared to neutrality [54]. The absolute values of the scores represent $-\log$ p-values under a null hypothesis of neutral evolution, i.e., the closer the value to "0", the more the evolutionary history of that site conforms to neutral expectations. The vertebrate track shows phyloP evolutionary conservation scores for individual nucleotides for 100 vertebrates, including 62 mammals and 12 primates.

For the DAMs dataset, we investigated the genomic position (hg19 assembly), amino acid change, COL1A1 protein domain, and phyloP score for each single nucleotide mutation (no length or splice variants) (Table S2). The same information was collected from the 1000G dataset, with additional data on the frequencies of the amino acid replacement SNPs within populations (Table S7). We also obtained the same information on amino acid variation within *COL1A1* for the *Pan* genomes dataset, as well as for our generated *Pan* sequence data. Finally, the "severity category" score was determined for each DAM (Table S2), following our previous criteria noted above [37].

We compared distributions of the data categories within and between the DAM and 1000G datasets, i.e., whether SNPs occurred at nucleotide sites with different phyloP scores with respect to severity category, domain, and amino acid change. These distributions were compared using nonparametric statistical tests evaluated for significance via permutations in the coin v. 1.4-2 R package [72].

Our previous evolutionary analyses of *COL1A1* investigated patterns of amino acid divergence across species [37], whereas here we contrast human and chimpanzee amino acid polymorphism. As levels of polymorphism and divergence are expected to be correlated under neutrality, we compared these two classes of variation at *COL1A1* amino acid replacement sites with *COL1A1* intronic sites as proxies for neutrality, using a 2×2 test of independence first applied by McDonald and Kreitman [73].

2.4. Statistical Analyses of COL1A1 Intronic Variation

As phyloP scores can be calculated for any nucleotide site, the vertebrate track for individual sites was also downloaded (similar to exons above) for the 50 introns of *COL1A1* to determine whether introns and sites underlying SNPs showed unusual evolutionary conservation. Similar to statistical analyses with the coding sequences above, distributions of phyloP scores were compared across datasets using nonparametric statistical tests and permutations to evaluate significance.

We conducted several population genetic analyses to investigate patterns of intronic variation within and between populations and species. First, population-specific estimates of nucleotide diversity for *COL1A1* introns were calculated as Watterson's [74] θ_S. Second, we used F_{ST} analyses to investigate patterns of genetic differentiation among human populations and among chimpanzee subspecies. Calculations were conducted using an R script based on the F_{ST} estimator from Hudson et al. [75] averaged across SNPs between population pairs and, as described by Hudson [76], we used a permutation analysis to identify significant outliers. This analysis pooled populations and randomly sampled allelic

variation using our same sample sizes to reconstitute the populations, after which pairwise F_{ST} values were calculated. This simulation was repeated 1000 times, with observed F_{ST} values compared to the simulated distributions. These analyses were conducted for the *COL1A1* locus and for SNPs up to 100 Kb upstream and downstream, using the 1000G data for humans and the *Pan* genomes data for chimpanzees.

Finally, to investigate patterns of haplotype structure, we also conducted population-specific LD analyses. For humans, we used the 1000G phased data, and SNP pairwise r^2 values were downloaded using VCFTOOLS v. 0.1.14 [77], after filtering out SNPs with minor allele frequency (MAF) <5%. Population-specific LD data were generated for the *COL1A1* locus, as well as for SNPs up to 100 Kb away in each of the upstream and downstream regions, to investigate long-range haplotype structure. For chimpanzees, we used PHASE v. 2.1.1 [78] to statistically resolve heterozygous sequences for each individual into two haplotypes, after which SNP pairwise r^2 values were generated using DnaSP v. 6.12.03 [79]. The PHASE analysis was repeated with 500, 750, and 1000 iterations with phased haplotypes from the run with the highest average goodness-of-fit used in subsequent analyses. The PHASE and LD analyses were first performed for the *Pan* genomes dataset for each of the four subspecies for *COL1A1*, as well as for SNPs up to 100 Kb away in each of the upstream and downstream regions to investigate long-range haplotype structure. We also performed the same PHASE and LD analyses for our high-coverage western chimpanzee sequence data at *COL1A1*, and for the intergenic regions we collected spanning 100 Kb each upstream and downstream (Tables S4–S6).

3. Results

3.1. Contrasts of Amino Acid Variation between the DAM and 1000G Datasets

From the 1000G dataset of 5008 global gene copies, we identified 60 amino acid replacement SNPs (Table S7) to compare with the 422 DAMs identified from the clinical database that met the criteria of documented severity categories of 1–4 (Table S2). Each of the 26 populations in the 1000G dataset includes at least one variant, and some include as many as seven (Table S1). Overall, these SNPs are rare in frequency (Table S8); however, we note that 14 of the 60 are found segregating in at least two populations, with one such variant shared across 17 populations and reaching a frequency as high as 7% (Table S7). Interestingly, 6 of these 60 SNPs (10%) occur at the same residues as documented DAMs (Table S9), yet severity can differ across ethnic and age groups [80], especially for "lethal" ones. For example, the Ala-390-Thr variant noted as a category 4 DAM is found segregating in six different populations, five of which are of African ancestry, where we typically expect strength of selection to be higher. However, we note that this variant has a low phyloP score (Table S9), implying a history closer to neutrality, and not strong purifying selection.

We next examined comparisons within and between the DAM categories and 1000G datasets to determine whether evolutionary site conservation distinguishes these groups. First, for some perspective, phyloP scores in the vertebrate track across the genome reach maximum positive values of approximately 10, whereas they can reach maximum negative values of −20 (data unpublished). Even for *COL1A1*'s relatively unusual short exons and introns [37], exon sites have become uncoupled evolutionarily, with significantly higher conservation (reaching the local maximum; Figure S1) compared to introns (Tables S10 and S11). Looking at these sites with respect to DAMs, we find that overall DAMs occur at *COL1A1* sites that are significantly more conserved, but that category 1 DAMs occur at sites with significantly lower conservation compared to categories 2–4 DAMs (Tables S10–S12). Amino acid SNPs in 1000G occur at sites that are significantly less conserved compared to all four DAM categories, and do not occur at *COL1A1* coding sites that are unusually conserved overall (Tables S10–S12). However, amino acid SNPs in 1000G are found at significantly lower frequencies (Table S8; permutation test, $p = 0.0006$), and at sites with significantly higher conservation compared to intron SNPs, which are expected to be putatively neutral in comparison (Tables S10 and S11).

We also found that category 2–4 DAMs are more often associated with the triple-helix domain, and least associated with the N- and C-terminal domains, compared to category 1 DAMs (X_2 = 56.9, $p < 0.001$). In addition, category 1 DAMs in the triple helix involve a mutated glycine residue in significantly fewer cases (57%) compared to category 2–4 DAMs (91–98%; Table S13 and Figure S2; Fisher's exact tests, $p < 10^{-5}$). Although the 1000G amino acid SNPs are less associated with the triple helix than category 2–4 DAMs (Table S12, $X_2 > 9$, $p < 0.05$), they are distributed across domains similar to category 1 DAMs ($X_2 = 1.14$, $p = 0.56$). That said, the 1000G amino acid SNPs involve a mutated glycine residue in a significantly smaller number of the triple-helix SNPs (2.8%) compared with DAMs (Table S13, Figure 1), even in the least severe category 1 (Fisher's exact tests, $p < 10^{-7}$).
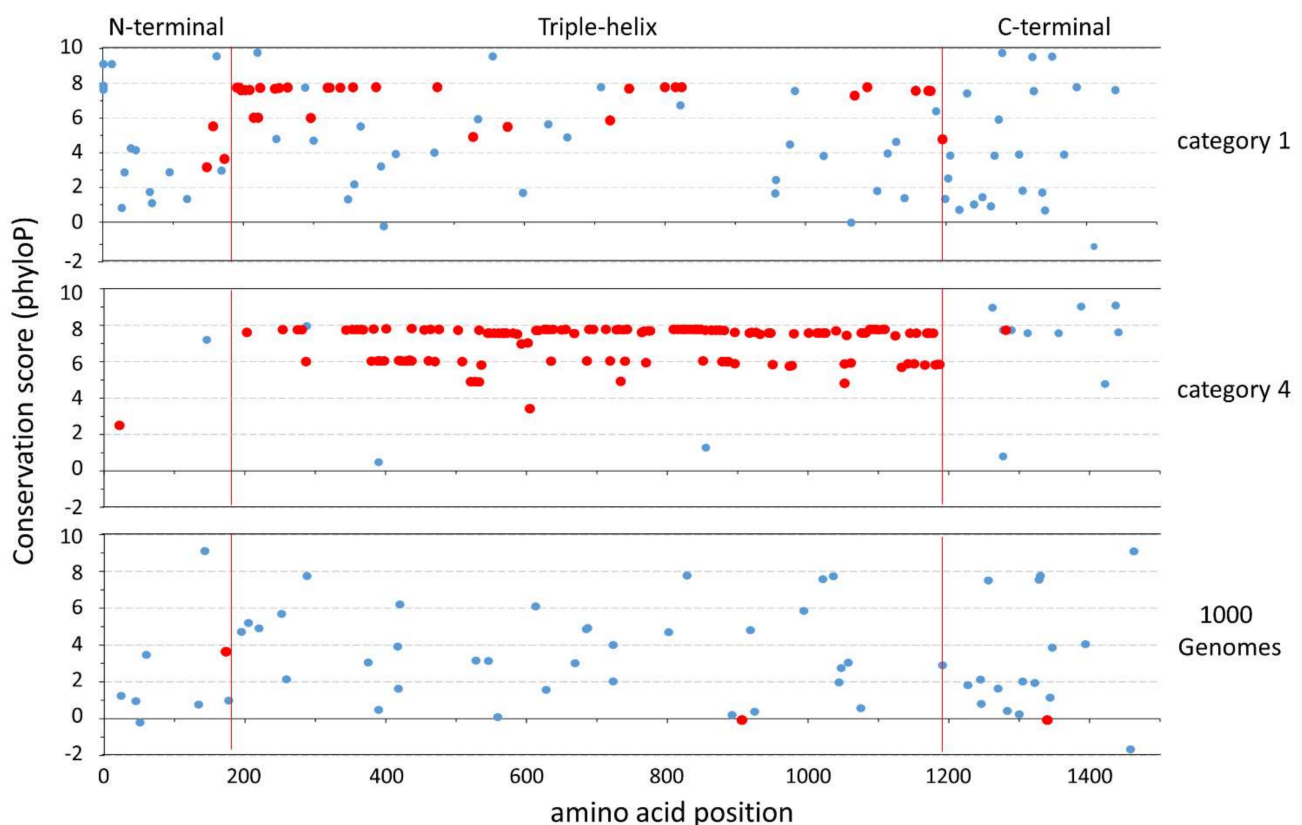

Figure 1. phyloP conservation scores plotted across human COL1A1: phyloP scores are shown for amino acid mutations in disease severity categories 1 (least severe) and 4 (most severe), and in the 1000 Genomes Project dataset across different protein domains (see Table S2). Mutations noted in "red" occur at glycine residues, whereas mutations noted in "blue" occur at non-glycine residues. As phyloP deviates from "0", the positive and negative scores reflect evolutionary conservation and acceleration, respectively, compared to a neutral model.

Lastly, McDonald–Kreitman tests found that 1000G amino acid SNPs are unexpectedly common in number (Table S14), especially since there is no amino acid divergence between humans and chimpanzees. Even though this result is typically explained by positive selection for amino acid polymorphism, this pattern may be explained by weak purifying selection on SNPs that remain at low frequencies [81,82]. Indeed, after individually omitting SNPs < 5% in frequency from all classes, there was only one amino acid SNP remaining (in Africans), and all McDonald–Kreitman tests were no longer significant.

3.2. Chimpanzee Amino Acid Variation

While we observed no *COL1A1* amino acid divergence between humans and chimpanzees, we also found no amino acid polymorphism within any of the four chimpanzee

subspecies in the *Pan* genomes dataset or in our generated sequence data—a total of 158 gene copies. However, we did observe a partial exon 35 duplication that appears to be the result of an unequal crossover event (Figure S3), with an allele frequency of 17.5% (including a homozygous individual), in our high-coverage dataset of 20 western chimpanzees. The exon 35 variant has 36 nucleotides, and the intron splice sites surrounding it are intact; thus, if encoded, it would result in an additional 12 amino acids to the COL1A1 protein in the triple-helix domain.

We further screened additional datasets to determine whether the exon 35 variant was isolated to western chimpanzees. First, the exon 35 variant was not found in the 26 bonobo *COL1A1* copies that we generated. Second, we obtained the raw sequence files for the *Pan* genomes dataset through the European Nucleotide Archive (ENA) Project PRJEB15086 and, through BLAST analysis, identified the exon 35 variant in one western chimpanzee, but not in the other three subspecies. The rare observation of the exon 35 variant in the *Pan* genomes dataset of western chimpanzees is likely the result of the much lower coverage we observed at *COL1A1* compared to the published genome-wide average [63]. This low coverage for *COL1A1* is not surprising, given its highly repetitive exons that are difficult to sequence.

Given the absence of COL1A1 protein coding-length polymorphisms in humans—at least unlinked to severe disease—the observation of one at high frequency in chimpanzees could reveal interesting information about the origin of COL1A1 functional variation. Using PCR analyses, we identified the exon 35 variant in two of the six pluripotent stem cell (iPSC) lines derived from western chimpanzees (Supplementary File S1) [83]. However, our follow-up qPCR analyses of RNA extracted from the two iPSC lines suggest that the exon 35 variant is not incorporated into the *COL1A1* RNA transcript, nor does it appear that it affects *COL1A1* gene expression levels.

3.3. Population Patterns of Human Intron Variation

Overall intron nucleotide diversity at *COL1A1* (Table S1) is consistent with the human genome average of ~0.1%, as well as with diversity in Africans being significantly higher than in non-Africans as a result of the aforementioned contrasts in their demographic histories. Previously we noted the overall skew of rare frequencies for amino acid SNPs compared to intron SNPs (Table S8). However, even intron SNP frequencies appeared to be highly skewed, with 87% of them found below 1% in frequency in the global dataset. The majority of this global pattern (88% of SNPs below 1%) is explained by a recent expansion in non-Africans [57]; however, the fact that Africans show very rare frequencies (64% below 1%) suggests that factors other than demography also shape *COL1A1* intron variation overall.

In looking at analyses of phyloP score distributions of introns (Tables S10 and S11), 1000G intron SNPs occur at sites with significantly more evolutionary acceleration (-0.36) compared to the distribution observed for all *COL1A1* intron sites (-0.05). However, when we examined the spatial distribution of intron conservation, we found significant heterogeneity (Figure 2a). The first 21 introns that cover a gene distance of ~6700 bp have significantly higher conservation than the latter 29 introns that cover ~8900 bp (randomization test comparing means of introns, while accounting for their lengths, $p = 0.00085$). This pattern was significant even when omitting the first intron ($p = 0.00057$) which, from a genome-wide perspective, is typically longer and more conserved [84].
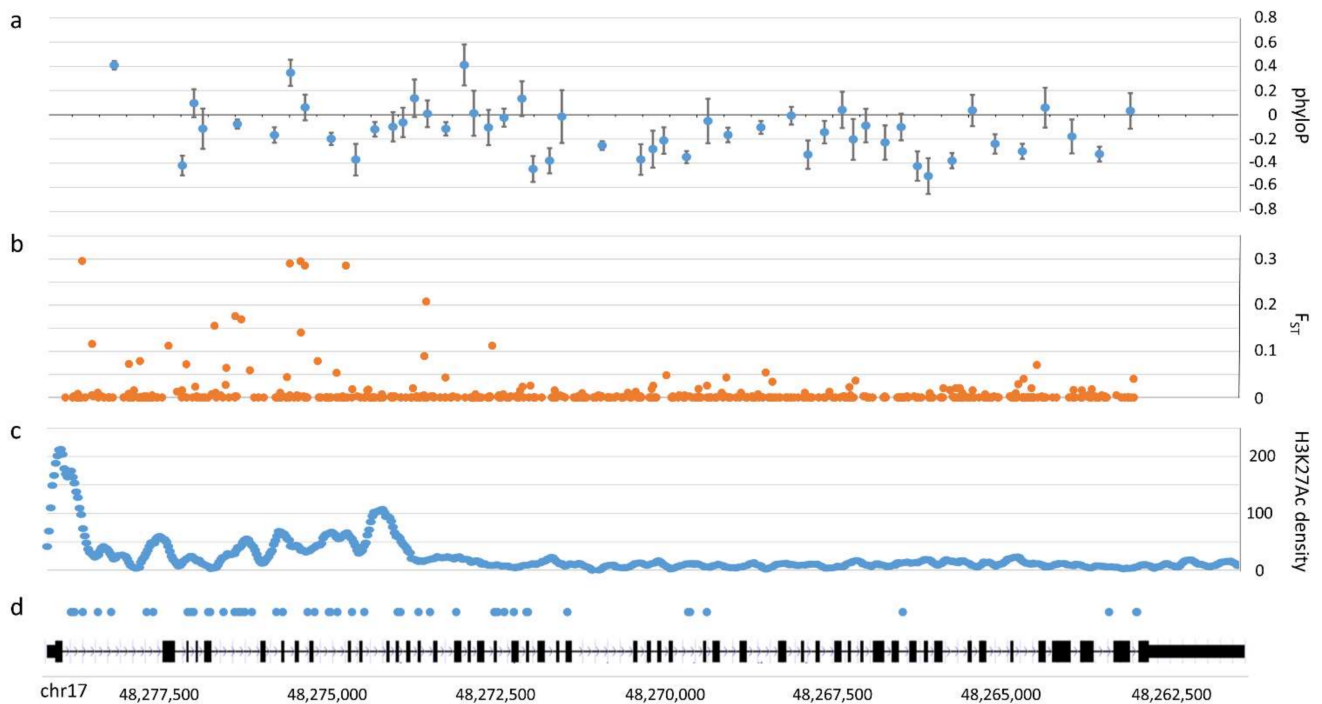
Figure 2. *COL1A1* patterns of variation in humans and chimpanzees: The ~17-kb *COL1A1* locus on chr17 (hg19 assembly) is shown with coding regions as black boxes interspersed with introns. (**a**) phyloP vertebrate track conservation scores for 50 introns (mean and SE plotted at their midpoints). As scores deviate from "0", the positive and negative scores reflect evolutionary conservation and acceleration, respectively, at nucleotide sites compared to a neutral model. (**b**) Plot of intronic SNP pairwise F_{ST} values between human populations of African and East Asian ancestry. (**c**) H3K27Ac marks, associated with regulatory elements, with plot reflecting the density calculated as the number of sequenced H3K27Ac tags overlapping a 25 bp window centered at that position. (**d**) Plot of the 45 intron SNPs fixed between the two core haplogroups in 40 western chimpanzee sequences.

As is typical with analyses of humans of African descent [85,86], we found significant genetic differentiation for *COL1A1* non-coding SNPs when compared to non-African groups. While the vast majority of SNPs exhibit typical patterns of differentiation among global populations, with $F_{ST} < 15\%$, there are a few SNPs between Africans and non-Africans that are significant outliers, with $F_{ST} > 30\%$ (Figures S4 and S5; Table S15). As noted previously, there are very few intron SNPs of high frequency, and they are evenly distributed across the *COL1A1* gene (Figure S6). However, SNPs that show significant population structure are located only within introns 1–15 (Figure 2b).

We found that all *COL1A1* SNPs of high population differentiation (Table S15) are included within a 5′ LD block, which spans the promoter region 2 Kb upstream through intron 15 (Figure 3), and is statistically independent from a 3′ LD block that includes the rest of the gene. We note that these patterns are similar in population-specific analyses (Figure S7), although correlations tend to be stronger and patterns obscured in some non-African populations (Figure S7g–r)—again, as a result of a more recent evolutionary history and less recombination associated with these ancestries [87]. While several intron SNPs are correlated within populations and form smaller blocks, it is clear that correlations among them alone cannot explain the significant population structure. For example, several SNPs that exhibit significant F_{ST} patterns reach high derived allele frequencies in Africans, or in non-Africans, but are very rare otherwise (Tables S15 and S16). Finally, in looking outside of *COL1A1* (Figure S8), we note that correlations with the 3′ LD block, although weak, continue in the downstream region no further than ~25 Kb, whereas correlations are completely absent between SNPs upstream and SNPs in introns 1–15 that show high F_{ST}.
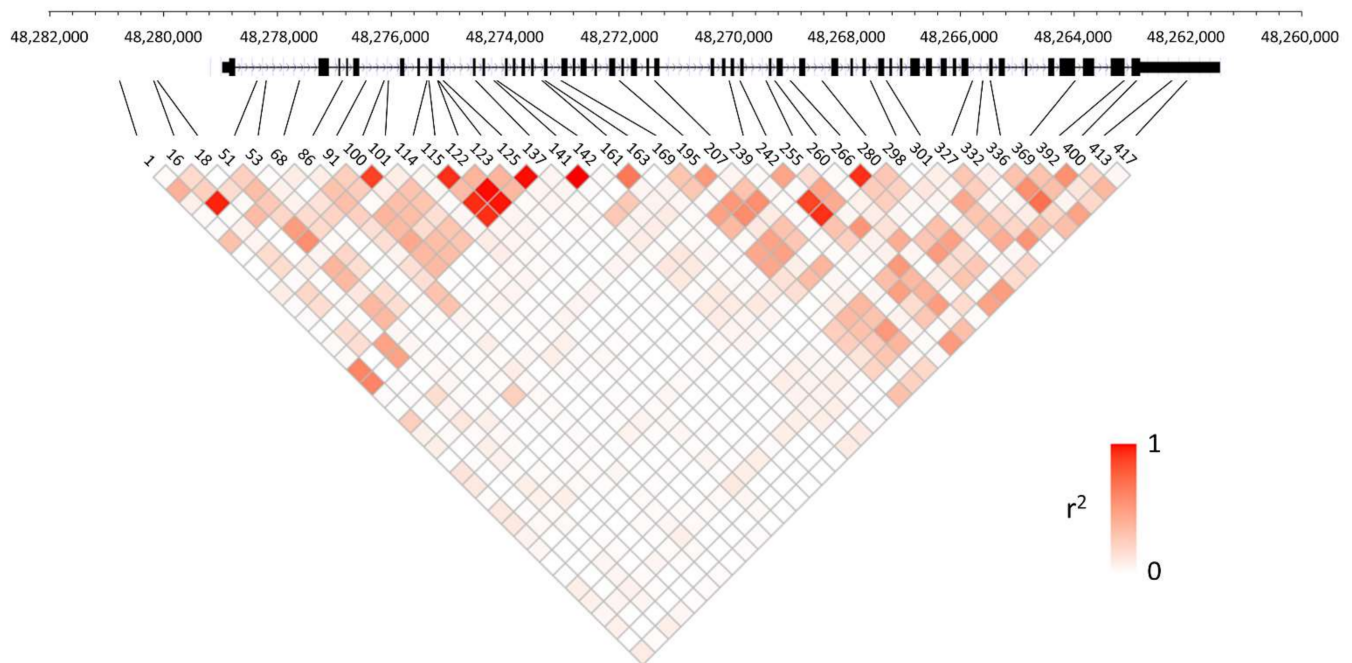
Figure 3. Global pattern of linkage disequilibrium in human *COL1A1*: Linkage disequilibrium pairwise plot of 39 non-coding SNPs (MAF > 5%) in the phased 1000 Genomes Project dataset is plotted as a function of nucleotide distance (see Table S16 for SNP numbering). The ~17 kb *COL1A1* locus on chr17 (hg19 assembly) is shown with coding regions as black boxes interspersed with introns. Shaded boxes in red reflect the strength of pairwise correlations among SNPs.

3.4. Population Patterns of Chimpanzee Intron Variation

Intron nucleotide diversity in *COL1A1* in our sample of 20 western chimpanzees (Table S4) was similar to that of humans, as well as to previously published autosomal loci in these and other samples [66–70]. The phased *Pan* genomes *COL1A1* haplotypes (Table S3) revealed less than one-third of the SNPs in our western chimpanzee dataset (Table S17). This result cannot be explained by the SNPs being rare in frequency (see below), and there is little evidence for population structure in western chimpanzees from the published *Pan* genomes analysis [63]. Instead, this result reflects the significantly reduced sequence coverage of *COL1A1* in the *Pan* genomes dataset. Thus, we used our western chimpanzee sample for standard population genetic analyses, while the *Pan* genomes dataset served for analyses such as LD, where estimates are not biased by missing SNPs.

Unlike the skew in rare allele frequencies for humans, our phased western chimpanzee data revealed two intermediate-frequency haplotypes with strong LD across the *COL1A1* locus (Table S17). In fact, of the 63 intron SNPs in the phased sample, 45 are fixed between the two core haplogroups. In contrast, in our most polymorphic African human sample, with >100 SNPs, only 12 intron SNPs reach an MAF > 30%. Interestingly, 38 of the 45 fixed differences are found in introns 1–21 (Figure 2d)—a result that was significantly unexpected ($p < 10^{-7}$) based on a permutation analysis that simulated random distributions of 45 mutations across the >11 Kb of 50 introns. This result shows that the high proportion of fixed differences between haplogroups occurring in the first 21 introns is highly unlikely to be explained by chance alone.

Although far fewer SNPs exist in the *Pan* genomes dataset of *COL1A1*, we can see that the unusual haplotype structure in western chimpanzees is also similar across the other three subspecies (Table S3), suggesting a relatively old age that predates chimpanzee subspecies divergence. To test this hypothesis, we employed the model of Thomson et al. [88], wherein assumptions of population equilibrium and recombination are relaxed. Owing to the unusually high LD, even across subspecies, very few recombinants are

obvious, enabling more accurate age estimates under such a coalescent approach [89,90]. The age estimate (*t*) involves the relationship in (1):

$$t = \sum_{i=1}^{n} \frac{x_i}{(n\mu)} \tag{1}$$

where x_i is the number of mutational differences between the *i*th sequence and the most recent common ancestor (MRCA) of all sequences, *n* is the number of sequences, and μ is the mutation rate. Here, μ is estimated as the number of substitutions between human and chimpanzee divided by the estimated molecular divergence time between species (5 ± 1 My) [26], multiplied by two. Alignments with bonobo and human data enabled estimates of x_i as the number of differences accumulated on each haplotype since the MRCA. Our estimate of the divergence time between the two haplotypes was 2.8 ± 0.6 My, which predates not only chimpanzee subspecies divergence times, but also the chimpanzee–bonobo divergence time [63].

Inspection of the bonobo data from the *Pan* genomes dataset shows that bonobos are not fixed for either of the two chimpanzee core haplotypes, but are a mix of the two (Table S17). This pattern is consistent with the two haplogroups existing in the ancestral population prior to the divergence of chimpanzee and bonobo, and with our estimated date above. One explanation for the relatively old age, yet such high LD, could be historical introgression of haplotypes from bonobos, as has been found for some chimpanzee subspecies, but very rarely in western chimpanzees. To test this hypothesis, we generated neighbor-joining (NJ) phylogenetic trees using MEGA-X [91] for (1) *COL1A1* in our 40 phased western chimpanzee sequences, with bonobo as an outgroup (Table S17); (2) sequenced regions 100 Kb upstream and downstream in our 40 phased western chimpanzee (Table S5) and 26 phased bonobo sequences (Table S6), using human as an outgroup; and (3) *COL1A1* in the *Pan* genomes' 118 phased subspecies sequences, using bonobo as an outgroup (Table S3). The western chimpanzee *COL1A1* NJ tree (Figure S9) reflects the two haplogroups seen in our high-coverage data, with divergence (albeit little) between chimpanzees and bonobos. This monophyletic pattern is even more apparent in the NJ tree that includes the upstream and downstream regions (Figure S10). Our final NJ tree analysis shows *COL1A1* haplotypes shared across the four subspecies (Figure 4), but as a monophyletic group separate from bonobos. Thus, we find no evidence of introgression of bonobo alleles into chimpanzees at *COL1A1* or areas in close proximity that can explain the unique *COL1A1* haplotype structure.

Finally, we used long-range LD and estimates of F_{ST} from the *Pan* genomes dataset to test the hypothesis that the haplotype structure of *COL1A1* is the result of linkage with other loci, or even potentially the result of hybridization among chimpanzees. Calculations of LD between a core set of SNPs at *COL1A1* and other SNPs in each of the four subspecies samples show a striking pattern of localized LD at *COL1A1* that variably decays, moving 100 Kb each upstream and downstream (Figure S11). This decay is expected, as recombination reduces correlations as a function of time and population size and, as predicted, is most abrupt in central chimpanzees, but with the least decay in Nigeria–Cameroon samples [63,92]. This same pattern can be visualized via the F_{ST} analyses that, despite the high sharing of haplotype diversity in *COL1A1* among subspecies, show high levels of differentiation between subspecies over a 1 Mb region outside of *COL1A1* (Figure S12). These results demonstrate that the old chimpanzee haplotype structure localized to the *COL1A1* locus cannot be explained simply by demographic or chromosomal factors.
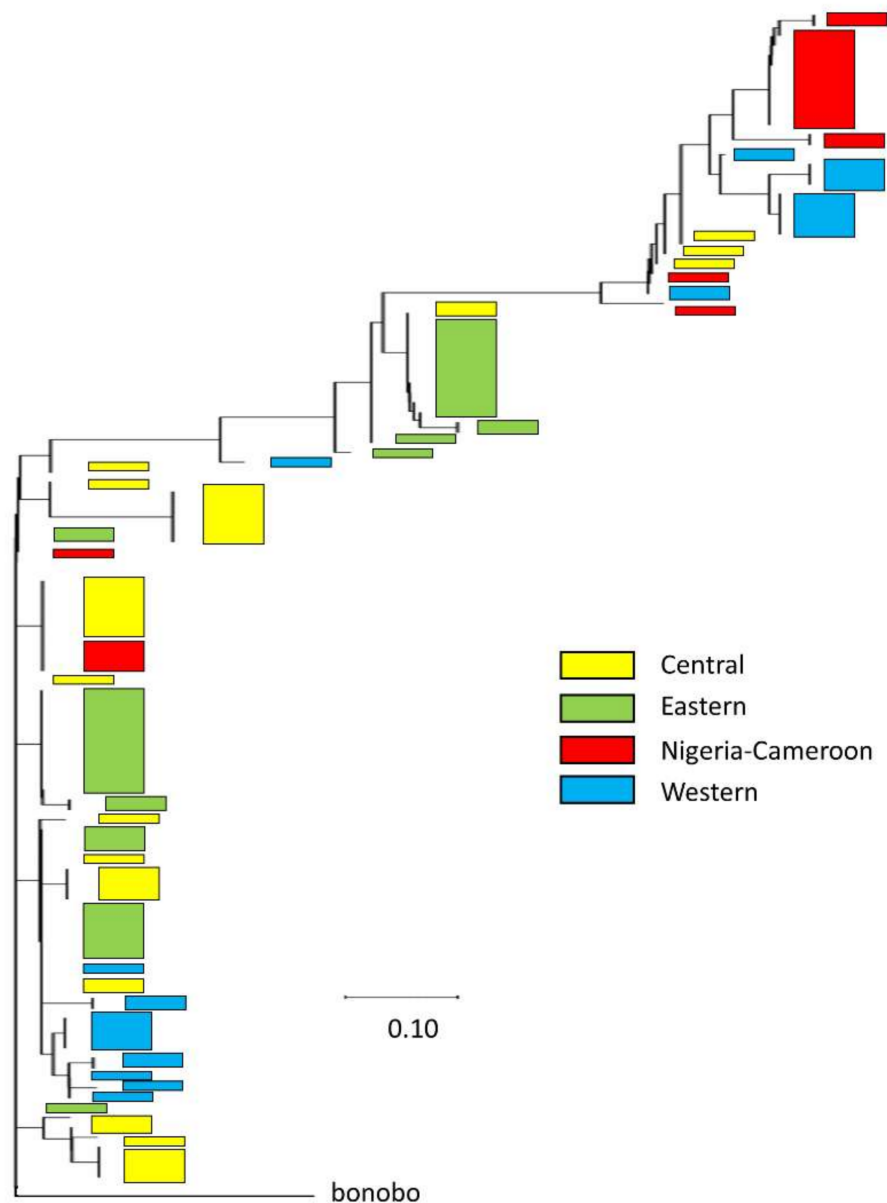
Figure 4. Neighbor-joining tree of 118 chimpanzee *COL1A1* haplotypes: Haplotypes and their relative phylogenetic positions for each of the four subspecies are shown as color blocks, with bonobo as an outgroup. Evolutionary distances reflect the number of base substitutions per variable site.

4. Discussion

The present study takes advantage of phylogenetic conservation analyses, whole-genome databases, and molecular functional tools that have emerged in the 10 years since our initial deep-time evolutionary analyses of *COL1A1* suggested that the general population likely harbors significantly more variation than clinical studies reveal [37]. We found contrasting patterns of *COL1A1* amino acid variation between clinical and natural human populations, and multiple evolutionary signatures of potentially adaptive functional variation associated with introns in humans and chimpanzees. Here, we discuss the implications that these observations may have for the selective pressures shaping bone-related phenotypes, and consequences for disease prevalence.

4.1. COL1A1 Protein Variation Is Higher Than Expected

COL1A1 protein variation in the natural population appears more common than would be expected if it were simply linked to severe disease. From a broader perspective, this pattern is unusual even in comparison to genome-wide patterns for "disease genes" [90,93–95]. One consideration is that studies have shown that sequencing anomalies in the 1000G data impact estimates of rare alleles [96,97]; however, there are several reasons why these anomalies alone cannot explain our results. First, we would expect all populations and sites to be similarly impacted, yet we see that patterns differ between Africans and non-Africans, and between coding and non-coding variants. Second, the patterns of human population differentiation for intronic variation discussed below all involve common variants. Lastly, as we presented high-coverage data in a resequencing of chimpanzees and bonobos, we also conducted a similar resequencing effort of *COL1A1* in a global sample of humans (data unpublished, Table S18), finding that 9% of individuals carry at least one amino acid variant. This result is consistent with the 1000G data—specifically, the global human sample predicts that ~5% of individuals, and as many as 9% of individuals with African ancestry, carry an amino acid variant in *COL1A1*.

Compared to DAMs, patterns for natural populations are consistent with the hypothesis that *COL1A1* variants do not always reflect disease [80] but, rather, they represent a different category of COL1A1 protein variation altogether. In regions of the protein such as the triple helix, where change is expected to be deleterious, they fall almost completely at sites different from DAMs, and have significantly lower conservation scores, even compared to the least severe category of osteoporotic-like disease (Figure 1). In other words, amino acid polymorphism at *COL1A1* appears to have varying functional effects insufficient to result in severe disease and subsequent detection in clinical studies, and may thus contribute to natural phenotypic variation in type I collagen that is not deleterious with respect to fitness. In this respect, patterns of polymorphism and divergence in *COL1A1* are actually consistent with the same evolutionary process. That is, certain mutations are rapidly removed by purifying selection (i.e., fatal mutations impairing the triple-helix domain [32,33]), resulting in little fixation over deep time, while others with subtle to no effect on type I collagen can accumulate as polymorphisms in the general population. As such, purifying selection at *COL1A1* would not be considered "weak", as has been suggested as an explanation for similar patterns of genome-wide amino acid variation [93–95]; rather, selection is of varying strength across the protein sequence over time. This scenario was initially proposed by us and others from deep-time evolutionary analyses [37,47]. Specifically, variation in constraint along the COL1A1 triple-helix domain implies more flexibility in this region over evolutionary time, which makes sense, as this region is responsible for structural and mechanical variation in bone, including mineral content and organization of collagen fibers that vary among vertebrates [98].

These combined deep-time and population-based evolutionary analyses make predictions about where mutations of variable functional impact are likely to occur and contribute to population variation related to bone morphology. For example, only 5% of DAMs are found in the N-domain, with almost all of them related to osteoporotic-like diseases; in contrast, when looking at the natural population, we find that 13% are found in the N-domain. We see a similar pattern for the C-domain, with 15% and 28% found in DAMs and the natural population, respectively. While the N-domain helps keep the protein soluble until translation and processing are complete, the C-domain is responsible for the recognition and assembly of type I collagen subunits, and where mutations can be severe, as they prevent formation of the triple-helix [44,45]. This result may be expected for the C-domain which, as previously mentioned, shows the strongest signature of deep-time evolutionary constraint in our previous analyses [37], supporting the patterns seen here at the population level. Finally, we also identified mutations—even shared across populations—in specific regions of the triple-helix domain that are predicted to be "lethal" because they are related to major ligand binding [33]. Thus, while *COL1A1* mutations have been linked to disease severity, natural populations represent complex interactions between the environment and

natural selection that have resulted in some mutations being common. While these patterns cannot explain the overall variation in phenotypic and functional differences, such as BMD, across populations [22], they do represent variation that can be evolutionarily vital to bone-related phenotypes over time. That is, factors such as amino acid size and thermostability are important to bone-related disease prediction models [33,41], whereas subtle variation that does not greatly impact phenotypes has likely acted as adaptive potential for selection in historically shaping type I collagen [42,43].

A great example of this adaptive potential comes from our analyses of chimpanzees. The absence of observed amino acid polymorphism in all four subspecies of chimpanzees, and in bonobos, may reflect different environmental constraints between our species, such as in locomotion, diet, and skeletal growth periods [99,100]. However, the common variant resulting in a partial duplication of exon 35 was an unusual find compared to humans, where mutations that affect protein length—and particularly of the triple-helix domain—are exceedingly rare and highly deleterious [39–41]. Unsurprisingly, we found no evidence that the variant is currently encoded, as it would likely require a convergent change in COL1A2 to accurately form type I collagen. However, it should be noted that the current family of collagens all likely evolved from an ancestral collagen with a single 54 bp exon [42]; thus, mutation in length variation must have been available historically across vertebrate evolution. The nature of the exon 35 variant with a truncation of the exon that would code for 12 amino acids—with all splice sites intact, and at high frequency—represents an example of intriguing standing genetic variation of potential functional value for positive selection.

4.2. Signatures of Adaptive COL1A1 Intronic Variation within Humans and Chimpanzees

The patterns of deep-time evolutionary constraint, human population differentiation, and unusual chimpanzee haplotype structure all coincidentally isolated to the same intronic region of *COL1A1* cannot be explained by neutral forces such as shared demographic history, population structure, or low recombination. One explanation is that balancing selection—which favors intronic diversity within species, but reduces divergence over time [101,102]—maintains variation in *COL1A1* expression. In looking at ChIP-seq data for the *COL1A1* locus [103], enrichment of H3K27Ac—well recognized as a marker of enhancer activity and gene expression—is significantly over-represented not only in the first intron, but at least through intron 16 (Figure 2c), which coincides with the multiple signatures of selection seen here. The first intron includes numerous transcription-factor-binding sites [34,35,50], and frequencies of the Sp1 allele found here at 20% in Europeans and 9% in Africans confirm that it has long been segregating in the human population. However, the Sp1 allele is not one of the intron variants identified here that show patterns of significant population differentiation. Interestingly, these intronic patterns all include populations with sub-Saharan African ancestry, with intronic SNPs of intermediate derived allele frequencies that are virtually absent elsewhere. A previous study suggested that positive selection explains the over-representation of alleles that increase BMD in sub-Saharan African populations [22]. Thus, we might conclude that the signatures of positive selection associated with *COL1A1* intronic alleles over-represented in our sub-Saharan African sample also reflect increased BMD as a result of variation in *COL1A1* gene expression. In this respect, bone strength joins a list of phenotypes—such as malarial resistance [69,104], color vision [70,105], lactase persistence [106], and lipid and glycemic traits [107]—that have evolved from positive selection for adaptive immunity and subsistence in ancestral sub-Saharan Africans, yet are often linked to disease elsewhere [58,59,90,94]. These trait differences reflect variable selective pressures over time across ethnic groups and species, and merit caution in drawing conclusions about how phenotypic and genetic variation in one environment may have similar deleterious or advantageous effects in another.

5. Conclusions and Future Directions

Identifying the relationship between genotype and phenotype is a challenge for which evolutionary genetic studies provide a unique perspective, with additional insight into the adaptive potential of this variation. In this respect, disease phenotypes such as those associated with COL1A1 provide an excellent starting point, as we have intimate information about genotype–phenotype relationships that enables us to interpret natural population variation. Our results here are consistent with evolutionary theory in general, in that the *COL1A1* variation most likely to have adaptive potential would be subtle in nature with respect to phenotype [108], and would most likely go unnoticed in clinical screenings.

We can speculate as to whether the signatures of adaptive collagen-related variation here are the result of historical selection for bone strength related to locomotion, development, and wound repair [99,100]; however, a future step would be first to determine the extent to which these variants represent functional variation. Examples include targeted *COL1A1* SNPs in human- and chimpanzee-derived cell lines with RNA-seq, ChIP-seq, and reporter assay technologies to identify the effects of gene expression [109], as well as investigations of animal models and case–control cohorts that exhibit different BMD profiles [16,110]. As comparative primate functional genomics continues to evolve [111], we predict that using our study of *COL1A1* as a model for other gene-based evolutionary analyses will reveal cryptic variation underlying "disease genes" with potential functional and adaptive significance.

Author Contributions: Conceptualization, D.A.S. and B.C.V.; methodology, D.A.S., G.H., M.S.R. and B.C.V.; data curation, D.A.S., G.H., M.S.R. and B.C.V.; formal analysis, D.A.S., G.H., M.S.R. and B.C.V.; interpretation, D.A.S., G.H., A.C.S., M.S.R. and B.C.V.; writing—original draft preparation, D.A.S. and B.C.V.; writing—review and editing, D.A.S., G.H., A.C.S., M.S.R. and B.C.V.; supervision, project

References

1. Stoll, C.; Dott, B.; Roth, M.P.; Alembik, Y. Birth Prevalence Rates of Skeletal Dysplasias. *Clin. Genet.* **1989**, *35*, 88–92. [CrossRef]
2. Reginster, J.-Y.; Burlet, N. Osteoporosis: A Still Increasing Prevalence. *Bone* **2006**, *38*, 4–9. [CrossRef] [PubMed]
3. Ballane, G.; Cauley, J.A.; Luckey, M.M.; El-Hajj Fuleihan, G. Worldwide Prevalence and Incidence of Osteoporotic Vertebral Fractures. *Osteoporos. Int.* **2017**, *28*, 1531–1542. [CrossRef]
4. Looker, A.C.; Sarafrazi Isfahani, N.; Fan, B.; Shepherd, J.A. Trends in Osteoporosis and Low Bone Mass in Older US Adults, 2005–2006 through 2013–2014. *Osteoporos. Int.* **2017**, *28*, 1979–1988. [CrossRef] [PubMed]
5. Noel, S.E.; Santos, M.P.; Wright, N.C. Racial and Ethnic Disparities in Bone Health and Outcomes in the United States. *J. Bone. Min. Res.* **2021**, *36*, 1881–1905. [CrossRef] [PubMed]
6. Adami, G.; Cattani, G.; Rossini, M.; Viapiana, O.; Olivi, P.; Orsolini, G.; Bertoldo, E.; Fracassi, E.; Gatti, D.; Fassio, A. Association between Exposure to Fine Particulate Matter and Osteoporosis: A Population-Based Cohort Study. *Osteoporos. Int.* **2022**, *33*, 169–176. [CrossRef]
7. Min, C.; Yoo, D.M.; Wee, J.H.; Lee, H.-J.; Choi, H.G. High-Intensity Physical Activity with High Serum Vitamin D Levels Is Associated with a Low Prevalence of Osteopenia and Osteoporosis: A Population-Based Study. *Osteoporos. Int.* **2021**, *32*, 883–891. [CrossRef] [PubMed]
8. Brown, L.B.; Streeten, E.A.; Shapiro, J.R.; McBride, D.; Shuldiner, A.R.; Peyser, P.A.; Mitchell, B.D. Genetic and Environmental Influences on Bone Mineral Density in Pre- and Post-Menopausal Women. *Osteoporos. Int.* **2005**, *16*, 1849–1856. [CrossRef]
9. Videman, T.; Levälahti, E.; Battié, M.C.; Simonen, R.; Vanninen, E.; Kaprio, J. Heritability of BMD of Femoral Neck and Lumbar Spine: A Multivariate Twin Study of Finnish Men. *J. Bone Miner. Res.* **2007**, *22*, 1455–1462. [CrossRef]
10. Liu, C.-T.; Karasik, D.; Zhou, Y.; Hsu, Y.-H.; Genant, H.K.; Broe, K.E.; Lang, T.F.; Samelson, E.J.; Demissie, S.; Bouxsein, M.L.; et al. Heritability of Prevalent Vertebral Fracture and Volumetric Bone Mineral Density and Geometry at the Lumbar Spine in Three Generations of the Framingham Study. *J. Bone Miner. Res.* **2012**, *27*, 954–958. [CrossRef]
11. Estrada, K.; Styrkarsdottir, U.; Evangelou, E.; Hsu, Y.-H.; Duncan, E.L.; Ntzani, E.E.; Oei, L.; Albagha, O.M.E.; Amin, N.; Kemp, J.P.; et al. Genome-Wide Meta-Analysis Identifies 56 Bone Mineral Density Loci and Reveals 14 Loci Associated with Risk of Fracture. *Nat. Genet.* **2012**, *44*, 491–501. [CrossRef]
12. Wang, W.; Huang, S.; Hou, W.; Liu, Y.; Fan, Q.; He, A.; Wen, Y.; Hao, J.; Guo, X.; Zhang, F. Integrative Analysis of GWAS, EQTLs and MeQTLs Data Suggests That Multiple Gene Sets Are Associated with Bone Mineral Density. *Bone Joint Res.* **2017**, *6*, 572–576. [CrossRef]
13. Kim, S.K. Identification of 613 New Loci Associated with Heel Bone Mineral Density and a Polygenic Risk Score for Bone Mineral Density, Osteoporosis and Fracture. *PLoS ONE* **2018**, *13*, e0200785. [CrossRef] [PubMed]
14. Liu, A.; Liu, Y.; Su, K.-J.; Greenbaum, J.; Bai, Y.; Tian, Q.; Zhao, L.-J.; Deng, H.-W.; Shen, H. A Transcriptome-Wide Association Study to Detect Novel Genes for Volumetric Bone Mineral Density. *Bone* **2021**, *153*, 116106. [CrossRef]
15. Karasik, D.; Rivadeneira, F.; Johnson, M.L. The Genetics of Bone Mass and Susceptibility to Bone Diseases. *Nat. Rev. Rheumatol.* **2016**, *12*, 323–334. [CrossRef]
16. Formosa, M.M.; Bergen, D.J.M.; Gregson, C.L.; Maurizi, A.; Kämpe, A.; Garcia-Giralt, N.; Zhou, W.; Grinberg, D.; Ovejero Crespo, D.; Zillikens, M.C.; et al. A Roadmap to Gene Discoveries and Novel Therapies in Monogenic Low and High Bone Mass Disorders. *Front. Endocrinol.* **2021**, *12*, 709711. [CrossRef]
17. Nowlan, N.C.; Jepsen, K.J.; Morgan, E.F. Smaller, Weaker, and Less Stiff Bones Evolve from Changes in Subsistence Strategy. *Osteoporos. Int.* **2011**, *22*, 1967–1980. [CrossRef] [PubMed]

18. Kralick, A.E.; Zemel, B.S. Evolutionary Perspectives on the Developing Skeleton and Implications for Lifelong Health. *Front. Endocrinol.* **2020**, *11*, 99. [CrossRef] [PubMed]

19. Liu, C.-T.; Estrada, K.; Yerges-Armstrong, L.M.; Amin, N.; Evangelou, E.; Li, G.; Minster, R.L.; Carless, M.A.; Kammerer, C.M.; Oei, L.; et al. Assessment of Gene-by-Sex Interaction Effect on Bone Mineral Density. *J. Bone Miner. Res.* **2012**, *27*, 2051–2064. [CrossRef]

20. Yau, M.S.; Kuipers, A.L.; Price, R.; Nicolas, A.; Tajuddin, S.M.; Handelman, S.K.; Arbeeva, L.; Chesi, A.; Hsu, Y.-H.; Liu, C.-T.; et al. A Meta-Analysis of the Transferability of Bone Mineral Density Genetic Loci Associations from European to African Ancestry Populations. *J. Bone Miner. Res.* **2021**, *36*, 469–479. [CrossRef]

21. Agarwal, S.C. What Is Normal Bone Health? A Bioarchaeological Perspective on Meaningful Measures and Interpretations of Bone Strength, Loss, and Aging. *Am. J. Hum. Biol.* **2021**, *33*, e23647. [CrossRef] [PubMed]

22. Medina-Gómez, C.; Chesi, A.; Heppe, D.H.M.; Zemel, B.S.; Yin, J.-L.; Kalkwarf, H.J.; Hofman, A.; Lappe, J.M.; Kelly, A.; Kayser, M.; et al. BMD Loci Contribute to Ethnic and Developmental Differences in Skeletal Fragility across Populations: Assessment of Evolutionary Selection Pressures. *Mol. Biol. Evol.* **2015**, *32*, 2961–2972. [CrossRef]

23. Black, A.; Tilmont, E.M.; Handy, A.M.; Scott, W.W.; Shapses, S.A.; Ingram, D.K.; Roth, G.S.; Lane, M.A. A Nonhuman Primate Model of Age-Related Bone Loss: A Longitudinal Study in Male and Premenopausal Female Rhesus Monkeys. *Bone* **2001**, *28*, 295–302. [CrossRef]

24. Lipkin, E.W.; Aumann, C.A.; Newell-Morris, L.L. Evidence for Common Controls over Inheritance of Bone Quantity and Body Size from Segregation Analysis in a Pedigreed Colony of Nonhuman Primates (*Macaca Nemestrina*). *Bone* **2001**, *29*, 249–257. [CrossRef]

25. Havill, L.M.; Allen, M.R.; Harris, J.a.K.; Levine, S.M.; Coan, H.B.; Mahaney, M.C.; Nicolella, D.P. Intracortical Bone Remodeling Variation Shows Strong Genetic Effects. *Calcif. Tissue Int.* **2013**, *93*, 472–480. [CrossRef]

26. Kumar, S.; Filipski, A.; Swarna, V.; Walker, A.; Hedges, S.B. Placing Confidence Limits on the Molecular Age of the Human-Chimpanzee Divergence. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 18842–18847. [CrossRef]

27. Kikuchi, Y.; Udono, T.; Hamada, Y. Bone Mineral Density in Chimpanzees, Humans, and Japanese Macaques. *Primates* **2003**, *44*, 151–155. [CrossRef]

28. Matsumura, A.; Gunji, H.; Takahashi, Y.; Nishida, T.; Okada, M. Cross-Sectional Morphology of the Femoral Neck of Wild Chimpanzees. *Int. J. Primatol.* **2010**, *31*, 219–238. [CrossRef]

29. Mulhern, D.M.; Ubelaker, D.H. Bone Microstructure in Juvenile Chimpanzees. *Am. J. Phys. Anthropol.* **2009**, *140*, 368–375. [CrossRef]

30. Zhu, X.; Bai, W.; Zheng, H. Twelve Years of GWAS Discoveries for Osteoporosis and Related Traits: Advances, Challenges and Applications. *Bone Res.* **2021**, *9*, 23. [CrossRef]

31. Viguet-Carrin, S.; Garnero, P.; Delmas, P.D. The Role of Collagen in Bone Strength. *Osteoporos. Int.* **2006**, *17*, 319–336. [CrossRef] [PubMed]

32. Dalgleish, R. The Human Type I Collagen Mutation Database. *Nucleic Acids Res.* **1997**, *25*, 181–187. [CrossRef]

33. Marini, J.C.; Forlino, A.; Cabral, W.A.; Barnes, A.M.; San Antonio, J.D.; Milgrom, S.; Hyland, J.C.; Körkkö, J.; Prockop, D.J.; de Paepe, A.; et al. Consortium for Osteogenesis Imperfecta Mutations in the Helical Domain of Type I Collagen: Regions Rich in Lethal Mutations Align with Collagen Binding Sites for Integrins and Proteoglycans. *Hum. Mutat.* **2007**, *28*, 209–221. [CrossRef]

34. Garcia-Giralt, N.; Nogués, X.; Enjuanes, A.; Puig, J.; Mellibovsky, L.; Bay-Jensen, A.; Carreras, R.; Balcells, S.; Díez-Pérez, A.; Grinberg, D. Two New Single-Nucleotide Polymorphisms in the COL1A1 Upstream Regulatory Region and Their Relationship to Bone Mineral Density. *J. Bone Miner. Res.* **2002**, *17*, 384–393. [CrossRef]

35. Stewart, T.L.; Jin, H.; McGuigan, F.E.A.; Albagha, O.M.E.; Garcia-Giralt, N.; Bassiti, A.; Grinberg, D.; Balcells, S.; Reid, D.M.; Ralston, S.H. Haplotypes Defined by Promoter and Intron 1 Polymorphisms of the COLIA1 Gene Regulate Bone Mineral Density in Women. *J. Clin. Endocrinol. Metab.* **2006**, *91*, 3575–3583. [CrossRef]

36. Jiang, H.; Lei, S.-F.; Xiao, S.-M.; Chen, Y.; Sun, X.; Yang, F.; Li, L.-M.; Wu, S.; Deng, H.-W. Association and Linkage Analysis of COL1A1 and AHSG Gene Polymorphisms with Femoral Neck Bone Geometric Parameters in Both Caucasian and Chinese Nuclear Families. *Acta Pharmacol. Sin.* **2007**, *28*, 375–381. [CrossRef]

37. Stover, D.A.; Verrelli, B.C. Comparative Vertebrate Evolutionary Analyses of Type I Collagen: Potential of COL1a1 Gene Structure and Intron Variation for Common Bone-Related Diseases. *Mol. Biol. Evol.* **2011**, *28*, 533–542. [CrossRef]

38. Fokkema, I.F.A.C.; Taschner, P.E.M.; Schaafsma, G.C.P.; Celli, J.; Laros, J.F.J.; den Dunnen, J.T. LOVD v.2.0: The next Generation in Gene Variant Databases. *Hum. Mutat.* **2011**, *32*, 557–563. [CrossRef] [PubMed]

39. Pace, J.M.; Atkinson, M.; Willing, M.C.; Wallis, G.; Byers, P.H. Deletions and Duplications of Gly-Xaa-Yaa Triplet Repeats in the Triple Helical Domains of Type I Collagen Chains Disrupt Helix Formation and Result in Several Types of Osteogenesis Imperfecta. *Hum. Mutat.* **2001**, *18*, 319–326. [CrossRef]

40. Cabral, W.A.; Mertts, M.V.; Makareeva, E.; Colige, A.; Tekin, M.; Pandya, A.; Leikin, S.; Marini, J.C. Type I Collagen Triplet Duplication Mutation in Lethal Osteogenesis Imperfecta Shifts Register of α Chains throughout the Helix and Disrupts Incorporation of Mutant Helices into Fibrils and Extracellular Matrix. *J. Biol. Chem.* **2003**, *278*, 10006–10012. [CrossRef] [PubMed]

41. Bodian, D.L.; Chan, T.-F.; Poon, A.; Schwarze, U.; Yang, K.; Byers, P.H.; Kwok, P.-Y.; Klein, T.E. Mutation and Polymorphism Spectrum in Osteogenesis Imperfecta Type II: Implications for Genotype-Phenotype Relationships. *Hum. Mol. Genet.* **2009**, *18*, 463–471. [CrossRef] [PubMed]

42. Slatter, D.A.; Farndale, R.W. Structural Constraints on the Evolution of the Collagen Fibril: Convergence on a 1014-Residue COL Domain. *Open Biol.* **2015**, *5*, 140220. [CrossRef] [PubMed]

43. Kleinnijenhuis, A.J. Visualization of Genetic Drift Processes Using the Conserved Collagen 1α1 GXY Domain. *J. Mol. Evol.* **2019**, *87*, 106–130. [CrossRef]

44. Boot-Handford, R.P.; Tuckwell, D.S. Fibrillar Collagen: The Key to Vertebrate Evolution? A Tale of Molecular Incest. *Bioessays* **2003**, *25*, 142–151. [CrossRef] [PubMed]

45. Aouacheria, A.; Cluzel, C.; Lethias, C.; Gouy, M.; Garrone, R.; Exposito, J.-Y. Invertebrate Data Predict an Early Emergence of Vertebrate Fibrillar Collagen Clades and an Anti-Incest Model. *J. Biol. Chem.* **2004**, *279*, 47711–47719. [CrossRef] [PubMed]

46. Wada, H.; Okuyama, M.; Satoh, N.; Zhang, S. Molecular Evolution of Fibrillar Collagen in Chordates, with Implications for the Evolution of Vertebrate Skeletons and Chordate Phylogeny. *Evol. Dev.* **2006**, *8*, 370–377. [CrossRef]

47. Morgan, C.C.; Loughran, N.B.; Walsh, T.A.; Harrison, A.J.; O'Connell, M.J. Positive Selection Neighboring Functionally Essential Sites and Disease-Implicated Regions of Mammalian Reproductive Proteins. *BMC Evol. Biol.* **2010**, *10*, 39. [CrossRef] [PubMed]

48. Gelse, K.; Pöschl, E.; Aigner, T. Collagens—Structure, Function, and Biosynthesis. *Adv. Drug Deliv. Rev.* **2003**, *55*, 1531–1546. [CrossRef]

49. Hildebrand, K.A.; Gallant-Behm, C.L.; Kydd, A.S.; Hart, D.A. The Basics of Soft Tissue Healing and General Factors That Influence Such Healing. *Sports Med. Arthrosc. Rev.* **2005**, *13*, 136–144. [CrossRef]

50. Bornstein, P.; McKay, J.; Morishima, J.K.; Devarayalu, S.; Gelinas, R.E. Regulatory Elements in the First Intron Contribute to Transcriptional Control of the Human α 1(I) Collagen Gene. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 8869–8873. [CrossRef]

51. Mann, V.; Hobson, E.E.; Li, B.; Stewart, T.L.; Grant, S.F.; Robins, S.P.; Aspden, R.M.; Ralston, S.H. A COL1A1 Sp1 Binding Site Polymorphism Predisposes to Osteoporotic Fracture by Affecting Bone Density and Quality. *J. Clin. Investig.* **2001**, *107*, 899–907. [CrossRef] [PubMed]

52. Jin, H.; van't Hof, R.J.; Albagha, O.M.E.; Ralston, S.H. Promoter and Intron 1 Polymorphisms of COL1A1 Interact to Regulate Transcription and Susceptibility to Osteoporosis. *Hum. Mol. Genet.* **2009**, *18*, 2729–2738. [CrossRef]

53. Wang, C.; Li, H.; Chen, K.; Wu, B.; Liu, H. Association of Polymorphisms Rs1800012 in COL1A1 with Sports-Related Tendon and Ligament Injuries: A Meta-Analysis. *Oncotarget* **2017**, *8*, 27627–27634. [CrossRef]

54. Pollard, K.S.; Hubisz, M.J.; Rosenbloom, K.R.; Siepel, A. Detection of Nonneutral Substitution Rates on Mammalian Phylogenies. *Genome Res.* **2010**, *20*, 110–121. [CrossRef]

55. Chan, T.-F.; Poon, A.; Basu, A.; Addleman, N.R.; Chen, J.; Phong, A.; Byers, P.H.; Klein, T.E.; Kwok, P.-Y. Natural Variation in Four Human Collagen Genes across an Ethnically Diverse Population. *Genomics* **2008**, *91*, 307–314. [CrossRef]

56. Byrska-Bishop, M.; Evani, U.S.; Zhao, X.; Basile, A.O.; Abel, H.J.; Regier, A.A.; Corvelo, A.; Clarke, W.E.; Musunuri, R.; Nagulapalli, K.; et al. High coverage whole genome sequencing of the expanded 1000 genomes project cohort including 602 trios. *bioRxiv* **2021**. [CrossRef]

57. Keinan, A.; Clark, A.G. Recent Explosive Human Population Growth Has Resulted in an Excess of Rare Genetic Variants. *Science* **2012**, *336*, 740–743. [CrossRef] [PubMed]

58. Tishkoff, S.A.; Verrelli, B.C. Patterns of Human Genetic Diversity: Implications for Human Evolutionary History and Disease. *Annu. Rev. Genom. Hum. Genet.* **2003**, *4*, 293–340. [CrossRef] [PubMed]

59. Campbell, M.C.; Tishkoff, S.A. African Genetic Diversity: Implications for Human Demographic History, Modern Human Origins, and Complex Disease Mapping. *Annu. Rev. Genom. Hum. Genet.* **2008**, *9*, 403–433. [CrossRef]

60. Henn, B.M.; Gignoux, C.R.; Jobin, M.; Granka, J.M.; Macpherson, J.M.; Kidd, J.M.; Rodríguez-Botigué, L.; Ramachandran, S.; Hon, L.; Brisbin, A.; et al. Hunter-Gatherer Genomic Diversity Suggests a Southern African Origin for Modern Humans. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 5154–5162. [CrossRef]

61. Basel, D.; Steiner, R.D. Osteogenesis Imperfecta: Recent Findings Shed New Light on This Once Well-Understood Condition. *Genet. Med.* **2009**, *11*, 375–385. [CrossRef]

62. Prado-Martinez, J.; Sudmant, P.H.; Kidd, J.M.; Li, H.; Kelley, J.L.; Lorente-Galdos, B.; Veeramah, K.R.; Woerner, A.E.; O'Connor, T.D.; Santpere, G.; et al. Great Ape Genetic Diversity and Population History. *Nature* **2013**, *499*, 471–475. [CrossRef] [PubMed]

63. De Manuel, M.; Kuhlwilm, M.; Frandsen, P.; Sousa, V.C.; Desai, T.; Prado-Martinez, J.; Hernandez-Rodriguez, J.; Dupanloup, I.; Lao, O.; Hallast, P.; et al. Chimpanzee Genomic Diversity Reveals Ancient Admixture with Bonobos. *Science* **2016**, *354*, 477–481. [CrossRef]

64. Won, Y.-J.; Hey, J. Divergence Population Genetics of Chimpanzees. *Mol. Biol. Evol.* **2005**, *22*, 297–307. [CrossRef] [PubMed]

65. Becquet, C.; Przeworski, M. A New Approach to Estimate Parameters of Speciation Models with Application to Apes. *Genome Res.* **2007**, *17*, 1505–1519. [CrossRef] [PubMed]

66. Stone, A.C.; Griffiths, R.C.; Zegura, S.L.; Hammer, M.F. High Levels of Y-Chromosome Nucleotide Diversity in the Genus Pan. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 43–48. [CrossRef]

67. Fischer, A.; Wiebe, V.; Pääbo, S.; Przeworski, M. Evidence for a Complex Demographic History of Chimpanzees. *Mol. Biol. Evol.* **2004**, *21*, 799–808. [CrossRef] [PubMed]

68. Claw, K.G.; Tito, R.Y.; Stone, A.C.; Verrelli, B.C. Haplotype Structure and Divergence at Human and Chimpanzee Serotonin Transporter and Receptor Genes: Implications for Behavioral Disorder Association Analyses. *Mol. Biol. Evol.* **2010**, *27*, 1518–1529. [CrossRef]

69. Verrelli, B.C.; Tishkoff, S.A.; Stone, A.C.; Touchman, J.W. Contrasting Histories of G6PD Molecular Evolution and Malarial Resistance in Humans and Chimpanzees. *Mol. Biol. Evol.* **2006**, *23*, 1592–1601. [CrossRef]

70. Verrelli, B.C.; Lewis, C.M.; Stone, A.C.; Perry, G.H. Different Selective Pressures Shape the Molecular Evolution of Color Vision in Chimpanzee and Human Populations. *Mol. Biol. Evol.* **2008**, *25*, 2735–2743. [CrossRef]

71. Kuhn, R.M.; Haussler, D.; Kent, W.J. The UCSC Genome Browser and Associated Tools. *Brief Bioinform.* **2013**, *14*, 144–161. [CrossRef]

72. Hothorn, T.; Hornik, K.; van de Wiel, M.A.; Zeileis, A. Implementing a Class of Permutation Tests: The coin Package. *J. Stat. Soft.* **2008**, *28*, 1–23. [CrossRef]

73. McDonald, J.H.; Kreitman, M. Adaptive Protein Evolution at the Adh Locus in Drosophila. *Nature* **1991**, *351*, 652–654. [CrossRef]

74. Watterson, G.A. On the Number of Segregating Sites in Genetical Models without Recombination. *Theor. Popul. Biol.* **1975**, *7*, 256–276. [CrossRef]

75. Hudson, R.R.; Slatkin, M.; Maddison, W.P. Estimation of Levels of Gene Flow from DNA Sequence Data. *Genetics* **1992**, *132*, 583–589. [CrossRef] [PubMed]

76. Hudson, R.R. A New Statistic for Detecting Genetic Differentiation. *Genetics* **2000**, *155*, 2011–2014. [CrossRef]

77. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. 1000 Genomes Project Analysis Group. The Variant Call Format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [CrossRef] [PubMed]

78. Stephens, M.; Smith, N.J.; Donnelly, P. A New Statistical Method for Haplotype Reconstruction from Population Data. *Am. J. Hum. Genet.* **2001**, *68*, 978–989. [CrossRef]

79. Rozas, J.; Ferrer-Mata, A.; Sánchez-DelBarrio, J.C.; Guirao-Rico, S.; Librado, P.; Ramos-Onsins, S.E.; Sánchez-Gracia, A. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol. Biol. Evol.* **2017**, *34*, 3299–3302. [CrossRef]

80. Sałacińska, K.; Pinkier, I.; Rutkowska, L.; Chlebna-Sokół, D.; Jakubowska-Pietkiewicz, E.; Michałus, I.; Kępczyński, Ł.; Salachna, D.; Jamsheer, A.; Bukowska-Olech, E.; et al. Novel Mutations Within Collagen Alpha1(I) and Alpha2(I) Ligand-Binding Sites, Broadening the Spectrum of Osteogenesis Imperfecta—Current Insights into Collagen Type I Lethal Regions. *Front. Genet.* **2021**, *12*, 692978. [CrossRef] [PubMed]

81. Ohta, T. The Nearly Neutral Theory of Molecular Evolution. *Annu. Rev. Ecol. Syst.* **1992**, *23*, 263–286. [CrossRef]

82. Charlesworth, J.; Eyre-Walker, A. The McDonald-Kreitman Test and Slightly Deleterious Mutations. *Mol. Biol. Evol.* **2008**, *25*, 1007–1015. [CrossRef] [PubMed]

83. Housman, G.; Briscoe, E.; Gilad, Y. Evolutionary insights into primate skeletal gene regulation using a comparative cell culture model. *bioRxiv* **2021**. [CrossRef]

84. Bradnam, K.R.; Korf, I. Longer First Introns Are a General Property of Eukaryotic Gene Structure. *PLoS ONE* **2008**, *3*, e3093. [CrossRef] [PubMed]

85. Rosenberg, N.A.; Pritchard, J.K.; Weber, J.L.; Cann, H.M.; Kidd, K.K.; Zhivotovsky, L.A.; Feldman, M.W. Genetic Structure of Human Populations. *Science* **2002**, *298*, 2381–2385. [CrossRef]

86. Tishkoff, S.A.; Reed, F.A.; Friedlaender, F.R.; Ehret, C.; Ranciaro, A.; Froment, A.; Hirbo, J.B.; Awomoyi, A.A.; Bodo, J.-M.; Doumbo, O.; et al. The Genetic Structure and History of Africans and African Americans. *Science* **2009**, *324*, 1035–1044. [CrossRef] [PubMed]

87. Tishkoff, S.A.; Verrelli, B.C. Role of Evolutionary History on Haplotype Block Structure in the Human Genome: Implications for Disease Mapping. *Curr. Opin. Genet. Dev.* **2003**, *13*, 569–575. [CrossRef]

88. Thomson, R.; Pritchard, J.K.; Shen, P.; Oefner, P.J.; Feldman, M.W. Recent Common Ancestry of Human Y Chromosomes: Evidence from DNA Sequence Data. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 7360–7365. [CrossRef]

89. Griffiths, R.C.; Tavaré, S. The Age of a Mutation in a General Coalescent Tree. *Commun. Statistics. Stoch. Models* **1998**, *14*, 273–295. [CrossRef]

90. Fu, W.; O'Connor, T.D.; Jun, G.; Kang, H.M.; Abecasis, G.; Leal, S.M.; Gabriel, S.; Rieder, M.J.; Altshuler, D.; Shendure, J.; et al. Analysis of 6,515 Exomes Reveals the Recent Origin of Most Human Protein-Coding Variants. *Nature* **2013**, *493*, 216–220. [CrossRef]

91. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef]

92. Gonder, M.K.; Locatelli, S.; Ghobrial, L.; Mitchell, M.W.; Kujawski, J.T.; Lankester, F.J.; Stewart, C.-B.; Tishkoff, S.A. Evidence from Cameroon Reveals Differences in the Genetic Structure and Histories of Chimpanzee Populations. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 4766–4771. [CrossRef] [PubMed]

93. Boyko, A.R.; Williamson, S.H.; Indap, A.R.; Degenhardt, J.D.; Hernandez, R.D.; Lohmueller, K.E.; Adams, M.D.; Schmidt, S.; Sninsky, J.J.; Sunyaev, S.R.; et al. Assessing the Evolutionary Impact of Amino Acid Mutations in the Human Genome. *PLoS Genet.* **2008**, *4*, e1000083. [CrossRef] [PubMed]

94. Lohmueller, K.E.; Indap, A.R.; Schmidt, S.; Boyko, A.R.; Hernandez, R.D.; Hubisz, M.J.; Sninsky, J.J.; White, T.J.; Sunyaev, S.R.; Nielsen, R.; et al. Proportionally More Deleterious Genetic Variation in European than in African Populations. *Nature* **2008**, *451*, 994–997. [CrossRef] [PubMed]

95. Bustamante, C.D.; Fledel-Alon, A.; Williamson, S.; Nielsen, R.; Hubisz, M.T.; Glanowski, S.; Tanenbaum, D.M.; White, T.J.; Sninsky, J.J.; Hernandez, R.D.; et al. Natural Selection on Protein-Coding Genes in the Human Genome. *Nature* **2005**, *437*, 1153–1157. [CrossRef]

96. Anderson-Trocmé, L.; Farouni, R.; Bourgey, M.; Kamatani, Y.; Higasa, K.; Seo, J.-S.; Kim, C.; Matsuda, F.; Gravel, S. Legacy Data Confound Genomics Studies. *Mol. Biol. Evol.* **2020**, *37*, 2–10. [CrossRef]

97. Belsare, S.; Levy-Sakin, M.; Mostovoy, Y.; Durinck, S.; Chaudhuri, S.; Xiao, M.; Peterson, A.S.; Kwok, P.-Y.; Seshagiri, S.; Wall, J.D. Evaluating the Quality of the 1000 Genomes Project Data. *BMC Genom.* **2019**, *20*, 620. [CrossRef]

98. Rensberger, J.M.; Watabe, M. Fine Structure of Bone in Dinosaurs, Birds and Mammals. *Nature* **2000**, *406*, 619–622. [CrossRef]

99. Abbott, S.; Trinkaus, E.; Burr, D.B. Dynamic Bone Remodeling in Later Pleistocene Fossil Hominids. *Am. J. Phys. Anthropol.* **1996**, *99*, 585–601. [CrossRef]

100. Larsen, C.S. Biological Changes in Human Populations with Agriculture. *Annu. Rev. Anthropol.* **1995**, *24*, 185–213. [CrossRef]

101. Przeworski, M.; Coop, G.; Wall, J.D. The Signature of Positive Selection on Standing Genetic Variation. *Evolution* **2005**, *59*, 2312–2323. [CrossRef] [PubMed]

102. Bitarello, B.D.; de Filippo, C.; Teixeira, J.C.; Schmidt, J.M.; Kleinert, P.; Meyer, D.; Andrés, A.M. Signatures of Long-Term Balancing Selection in Human Genomes. *Genome Biol. Evol.* **2018**, *10*, 939–955. [CrossRef] [PubMed]

103. Ram, O.; Goren, A.; Amit, I.; Shoresh, N.; Yosef, N.; Ernst, J.; Kellis, M.; Gymrek, M.; Issner, R.; Coyne, M.; et al. Combinatorial Patterning of Chromatin Regulators Uncovered by Genome-Wide Location Analysis in Human Cells. *Cell* **2011**, *147*, 1628–1639. [CrossRef]

104. Verrelli, B.C.; McDonald, J.H.; Argyropoulos, G.; Destro-Bisol, G.; Froment, A.; Drousiotou, A.; Lefranc, G.; Helal, A.N.; Loiselet, J.; Tishkoff, S.A. Evidence for Balancing Selection from Nucleotide Sequence Analyses of Human G6PD. *Am. J. Hum. Genet.* **2002**, *71*, 1112–1128. [CrossRef] [PubMed]

105. Verrelli, B.C.; Tishkoff, S.A. Signatures of Selection and Gene Conversion Associated with Human Color Vision Variation. *Am. J. Hum. Genet.* **2004**, *75*, 363–375. [CrossRef]

106. Tishkoff, S.A.; Reed, F.A.; Ranciaro, A.; Voight, B.F.; Babbitt, C.C.; Silverman, J.S.; Powell, K.; Mortensen, H.M.; Hirbo, J.B.; Osman, M.; et al. Convergent Adaptation of Human Lactase Persistence in Africa and Europe. *Nat. Genet.* **2007**, *39*, 31–40. [CrossRef]

107. Gurdasani, D.; Carstensen, T.; Fatumo, S.; Chen, G.; Franklin, C.S.; Prado-Martinez, J.; Bouman, H.; Abascal, F.; Haber, M.; Tachmazidou, I.; et al. Uganda Genome Resource Enables Insights into Population History and Genomic Discovery in Africa. *Cell* **2019**, *179*, 984–1002.e36. [CrossRef]

108. Hancock, A.M.; Witonsky, D.B.; Ehler, E.; Alkorta-Aranburu, G.; Beall, C.; Gebremedhin, A.; Sukernik, R.; Utermann, G.; Pritchard, J.; Coop, G.; et al. Colloquium Paper: Human Adaptations to Diet, Subsistence, and Ecoregion Are Due to Subtle Shifts in Allele Frequency. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 8924–8930. [CrossRef] [PubMed]

109. Wong, H.H.; Seet, S.H.; Bascom, C.C.; Isfort, R.J.; Bard, F. Red-COLA1: A Human Fibroblast Reporter Cell Line for Type I Collagen Transcription. *Sci. Rep.* **2020**, *10*, 19723. [CrossRef]

110. Al-Barghouthi, B.M.; Mesner, L.D.; Calabrese, G.M.; Brooks, D.; Tommasini, S.M.; Bouxsein, M.L.; Horowitz, M.C.; Rosen, C.J.; Nguyen, K.; Haddox, S.; et al. Systems Genetics in Diversity Outbred Mice Inform BMD GWAS and Identify Determinants of Bone Strength. *Nat. Commun.* **2021**, *12*, 3408. [CrossRef]

111. Stone, A.C.; Verrelli, B.C. Focusing on Comparative Ape Population Genetics in the Post-Genomic Age. *Curr. Opin. Genet. Dev.* **2006**, *16*, 586–591. [CrossRef] [PubMed]