Please cite this article in press as: Payseur et al., Linkage Disequilibrium between STRPs and SNPs across the Human Genome, The American Journal of Human Genetics (2008), doi:10.1016/j.ajhg.2008.02.018

ARTICLE

Linkage Disequilibrium between STRPs and SNPs across the Human Genome

Bret A. Payseur,[1,*] Michael Place,[1] and James L. Weber[2]

Patterns of linkage disequilibrium (LD) reveal the action of evolutionary processes and provide crucial information for association mapping of disease genes. Although recent studies have described the landscape of LD among single nucleotide polymorphisms (SNPs) from across the human genome, associations involving other classes of molecular variation remain poorly understood. In addition to recombination and population history, mutation rate and process are expected to shape LD. To test this idea, we measured associations between short-tandem-repeat polymorphisms (STRPs), which can mutate rapidly and recurrently, and SNPs in 721 regions across the human genome. We directly compared STRP-SNP LD with SNP-SNP LD from the same genomic regions in the human HapMap populations. The intensity of STRP-SNP LD, measured by the average of D', was reduced, consistent with the action of recurrent mutation. Nevertheless, a higher fraction of STRP-SNP pairs than SNP-SNP pairs showed significant LD, on both short (up to 50 kb) and long (cM) scales. These results reveal the substantial effects of mutational processes on LD at STRPs and provide important measures of the potential of STRPs for association mapping of disease genes.

Introduction

Linkage disequilibrium (LD), the correlation among DNA polymorphisms in populations, is a key quantity in human genetics. Because LD is broken down by recombination and shaped by demographic and selective history, patterns of LD can provide detailed information about these evolutionary forces.[1–8] The level of LD in a genomic region also predicts the power to locate genetic variants that underlie phenotypic differences through association mapping.[9–12] Along with advances in high-density genotyping, these insights have spurred successful efforts to describe and interpret patterns of LD across the human genome.[13–15]

In addition to being shaped by recombination and population history, LD is also shaped by mutation. Markers with higher mutation rates have the potential to detect LD with greater power because more branches of the sample genealogy are "marked" by mutations.[16–18] Additionally, multiple mutations to alleles with the same lengths can erase the record of genealogical history, thereby reducing LD. A class of molecular markers widely used in human genetics, short-tandem-repeat polymorphisms (STRPs), have these characteristics. STRPs mutate rapidly (typically 10^{-3}–10^{-5} per generation),[19,20] primarily through replication slippage.[19,21] As a result, human populations segregate many alleles at individual STRPs, and some fraction of these alleles is identical by state but not identical by descent. These attributes contrast with single nucleotide polymorphisms (SNPs), which arise at a low rate (10^{-8}–10^{-9} per generation)[22] and usually represent unique mutational events. Differences in mutational dynamics therefore translate into contrasting levels of marker informativeness for STRPs and SNPs.[23]

Genomic analyses of LD in humans have focused primarily on SNPs,[13–15,24–30] with the emergence of several notable patterns. The spatial extent of SNP-SNP LD (1) is on the order of tens of kb (on average), (2) decreases with recombination rate, (3) varies among genomic regions, and (4) differs between populations.

LD involving STRPs has also been measured in human populations. Genomic examinations of STRP-STRP LD include a study of 5048 markers in the CEU (individuals of northern and western European ancestry living in Utah from the Centre d'Etude du Polymorphisme Humain [CEPH] collection) panel[31] and an analysis of 179 markers in a large sample from the Icelandic population.[31] Both studies showed that STRP-STRP LD decays with recombinational distance and varies significantly among genomic regions, like SNP-SNP LD. LD between STRPs and SNPs has also been measured. Detailed investigations of several genomic regions have revealed that statistically significant STRP-SNP LD extends further than does SNP-SNP LD.[32,33] However, LD involving STRPs has never been directly compared to SNP-SNP LD on a genomic scale in the same set of individuals. Such an investigation is motivated by several goals.

First, because STRPs and SNPs are known to mutate differently, comparisons among these markers allow the effects of the mutational process on LD to be empirically examined. Second, relative patterns of LD at SNPs and STRPs provide guidance concerning marker choice for studies that associate genotype and phenotype in human populations.[34,35] The integration of STRPs with SNPs should help in the identification of disease mutations, as do other copy number variants.[36–38] Finally, LD between STRPs and SNPs provides important information for population-genetic approaches that combine data from both marker classes.[39,40]

Here, we report patterns of LD between STRPs and SNPs in three human populations. By comparing STRP-SNP LD with SNP-SNP LD in the same set of individuals, we extend to the genomic scale the observation that STRPs more

[1]Laboratory of Genetics, University of Wisconsin, Madison, WI 53706, USA; [2]Prevention Genetics, Marshfield, WI 54449, USA
*Correspondence: payseur@wisc.edu
DOI 10.1016/j.ajhg.2008.02.018.

Please cite this article in press as: Payseur et al., Linkage Disequilibrium between STRPs and SNPs across the Human Genome, The American Journal of Human Genetics (2008), doi:10.1016/j.ajhg.2008.02.018

readily detect statistically significant associations. We also demonstrate that STRP-SNP LD is reduced by recurrent mutation and dependent on repeat type. Our results highlight the effects of mutational mechanisms on LD and motivate a population-genetic framework that combines patterns of variation at SNPs and STRPs.

Material and Methods

STRP Genotyping and Selection of SNPs

STRPs for genotyping were from Marshfield 5 cM genomic linkage screening sets (see Web Resources). These markers were chosen to be uniformly spaced, highly informative, and easy to type accurately.[41] Genotyping was performed in the Mammalian Genotyping Service as previously described.[42]

We determined the genomic positions of 721 autosomal STRPs from the screening sets by BLATing the consensus sequence to the human genome sequence at the UCSC website (hg17; Build 35). Of these 721 STRPs, 51 were dinucleotide repeats, 149 were trinucleotide repeats, 511 were tetranucleotide repeats, and 10 were pentanucleotide repeats. For phased analyses, genotypes of all SNPs within 50kb of each microsatellite were downloaded from the HapMap website (public release 21). To conduct longer-range, unphased analyses, the cM position of each STRP was estimated using the high-density STRP human genetic map.[43] Two hundred seventy three of the STRPs were directly placed on this map; the cM position of each remaining STRP was estimated as the position of the closest mapped STRP in the sequence. The sequence positions of mapped STRPs nearest to 2 cM on either side of each STRP, assuming a constant recombination rate in each region (but allowing rates to vary among regions), were used to delineate a window of approximately 4 cM in size centered on each STRP. All SNPs falling within these windows were obtained from the HapMap website.

Analyses

Individuals from the CHB (Han Chinese individuals living in Beijing, China) and JPT (Japanese individuals living in Tokyo, Japan) populations were combined (denoted hereafter as "CHB+JPT") for purposes of this study.[14] The CEU, YRI (individuals from the Yoruba population in Ibadan, Nigeria), and CHB+JPT populations were considered separately in all analyses. Autosomal haplotypes including each STRP and all non-singleton SNPs within 50 kb were computationally phased using PHASE v.2.1.[44,45] This distance was selected based on average haplotype block sizes reported for the SNP-dense ENCODE regions in these individuals[14] and computational constraints associated with phasing. PHASE assumes an infinite-sites model for SNPs and a symmetrical, one-step stepwise-mutation model for STRPs. For the CEU and YRI populations, genotypes from children were used in haplotype reconstruction.[46] All genotypes that departed from Mendelian transmission were re-coded as missing. If a genotype was absent in one or both parents, the genotype of the corresponding child was also re-coded as missing. One CEU parent was not genotyped for STRPs and the matching parent and child were removed prior to all analyses. For each individual, the haplotype pair with the highest posterior probability estimated by PHASE was used for subsequent analyses. All genomic regions except two in CEU and three in YRI were successfully phased. The remaining 719, 718, and 721 autosomal regions (in CEU, YRI, and CHB+JPT, respectively) were the focus of our analyses. We analyzed an additional 31 X-linked STRPs in males only, where haplotypes were directly observed, to investigate LD in the absence of phasing. All LD analyses involved only unrelated individuals: 58 parents from CEU trios, 60 parents from YRI trios, and 90 (45+45) CHB+JPT individuals.

For phased haplotypes, D' [47] was estimated for each pair of alleles from each STRP-SNP (designated as D'_{ind}) or SNP-SNP combination. For each STRP-SNP combination, the multi-allelic, average D' (designated as D'_{avg}) was also estimated as

$$D'_{avg} = \sum_{i=1}^{k} \sum_{j=1}^{l} p_i q_j \, |D'_{ij}|$$

(ref. [48]) with the Haploxt program in the GOLD package.[49] Statistical evidence for association was measured with contingency tables of haplotype counts. p values for SNP-SNP associations (2×2 tables) were estimated with Fisher's Exact Test (FET). p values for STRP-SNP associations were also estimated with FET, but the null distribution was obtained from 10,000 randomized tables because the number of STRP alleles was large.

To evaluate long-range LD, composite genotypic disequilibrium (CGD)[50–52] between STRPs and all SNPs within 2 cM was calculated with the use of unphased diploid genotypes. CGD between the SNP closest to each STRP and all remaining SNPs in the window was estimated for comparison. Statistical significance was measured with a χ^2 test. CGD analyses were restricted to alleles with at least 5% frequency to minimize departures from the asymptotic χ^2 approximation caused by sparse contingency tables. The CGD approach directly uses unphased diploid genotypes, obviating the need for phasing (which can be very difficult at larger recombinational distances) and thus avoiding effects of phasing error. These analyses were implemented with an R script kindly provided by Daniel Schaid.[50]

For both phased and unphased LD analyses, the overall (genomic) proportion of tests for which the null hypothesis of no association was rejected (m_1/m) was estimated from the pooled distribution of p values with the use of a false-discovery-rate approach[53–55] implemented in Storey's Qvalue package in R. Values of m_1/m were estimated separately for different data subsets (STRP versus SNP tests, STRP repeat types, etc.). Because SNP-SNP FET p value distributions were noncontinuous with a peak at 1 (a feature of FET[56]), we used the bootstrap method (rather than the smoother method) to estimate m_1/m.[55]

For each STRP, polymorphism summary statistics, including expected heterozygosity and variance in repeat number, were calculated separately for each population with Microsatellite Analyzer.[57]

Results

Genomic Distributions of STRP-SNP LD

To measure LD between STRPs and SNPs, we genotyped 721 autosomal and 31 X-linked STRPs in 268 individuals from the HapMap project.[14] These individuals had already been genotyped at more than 3.1 million SNPs. The STRPs were approximately uniformly spaced along each of the chromosomes.[41]

We first analyzed D' between autosomal STRPs and SNPs separated by less than 50 kb on haplotypes reconstructed with the use of PHASE.[44,45] Genomic distributions of D' between STRPs and SNPs were characterized by several

Please cite this article in press as: Payseur et al., Linkage Disequilibrium between STRPs and SNPs across the Human Genome, The American Journal of Human Genetics (2008), doi:10.1016/j.ajhg.2008.02.018

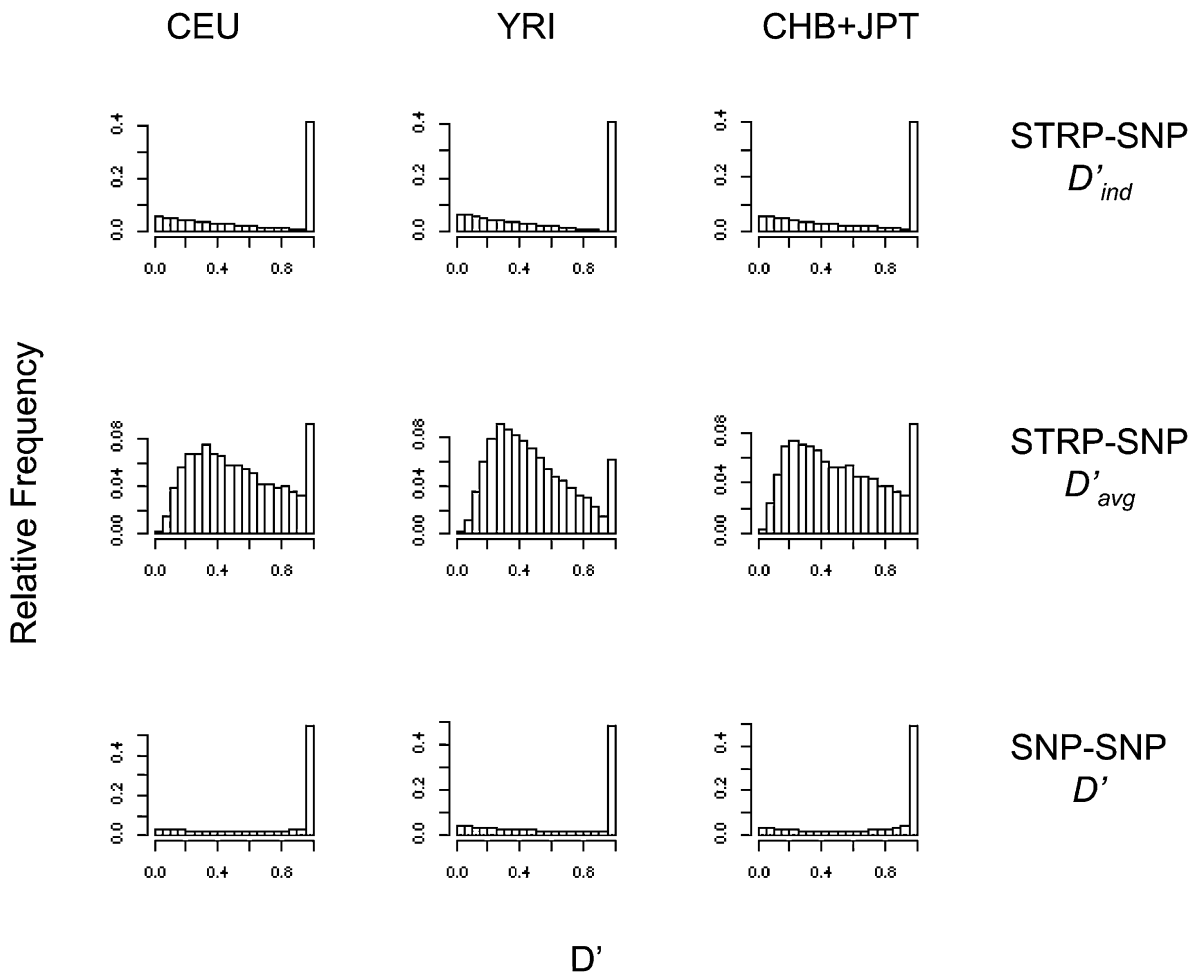

Figure 1. *D′* between Loci within 50 kb
Values are pooled across genomic regions. $D'_{ind} = D'$ with individual STRP alleles; D'_{avg} = multi-allelic D'.

patterns (Figure 1). First, many locus pairs were perfectly associated, with D' values of 1. Second, a substantial fraction of tests yielded D' values less than 1, consistent with the action of recombination or recurrent mutation. This pattern was observed in D' distributions among individual pairs of alleles (D'_{ind}) and was considerably stronger in distributions of D' averaged across allele pairs (D'_{avg}) (Figure 1), suggesting substantial heterogeneity among alleles within individual STRPs. Comparisons with D' among SNP pairs drawn from the same genomic regions indicated that: (1) STRP-SNP D'_{ind} was similar to but lower than SNP-SNP D' (Mann-Whitney U test; $p < 10^{-15}$ in all populations), and (2) STRP-SNP D'_{avg} was considerably lower than SNP-SNP D' ($p < 10^{-15}$ in all populations). STRP-SNP D' distributions differed among populations (Kruskal-Wallis test; $p < 10^{-15}$) (Figure 1). In particular, YRI showed lower D'_{avg} values (median: YRI = 0.43; CEU = 0.49; CHB+JPT = 0.47), in agreement with patterns of SNP-SNP LD (International HapMap Consortium, 2005) and presumably reflecting the larger effective size of the Yoruban population.

D' levels among STRPs and SNPs separated by up to 50 kb are likely to have been reduced by both recombination and recurrent mutation during the history of the population samples. To further gauge the contribution of recurrent mutation, we examined D' between the subset of loci located less than 5 kb apart (Figure 2), where effects of recombination should be less visible. STRP-SNP D' was increased relative to the 50 kb regions (compare to Figure 1) (median: CEU = 0.65; YRI = 0.57; CHB+JPT = 0.66), revealing the effects of tighter linkage. However, both D'_{ind} and D'_{avg} were still reduced in comparison to SNP-SNP pairs ($p < 10^{-15}$ in all tests), demonstrating that recurrent mutation had significantly shaped STRP-SNP LD.

Because repeat types mutate at different rates,[20,58] we also compared D' distributions among repeat types (excluding pentanucleotides, for which only ten loci were available) (Figure 3). Significant variation in D' among repeat types was observed in each population (Kruskal-Wallis test; $p < 10^{-15}$ in all tests), with dinucleotides (D'_{avg} median: CEU = 0.59; YRI = 0.52; CHB+JPT = 0.58) and trinucleotides (D'_{avg} median: CEU = 0.62; YRI = 0.53; CHB+JPT = 0.61) showing higher D'_{avg} than that of tetranucleotides (D'_{avg} median: CEU = 0.44; YRI = 0.40; CHB+JPT = 0.43). Comparisons to D' among SNP-SNP pairs from the corresponding genomic regions confirmed that STRP-SNP D'_{avg} was reduced in each repeat class ($p < 10^{-15}$ in all tests).

Please cite this article in press as: Payseur et al., Linkage Disequilibrium between STRPs and SNPs across the Human Genome, The American Journal of Human Genetics (2008), doi:10.1016/j.ajhg.2008.02.018

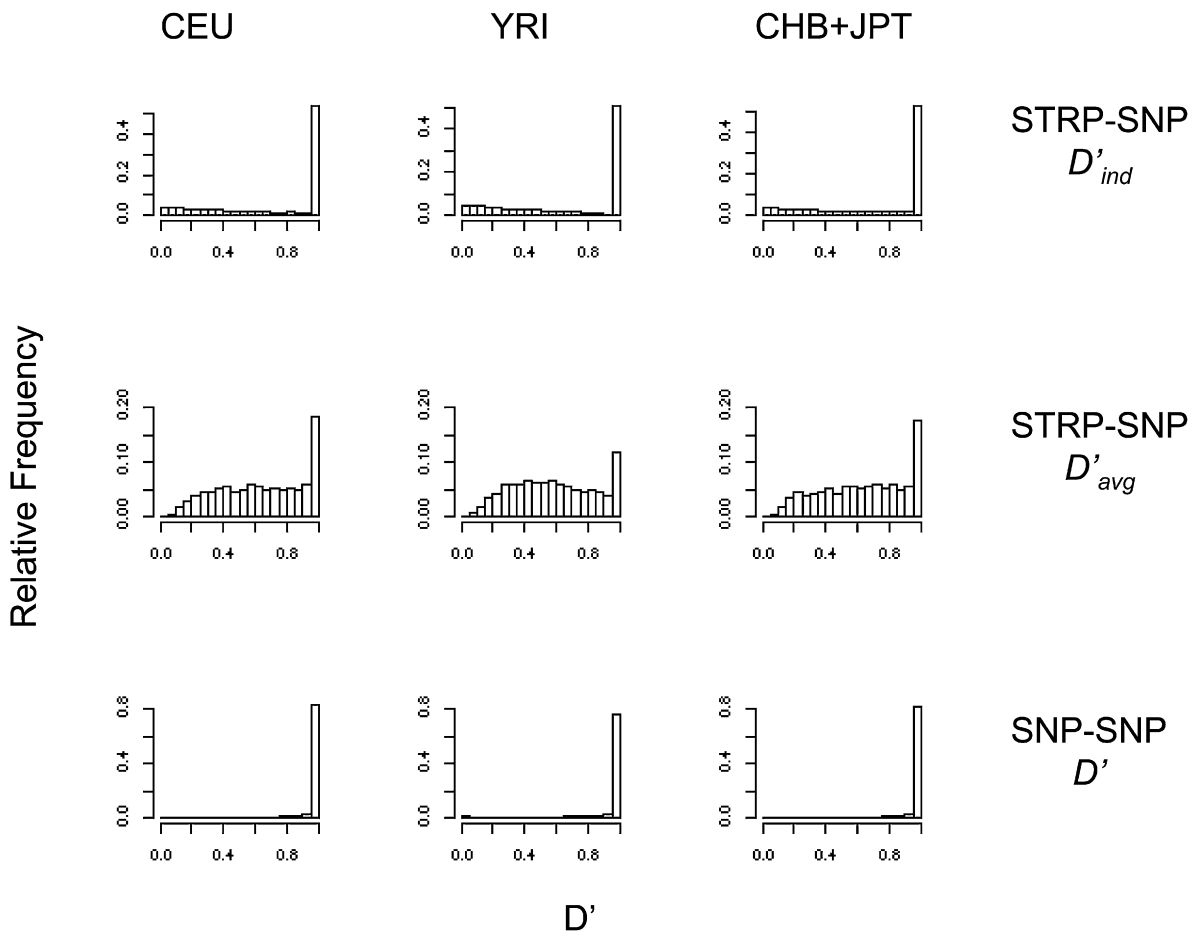

Figure 2. *D′* between Loci within 5 kb
Values are pooled across genomic regions. $D'_{ind} = D'$ with individual STRP alleles; D'_{avg} = multi-allelic D'.

Although D' describes the intensity of LD in useful ways, it does not measure statistical significance. We used Fisher's Exact Test to calculate p values for the tables of haplotype counts from all two-locus pairs located in the 50 kb intervals. Then, we separately estimated the genomic fractions of STRP-SNP and SNP-SNP tests for which the null hypothesis of no association was rejected (m_1/m) by using a false-discovery-rate approach.[53–55] This method accounts for the performance of multiple tests by considering the full distribution of p values.

Values of m_1/m for STRP-SNP pairs exceeded those for SNP-SNP pairs from corresponding genomic regions in all populations (Figure 4). The contrast was especially strong in the YRI population. Proportions of locus pairs showing significant LD differed substantially among STRP repeat types (Figure 4). In agreement with analyses of D', dinucleotides and trinucleotides showed more evidence for LD than did tetranucleotides.

Paired Comparisons between STRP-SNP and SNP-SNP LD

In addition to comparing the full distributions of STRP-SNP LD and SNP-SNP LD, we sought to test whether STRPs and SNPs located near each other differed in the ability to detect LD with the same SNPs. We conducted paired comparisons in which each (target) SNP within 50 kb of a genotyped STRP was considered with (1) the STRP and (2) the SNP located nearest the STRP. The average distances between STRPs and the closest SNPs were 589 bp (CEU), 518 bp (YRI), and 621 bp (CHB+JPT). Loci from such paired tests probably had identical genealogical histories and shared all properties of the target SNPs. As a result, variation in recombination rate and target-SNP allele frequency could not contribute to observed differences between STRPs and SNPs.

STRP-SNP and SNP-SNP D' values were positively correlated (Spearman's ρ: CEU = 0.43; YRI = 0.40; CHB+JPT = 0.48; $p < 10^{-15}$ in all populations), reflecting shared genealogical histories. Nevertheless, D' values were lower for STRPs than for SNPs (Wilcoxon matched-pairs signed-rank test; $p < 10^{-15}$ in all populations), and STRP p values were lower than SNP p values ($p < 10^{-15}$ in all populations). These patterns confirmed that the lower D' and stronger statistical significance of LD involving STRPs inferred from the full distributions (above) was not caused by differences in recombination rate, target-SNP allele frequency, or other factors that vary among genomic regions. STRP-SNP D'_{avg} showed a slightly weaker relationship with physical distance (Spearman's ρ: CEU = −0.22;

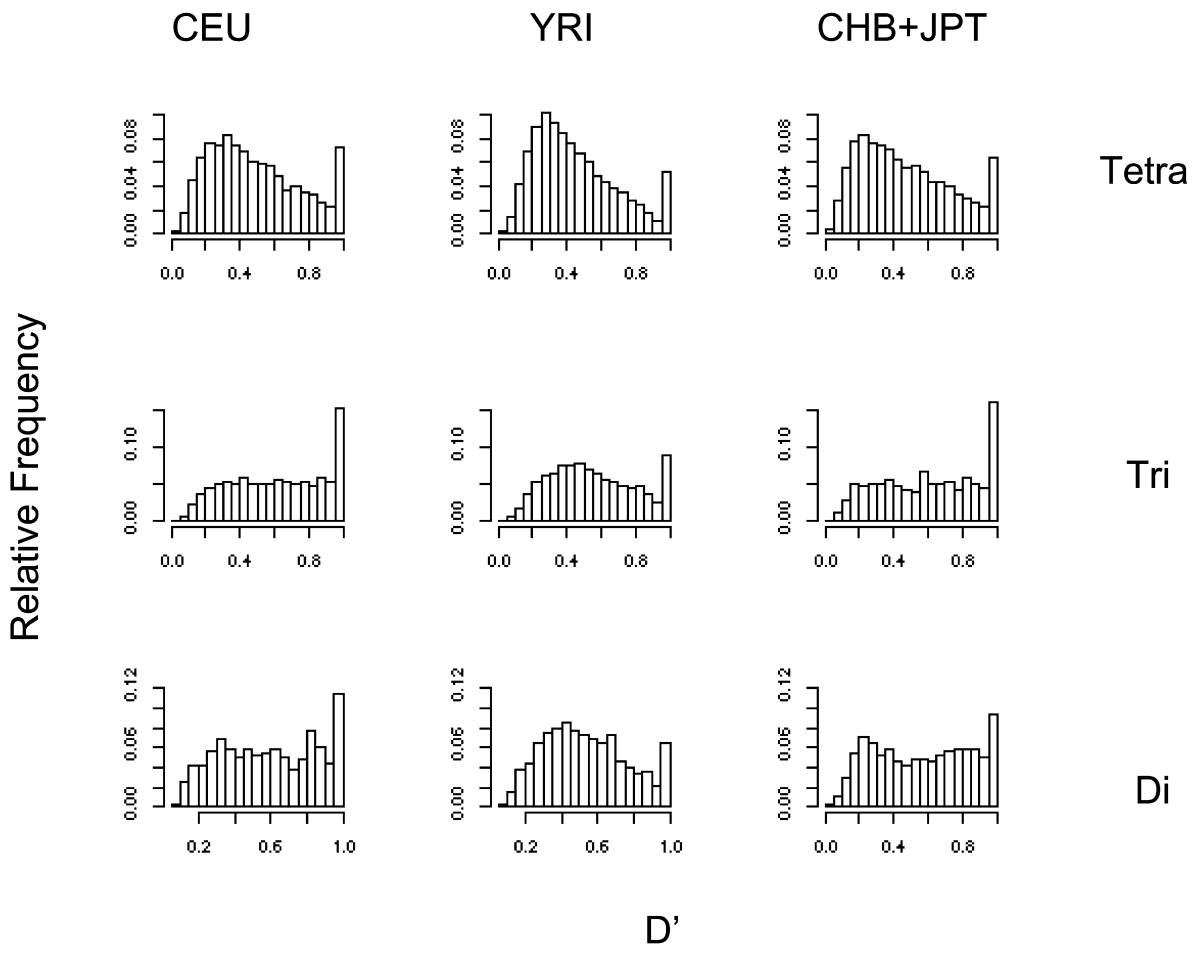

Figure 3. D'_{avg} between STRPs and SNPs by Repeat Type
Loci are within 50 kb and values are pooled across genomic regions. Distributions are shown separately for regions containing tetranucleotide ("Tetra"), trinucleotide ("Tri"), and dinucleotide ("Di") repeats.

YRI = −0.19; CHB+JPT = −0.25) than did SNP-SNP D' (CEU = −0.30; YRI = −0.27; CHB+JPT = −0.33).

Differences in LD patterns at STRPs and SNPs could be primarily attributable to the larger number of alleles at STRPs. Although individual SNPs only harbor two alleles, neighboring SNPs can be combined to capture more diversity. To compare STRP and SNP LD for markers with similar amounts of variation, we (1) selected contiguous SNPs that

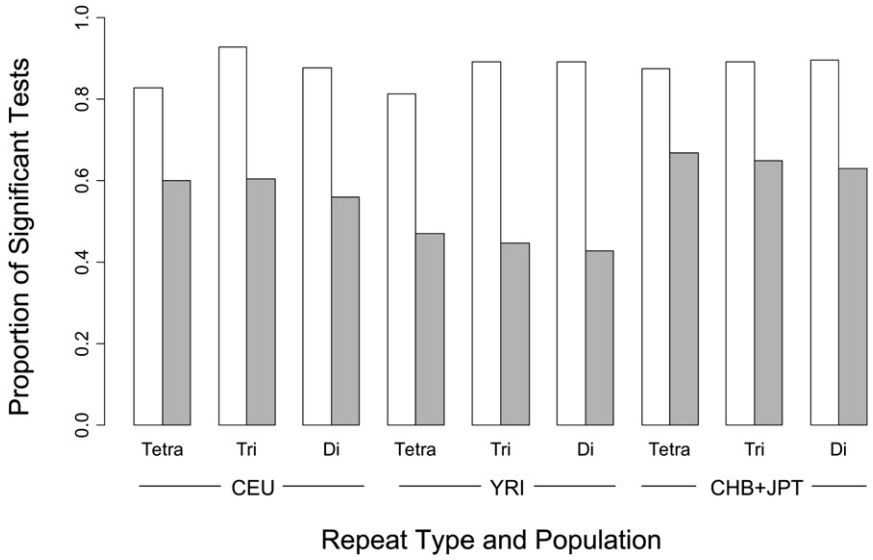

Figure 4. Genomic Proportions of Locus Pairs that Show Significant Linkage Disequilibrium
Proportions were estimated from analyses of phased haplotypes including loci within 50 kb using a false-discovery-rate approach. STRP-SNP tests are shown in white and SNP-SNP tests are shown in gray.

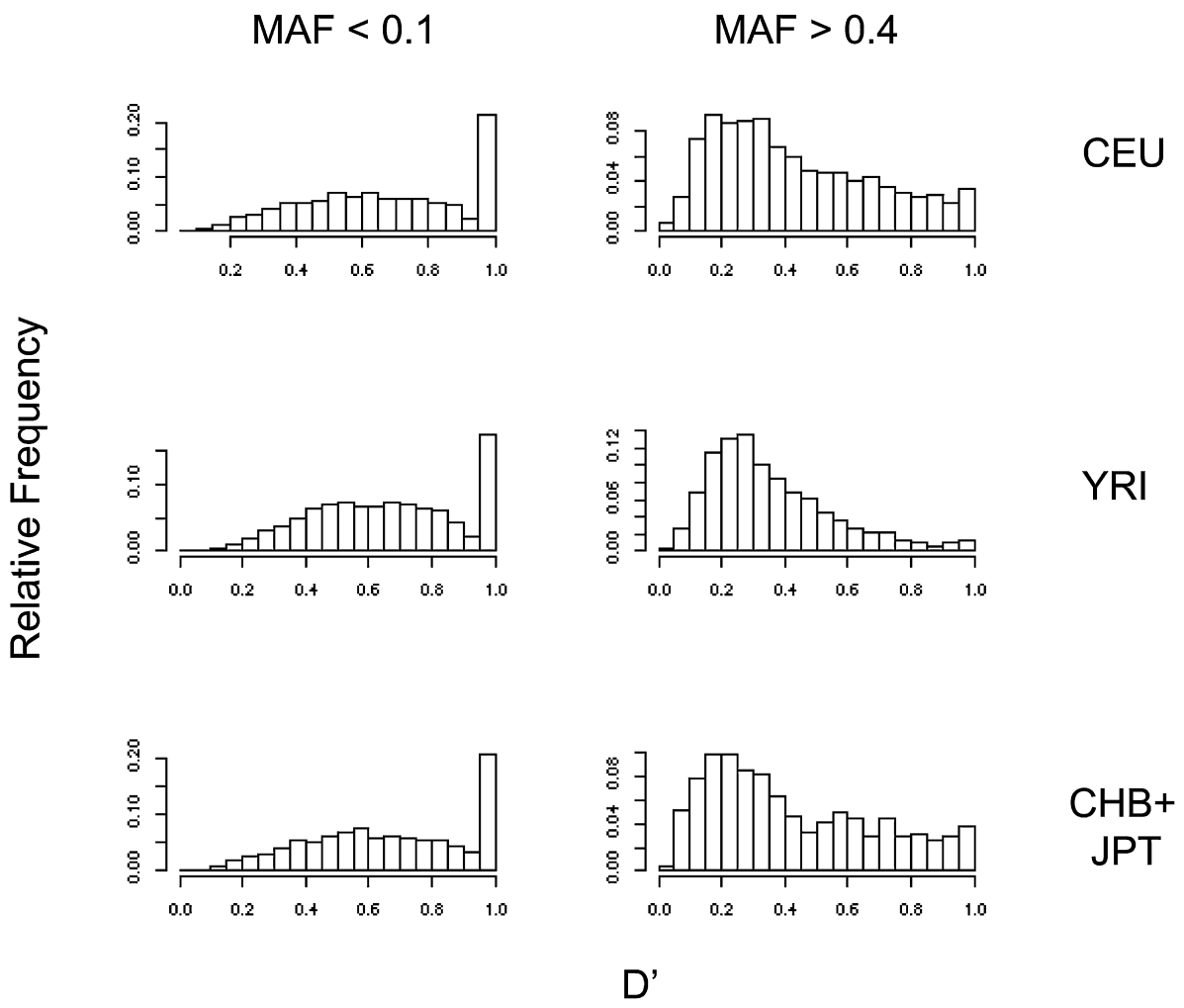

Figure 5. Effects of Allele Frequency on D'_{avg} between STRPs and SNPs
Loci are within 50kb. Values are separated into two categories defined by SNP minor allele frequency (MAF; < 0.1 versus > 0.4).

mapped closest to each STRP, (2) combined these SNPs to generate multi-SNP haplotypes, and (3) calculated LD between multi-SNP haplotypes and all remaining SNPs in the 50 kb regions. Seven-SNP haplotypes were used for these analyses because the average numbers of 7SNP haplotypes were similar to the average numbers of STRP alleles. Paired comparisons between 7SNP-SNP and STRP-SNP loci revealed lower p values at 7SNP-SNP combinations than at both 1SNP-SNP and STRP-SNP markers (Wilcoxon matched-pairs signed-rank test; $p < 10^{-15}$ in all populations). These results suggested that the increased statistical significance of LD at STRPs was largely driven by greater diversity.

Effects of Allele Frequency

Levels of LD depend on allele frequency, both for statistical reasons[48] and because alleles with higher frequencies tend to be older and are more likely to have experienced recombination. This phenomenon can be seen clearly in genomic SNP-SNP LD patterns.[7,13–15] We found that STRP-SNP D'_{avg} was also strongly influenced by SNP allele frequency, with clear reductions at higher minor-allele frequencies (Figure 5). This pattern probably reflected not only the increased recombination but also the higher probability of multiple mutations at STRP alleles paired with older SNPs.

Differences in frequency spectra between STRPs and SNPs could also contribute to observed patterns. STRPs harbor more rare alleles than do SNPs in human populations, and this difference is especially pronounced in the HapMap samples, where SNPs were ascertained to exhibit uniform frequency spectra.[14,59,60] To examine the cumulative effects of low-frequency alleles on relative patterns of LD, we repeated all analyses after removing STRP and SNP alleles with frequencies of less than 5%. In this filtered dataset, p values were slightly decreased for both STRP-SNP and SNP-SNP pairs. p values for SNP-SNP pairs were more similar to p values for STRP-SNP pairs than had occurred in unfiltered analyses. However, STRPs still retained a higher fraction of significant tests than did SNPs in all populations (results not shown).

To further account for potential effects of allele frequency on LD, we compared pairs of alleles matched by frequency. We chose the SNP closest to each STRP and then selected the STRP allele with the most similar

Please cite this article in press as: Payseur et al., Linkage Disequilibrium between STRPs and SNPs across the Human Genome, The American Journal of Human Genetics (2008), doi:10.1016/j.ajhg.2008.02.018

frequency to that SNP. Then, we measured D' between these alleles and all SNPs within 50 kb. The result was a paired set of tests with very similar allele frequencies at both loci and physical distances between them. On average, STRP and SNP alleles with frequency differences of less than 0.1 differed in D'_{ind} by 0.17 (CEU), 0.16 (YRI), and 0.17 (CHB+JPT) when considered with the same SNPs. These matched STRP-and-SNP D' distributions were significantly different (Wilcoxon matched-pairs signed-rank test; $p < 10^{-15}$ for all tests), demonstrating that the reduction in D' observed at STRP-SNP pairs was not caused by allele-frequency differences.

Effects of STRP-Polymorphism Levels

Markers with greater levels of variation are expected to detect LD with stronger statistical significance. This prediction was supported by negative correlations between p values and (1) STRP variance in allele size (Spearman's ρ: CEU = −0.26; YRI = −0.24; CHB+JPT = −0.24; $p < 10^{-15}$ in all populations) and (2) expected heterozygosity (CEU = −0.14; YRI = −0.19; CHB+JPT = −0.12; $p < 10^{-15}$ in all populations). Similarly, positive correlations between D' and (1) variance in allele size (CEU = 0.31; YRI = 0.30; CHB+JPT = 0.28; $p < 10^{-15}$ in all populations) and (2) expected heterozygosity (CEU = 0.17; YRI = 0.23; CHB+JPT = 0.11; $p < 10^{-15}$ in all populations) were observed.

Effects of Haplotype Phasing

We assumed that haplotypes were reconstructed without error, as in other large-scale analyses of LD in humans.[13–15,46] Although PHASE is expected to be highly accurate at the physical scale and SNP densities considered here,[46] especially in the CEU and YRI populations for which trios were used, there are reasons to suspect that phasing errors affected observed patterns of LD. First, the posterior probabilities of haplotype pairs provided evidence of uncertainty. Although many haplotype pairs had high posterior probabilities, some regions in some individuals had low probabilities. Use of the haplotype pairs with the highest probabilities ignored that uncertainty. Second, to infer haplotypes involving STRPs is a challenging task. These loci harbor many low-frequency alleles and sometimes mutate in ways that are inconsistent with the stepwise-mutation model assumed in PHASE. A heuristic measure of phasing uncertainty, the (across-individual) average of the highest posterior probabilities of haplotype pairs, was negatively correlated with STRP-SNP D'_{avg} (Spearman's rho: CEU = −0.13; YRI = −0.14; CHB+JPT = −0.31; $p < 10^{-15}$ in all populations), suggesting that LD estimates might have been biased by the phasing process.

We conducted two additional sets of analyses to address the effects of phasing error on the STRP-SNP LD patterns we observed. First, we estimated LD for 31 X-linked regions in males, in which phases were known without error. X-linked and autosomal D' distributions were similar (median: CEU X = 0.54, CEU autosomes = 0.49; YRI X = 0.48, YRI autosomes = 0.43; CHB+JPT X = 0.49, CHB+JPT autosomes = 0.47) (Figure 6), although D' levels were significantly higher on the X chromosome (Mann-Whitney U test: CEU = $p < 10^{-15}$; YRI = $p < 10^{-13}$; CHB+JPT = $p < 10^{-15}$). Because the smaller effective population size of the X chromosome and the lack of recombination in males should lead to greater LD among X-linked loci, the similarity in D' distributions suggests that phasing error was not a major contributor to observed LD patterns. As a further precaution against the effects of phasing error, we also estimated LD with the use of unphased genotypes.

Long-range LD among Unphased Genotypes

We used composite genotypic disequilibrium (CGD)[50–52] to study LD between loci separated by larger distances. We used the high-density human genetic map[43] to define windows with similar recombinational sizes and then calculated CGD between each STRP and every SNP within 2 cM. We also calculated CGD between the SNP closest to each STRP and every SNP within 2 cM. This design allowed us to directly compare the decay of LD in STRP-SNP and SNP-SNP pairs.

Table 1 shows the fractions of significant tests (m_1/m) in different cM intervals, separated by repeat type. STRPs detected significant LD more often than did SNPs across most cM scales and populations. Consistent with patterns observed in the shorter-scale phased analyses, dinucleotides and trinucleotides showed stronger statistical significance than did tetranucleotides. Paired comparisons of STRP-SNP and SNP-SNP tests revealed similar results. STRP-SNP p values were significantly lower than SNP-SNP p values across the 2 cM intervals (Wilcoxon matched-pairs signed-rank test; $p < 10^{-10}$ in all populations).

Discussion

Our study provides the first description of LD between STRPs and SNPs across the human genome. Some patterns are consistent with simple theoretical predictions and results from previous studies.[33] LD decays with recombinational distance, varies among populations, and depends on allele frequencies. These observations mirror empirical patterns seen in genome-wide examinations of SNP-SNP LD.

Our results also highlight the significance of mutational processes for LD. New mutations arise on particular haplotypes and remain perfectly associated with those variants until recombination or mutation disrupts this correlation. Under the infinite-sites model commonly applied to SNPs, only recombination contributes to the decay of D'. In contrast, STRPs routinely undergo recurrent mutation as replication slippage returns alleles to sizes previously realized in the population. Reduced D' levels at STRP-SNP pairs relative to SNP-SNP pairs demonstrate the ability of recurrent mutation to diminish associations among alleles. Lower values of D'_{avg} (LD averaged across STRP alleles) relative to D'_{ind} (LD at individual STRP alleles) indicate that much

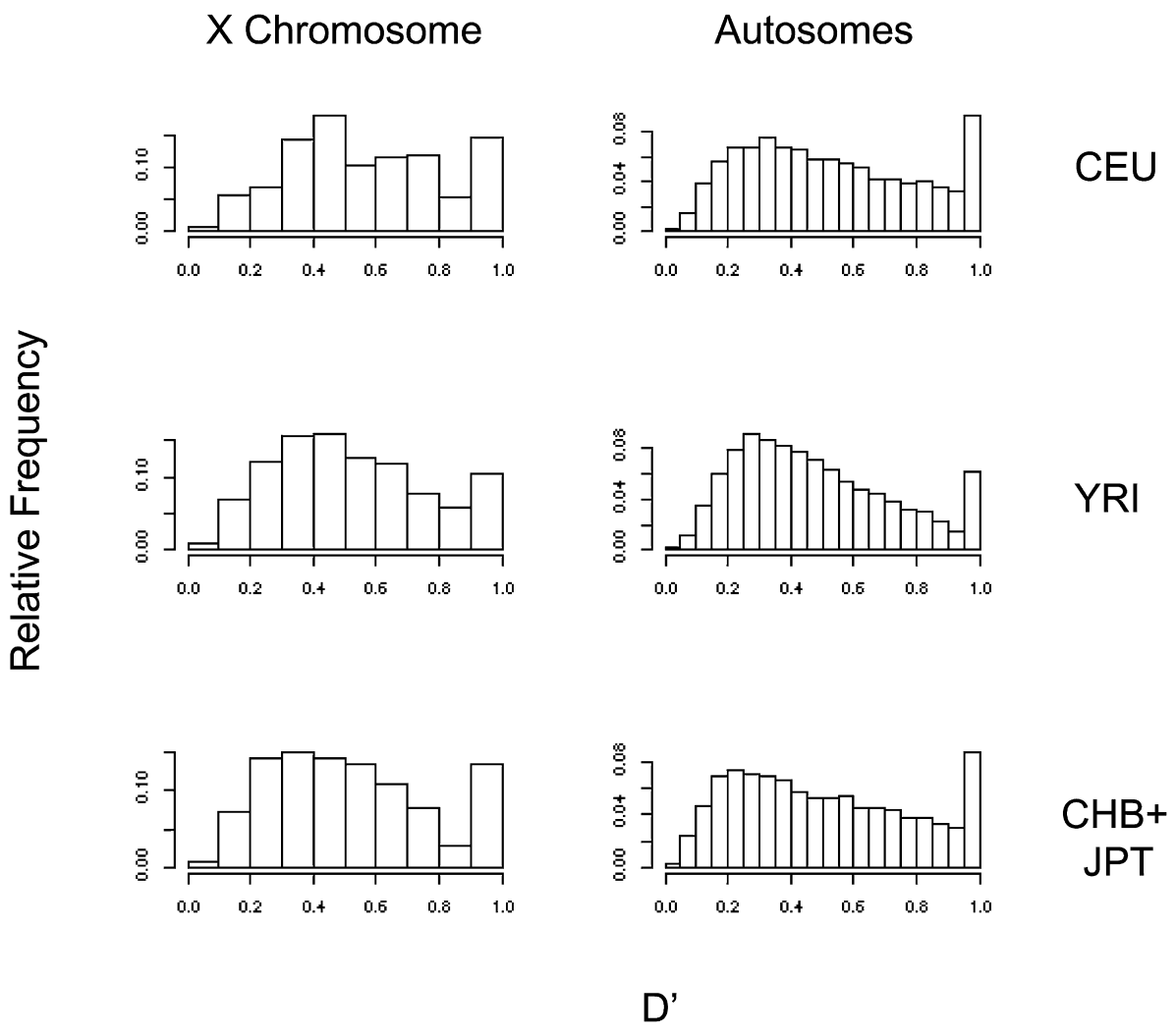

Figure 6. D'_{avg} between STRPs and SNPs on the X Chromosome versus the Autosomes
Loci are within 50 kb.

of the heterogeneity in LD occurs within individual STRPs. STRPs can harbor both alleles that show complete associations with an SNP ($D' = 1$) and alleles that show little to no association with the same SNP. Furthermore, differences in the frequency of recurrent mutation probably contribute to variation in LD among STRP repeat types. Additional complexities in the STRP mutational process, including multistep mutations,[61–64] biases toward expansion or contraction,[64–67] and allele-size-dependent dynamics,[63–65,68–70] probably shape observed patterns of LD as well.

Other aspects of the STRP mutational process are expected to affect LD. In particular, the mutation rate of STRPs exceeds that of individual SNPs by several orders of magnitude. The consequences of this difference for levels of variation can be seen in human populations, which typically segregate many alleles at an STRP[71] and just two alleles at an SNP. The higher mutation rate at STRPs is expected to confer increased power for the detection of significant LD because more branches of the genealogy are marked by mutations.[16–18] Previous comparisons among STRP-SNP LD and SNP-SNP LD in several regions of the human genome[33] and broad comparisons of the extent of SNP-SNP LD[14,15] and STRP-STRP LD[31,72] from different studies provided some support for this prediction. We have demonstrated that STRP-SNP LD is detected with stronger statistical significance than is SNP-SNP LD across the human genome. This difference is observed on both small (50 kb) and large (several cM) scales and in three populations. Several lines of evidence indicate that higher mutation rates underlie the stronger statistical significance of LD at STRPs. First, STRPs show lower p values than do SNPs, and these markers differ in mutation rate by orders of magnitude. Second, STRP repeat types differ in their ability to detect significant LD. Although human-pedigree studies suggest that longer repeats mutate more rapidly,[20] levels of polymorphism in human populations are generally greater at shorter repeats.[58] If differences in levels of variation reflect differences in mutation rates and loci with higher mutation rates offer more power to detect significant LD,[16–18] this could explain our finding that shorter repeats tend to have lower p values. Third, highly variable STRPs (regardless of repeat type) detect more

Please cite this article in press as: Payseur et al., Linkage Disequilibrium between STRPs and SNPs across the Human Genome, The American Journal of Human Genetics (2008), doi:10.1016/j.ajhg.2008.02.018

Table 1. Proportions of Locus Pairs Showing Statistically Significant Composite Genotypic Disequilibrium

		CEU		YRI		CHB+JPT	
cM Interval	Repeat Type	STRP-SNP	SNP-SNP	STRP-SNP	SNP-SNP	STRP-SNP	SNP-SNP
< 0.1	All	0.70	0.68	0.68	0.56	0.73	0.71
< 0.1	Tetranucleotide	0.67	0.69	0.64	0.56	0.69	0.72
< 0.1	Trinucleotide	0.82	0.68	0.81	0.56	0.84	0.70
< 0.1	Dinucleotide	0.76	0.62	0.79	0.55	0.77	0.64
0.1 – < 0.5	All	0.38	0.26	0.39	0.21	0.32	0.25
0.1 – < 0.5	Tetranucleotide	0.35	0.26	0.36	0.21	0.29	0.25
0.1 – < 0.5	Trinucleotide	0.49	0.28	0.49	0.23	0.43	0.25
0.1 – < 0.5	Dinucleotide	0.43	0.24	0.51	0.18	0.35	0.25
0.5 – < 1.0	All	0.22	0.07	0.25	0.07	0.15	0.08
0.5 – < 1.0	Tetranucleotide	0.21	0.06	0.23	0.07	0.14	0.09
0.5 – < 1.0	Trinucleotide	0.25	0.11	0.33	0.08	0.17	0.07
0.5 – < 1.0	Dinucleotide	0.23	0.10	0.34	0.07	0.14	0.05
1.0 – < 1.5	All	0.15	0.07	0.21	0.06	0.11	0.04
1.0 – < 1.5	Tetranucleotide	0.15	0.07	0.21	0.07	0.10	0.03
1.0 – < 1.5	Trinucleotide	0.17	0.07	0.18	0.06	0.16	0.06
1.0 – < 1.5	Dinucleotide	0.14	0.05	0.24	0.06	0.10	0.08
1.5 – < 2.0	All	0.14	0.04	0.19	0.03	0.13	0.05
1.5 – < 2.0	Tetranucleotide	0.13	0.04	0.18	0.04	0.13	0.05
1.5 – < 2.0	Trinucleotide	0.17	0.04	0.21	0.03	0.12	0.01
1.5 – < 2.0	Dinucleotide	0.21	0.10	0.29	0.09	0.10	0.12

Proportions were estimated separately by cM bin and STRP repeat type with a false-discovery-rate approach.

statistically significant LD. Finally, multi-SNP haplotypes and STRPs with similar levels of variation show more similar abilities to detect significant LD.

LD at STRPs reflects a balance between two mutational forces with opposing consequences. As STRP mutations accumulate, the fraction that is recurrent reduces LD (as evidenced by D') and the proportion that is new increases the power to detect LD. Consequently, an improved understanding of these proportions and other details of STRP mutational models will be crucial to our ability to explain observed patterns of STRP LD. We might expect, for example, that STRPs that mutate in larger steps will produce a higher fraction of unique alleles and better capture LD. In addition to providing improved predictions for STRPs, further modeling of the effects of mutational processes would be relevant to other classes of molecular variation, including CpG sites, where multiple mutations can segregate.

Our study also highlights challenges associated with the measurement of LD. First, our results emphasize the difference between measures of the intensity of LD, such as D', and tests of the null hypothesis of no association. The first measure describes the form of the association between a pair of loci, and the second measure describes the statistical significance of an association. Although these measures are correlated, they can differ. The relative usefulness of LD at STRPs and SNPs for specific applications therefore depends on which measure is most relevant. Furthermore, better descriptors of LD are needed for loci with many alleles. The commonly used metric of R^2, which features a theoretical relationship to the population-recombination parameter at equilibrium,[73] is undefined for loci with more than two alleles, and it can be difficult to compare D'_{avg} between loci with different numbers of alleles[48] (but see [74]).

Because identification of genetic variants that cause disease by association mapping requires detailed knowledge of LD, our study provides information on the relative merits of STRPs and SNPs for these efforts. SNPs offer several advantages over STRPs in the context of association mapping. The higher density of SNPs across the genome improves the capacity for fine-scale mapping. Modeling is simplified by the assumption that recombination is the primary force that causes LD to decay, an assumption that cannot be justified for STRPs. Finally, advances in genotyping technologies have made routine and affordable the task of surveying very large numbers of SNPs in many individuals. These factors suggest that SNPs will remain the marker of choice for association mapping.

Our results indicate that STRPs can provide an additional useful resource for association mapping. STRPs might offer greater power to detect LD than do individual SNPs.[33] Gains in the strength of statistical significance are most striking for dinucleotides and trinucleotides, suggesting that these markers might be particularly useful for association mapping. The genomic density of these repeats combined with the ability of STRPs to detect LD over large distances suggest that STRPs could be useful on this intermediate physical scale.[75] Genome-wide association studies using tens of thousands of STRPs have begun to appear.[76,77]

The relative performance of STRPs and SNPs in association mapping will also depend on the frequencies of disease variants. Marker alleles achieve maximal power for detecting associations when disease alleles are at similar frequencies.[78] As a result, STRPs have the potential to find rare disease variants that common SNPs will miss.[17,33] With growing evidence that rare alleles contribute to common

diseases,[79–81] this possibility deserves attention. It seems likely that we have under-estimated the relative performance of STRPs for association mapping by measuring LD in datasets that feature strong biases against rare SNPs. Because the STRPs were also chosen to be highly informative, their frequency spectra might have been biased as well. Furthermore, additional power conferred by low-frequency alleles at STRPs might have been eroded by our sample sizes, which were much smaller than those used in typical association studies.

In addition to using STRPs and SNPs separately, the contrasting properties of these markers suggest that methods that consider STRP-SNP haplotypes (or unphased multi-locus genotypes) might be useful for genome-wide association studies. Indeed, STRPs and SNPs are often combined to dissect associations between genotype and phenotype on a fine scale. Additionally, researchers could use the long-range LD at STRPs to reduce the number of initial association tests, following up candidate regions with dense SNP genotyping. The performance of these mixed marker strategies needs to be evaluated.

Patterns of LD at STRPs and SNPs also provide necessary background for integrating variation at these two marker classes for population genetic inference. Empirical and theoretical studies show that combining linked STRP and SNP variation provides novel insights into population structure, demographic history, and selection operating on different timescales.[39,40,82–86] Harnessing of the full power of molecular diversity for the understanding of human history will require the joint consideration of variation at STRPs and SNPs.

Acknowledgments

We thank Gonçalo Abecasis and Karl Broman for advice during the course of this study. We thank Daniel Schaid and Jason Sinwell for providing software for composite LD analyses. We thank Miron Livny and Zach Miller for access to computers and assistance with Condor high-throughput computing software. Aida M. Andrés provided helpful comments on the manuscript. This research was supported by a Medical Education and Research Committee New Investigator Award (School of Medicine and Public Health, University of Wisconsin) to B.A.P. and by funding from the National Heart, Lung, and Blood Institute (NHLBI) for the Mammalian Genotyping Service to J.L.W.

Received: October 21, 2007
Revised: January 6, 2008
Accepted: February 29, 2008
Published online: April 17, 2008

Web Resources

The URLs for data presented herein are as follows:

Hapmap, www.hapmap.org
Marshfield Mammalian Genotyping Service, http://research.marshfieldclinic.org/genetics/home/index.asp
UCSC Genome Browser, www.genome.ucsc.edu

References

1. Nordborg, M., and Tavare, S. (2002). Linkage disequilibrium: What history has to tell us. Trends Genet. *18*, 83–90.
2. Pritchard, J.K., and Przeworski, M. (2001). Linkage disequilibrium in humans: Models and data. Am. J. Hum. Genet. *69*, 1–14.
3. Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z.P., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J., et al. (2002). Detecting recent positive selection in the human genome from haplotype structure. Nature *419*, 832–837.
4. McVean, G.A., Myers, S.R., Hunt, S., Deloukas, P., Bentley, D.R., and Donnelly, P. (2004). The fine-scale structure of recombination rate variation in the human genome. Science *304*, 581–584.
5. Myers, S., Bottolo, L., Freeman, C., McVean, G., and Donnelly, P. (2005). A fine-scale map of recombination rates and hotspots across the human genome. Science *310*, 321–324.
6. Ptak, S.E., Voelpel, K., and Przeworski, M. (2004). Insights into recombination from patterns of linkage disequilibrium in humans. Genetics *167*, 387–397.
7. Eberle, M.A., Rieder, M.J., Kruglyak, L., and Nickerson, D.A. (2006). Allele frequency matching between SNPs reveals an excess of linkage disequilibrium in genic regions of the human genome. Plos Genetics *2*, 1319–1327.
8. Voight, B.F., Kudaravalli, S., Wen, X.Q., and Pritchard, J.K. (2006). A map of recent positive selection in the human genome. PLoS Biol. *4*, 446–458.
9. Hirschhorn, J.N., and Daly, M.J. (2005). Genome-wide association studies for common diseases and complex traits. Nat. Rev. Genet. *6*, 95–108.
10. Jorde, L.B. (2000). Linkage disequilibrium and the search for complex disease genes. Genome Res. *10*, 1435–1444.
11. Kruglyak, L. (1999). Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat. Genet. *22*, 139–144.
12. Risch, N., and Merikangas, K. (1996). The future of genetic studies of complex human diseases. Science *273*, 1516–1517.
13. The International HapMap Consortium (2007). A second generation human haplotype map of over 3.1 million SNPs. Nature *449*, 851–861.
14. The International HapMap Consortium (2005). A haplotype map of the human genome. Nature *437*, 1299–1320.
15. Hinds, D.A., Stuve, L.L., Nilsen, G.B., Halperin, E., Eskin, E., Ballinger, D.G., Frazer, K.A., and Cox, D.R. (2005). Whole-genome patterns of common DNA variation in three human populations. Science *307*, 1072–1079.
16. Chapman, N.H., and Wijsman, E.M. (1997). Optimal marker characteristics for genome screens using linkage disequilibrium tests. Am. J. Hum. Genet. *61*, A271–a271.
17. Ohashi, J., and Tokunaga, K. (2003). Power of genome-wide linkage disequilibrium testing by using microsatellite markers. J. Hum. Genet. *48*, 487–491.
18. Xiong, M., and Jin, L. (1999). Comparison of the power and accuracy of biallelic and microsatellite markers in population-based gene-mapping methods. Am. J. Hum. Genet. *64*, 629–640.
19. Ellegren, H. (2004). Microsatellites: Simple sequences with complex evolution. Nat. Rev. Genet. *5*, 435–445.
20. Weber, J.L., and Wong, C. (1993). Mutation of human short tandem repeats. Hum. Mol. Genet. *2*, 1123–1128.

21. Levinson, G., and Gutman, G.A. (1987). Slipped-strand mispairing: A major mechanism for DNA sequence evolution. Mol. Biol. Evol. *4*, 203–221.
22. Nachman, M.W., and Crowell, S.L. (2000). Estimate of the mutation rate per nucleotide in humans. Genetics *156*, 297–304.
23. Rosenberg, N.A., Li, L.M., Ward, R., and Pritchard, J.K. (2003). Informativeness of genetic markers for inference of ancestry. Am. J. Hum. Genet. *73*, 1402–1422.
24. Abecasis, G.R., Noguchi, E., Heinzmann, A., Traherne, J.A., Bhattacharyya, S., Leaves, N.I., Anderson, G.G., Zhang, Y.M., Lench, N.J., Carey, A., et al. (2001). Extent and distribution of linkage disequilibrium in three genomic regions. Am. J. Hum. Genet. *68*, 191–197.
25. Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., et al. (2002). The structure of haplotype blocks in the human genome. Science *296*, 2225–2229.
26. Reich, D.E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P.C., Richter, D.J., Lavery, T., Kouyoumjian, R., Farhadian, S.F., Ward, R., and Lander, E.S. (2001). Linkage disequilibrium in the human genome. Nature *411*, 199–204.
27. Clark, A.G., Nielsen, R., Signorovitch, J., Matise, T.C., Glanowski, S., Heil, J., Winn-Deen, E.S., Holden, A.L., and Lai, E. (2003). Linkage disequilibrium and inference of ancestral recombination in 538 single-nucleotide polymorphism clusters across the human genome. Am. J. Hum. Genet. *73*, 285–300.
28. Daly, M.J., Rioux, J.D., Schaffner, S.E., Hudson, T.J., and Lander, E.S. (2001). High-resolution haplotype structure in the human genome. Nat. Genet. *29*, 229–232.
29. Patil, N., Berno, A.J., Hinds, D.A., Barrett, W.A., Doshi, J.M., Hacker, C.R., Kautzer, C.R., Lee, D.H., Marjoribanks, C., McDonough, D.P., et al. (2001). Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science *294*, 1719–1723.
30. Conrad, D.F., Jakobsson, M., Coop, G., Wen, X., Wall, J.D., Rosenberg, N.A., and Pritchard, J.K. (2006). A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. Nat. Genet. *38*, 1251–1260.
31. Huttley, G.A., Smith, M.W., Carrington, M., and O'Brien, S.J. (1999). A scan for linkage disequilibrium across the human genome. Genetics *152*, 1711–1722.
32. Schulze, T.G., Chen, Y.S., Akula, N., Hennessy, K., Badner, J.A., McInnis, M.G., DePaulo, J.R., Schumacher, J., Cichon, S., Propping, P., et al. (2002). Can long-range microsatellite data be used to predict short-range linkage disequilibrium? Hum. Mol. Genet. *11*, 1363–1372.
33. Varilo, T., Paunio, T., Parker, A., Perola, M., Meyer, J., Terwilliger, J.D., and Peltonen, L. (2003). The interval of linkage disequilibrium (LD) detected with microsatellite and SNP markers in chromosomes of Finnish populations with different histories. Hum. Mol. Genet. *12*, 51–59.
34. Bahram, S., and Inoko, H. (2007). Microsatellite markers for genome-wide association studies. Nat. Rev. Genet. *8*.
35. Jorgenson, E., and Witte, J.S. (2007). Reply: Microsatellite markers for genome-wide association studies. Nat. Rev. Genet. *8*.
36. Conrad, D.F., Andrews, T.D., Carter, N.P., Hurles, M.E., and Pritchard, J.K. (2006). A high-resolution survey of deletion polymorphism in the human genome. Nat. Genet. *38*, 75–81.
37. Hinds, D.A., Kloek, A.P., Jen, M., Chen, X.Y., and Frazer, K.A. (2006). Common deletions and SNPs are in linkage disequilibrium in the human genome. Nat. Genet. *38*, 82–85.
38. Locke, D.P., Sharp, A.J., McCarroll, S.A., McGrath, S.D., Newman, T.L., Cheng, Z., Schwartz, S., Albertson, D.G., Pinkel, D., Altshuler, D.M., and Eichler, E.E. (2006). Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome. Am. J. Hum. Genet. *79*, 275–290.
39. Mountain, J.L., Knight, A., Jobin, M., Gignoux, C., Miller, A., Lin, A.A., and Underhill, P.A. (2002). SNPSTRs: Empirically derived, rapidly typed, autosomal haplotypes for inference of population history and mutational processes. Genome Res. *12*, 1766–1772.
40. Payseur, B.A., and Cutter, A.D. (2006). Integrating patterns of polymorphism at SNPs and STRs. Trends Genet. *22*, 424–429.
41. Ghebranious, N., Vaske, D., Yu, A.D., Zhao, C.F., Marth, G., and Weber, J.L. (2003). STRP Screening Sets for the human genome at 5 cM density. BMC Genomics *4*, 6.
42. Weber, J.L., and Broman, K. (2001). Genotyping for human whole-genome scans: Past, present and future. Adv. Genet. *42*, 77–96.
43. Kong, A., Gudbjartsson, D.F., Sainz, J., Jonsdottir, G.M., Gudjonsson, S.A., Richardsson, B., Sigurdardottir, S., Barnard, J., Hallbeck, B., Masson, G., et al. (2002). A high-resolution recombination map of the human genome. Nat. Genet. *31*, 241–247.
44. Stephens, M., and Scheet, P. (2005). Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. Am. J. Hum. Genet. *76*, 449–462.
45. Stephens, M., Smith, N.J., and Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. *68*, 978–989.
46. Marchini, J., Cutler, D., Patterson, N., Stephens, M., Eskin, E., Halperin, E., Lin, S., Qin, Z.S., Munro, H.M., Abecasis, G.R., and Donnelly, P. (2006). A comparison of phasing algorithms for trios and unrelated individuals. Am. J. Hum. Genet. *78*, 437–450.
47. Lewontin, R.C. (1964). The interaction of selection and linkage. I. General considerations; heterotic models. Genetics *49*, 49–67.
48. Hedrick, P.W. (1987). Gametic disequilibrium measures: Proceed with caution. Genetics *117*, 331–341.
49. Abecasis, G.R., and Cookson, W.O.C. (2000). GOLD - Graphical Overview of Linkage Disequilibrium. Bioinformatics *16*, 182–183.
50. Schaid, D.J. (2004). Linkage disequilibrium testing when linkage phase is unknown. Genetics *166*, 505–512.
51. Weir, B.S. (1979). Inferences about linkage disequilibrium. Biometrics *35*, 235–254.
52. Weir, B.S. (1996). Genetic Data Analysis II (Sunderland, MA: Sinauer Associates).
53. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B *57*, 289–300.
54. Storey, J.D. (2002). A direct approach to false discovery rates. J Roy Stat Soc B *64*, 479–498.
55. Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. USA *100*, 9440–9445.
56. Zapata, C., and Alvarez, G. (1997). On Fisher's exact test for detecting gametic disequilibrium between DNA polymorphisms. Ann. Hum. Genet. *61*, 71–77.

57. Dieringer, D., and Schlotterer, C. (2003). MICROSATELLITE ANALYSER (MSA): A platform independent analysis tool for large microsatellite data sets. Mol. Ecol. Notes *3*, 167–169.
58. Chakraborty, R., Kimmel, M., Stivers, D.N., Davison, L.J., and Deka, R. (1997). Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. Proc. Natl. Acad. Sci. USA *94*, 1041–1046.
59. Clark, A.G., Hubisz, M.J., Bustamante, C.D., Williamson, S.H., and Nielsen, R. (2005). Ascertainment bias in studies of human genome-wide polymorphism. Genome Res. *15*, 1496–1502.
60. Pe'er, I., Chretien, Y.R., de Bakker, P.I.W., Barrett, J.C., Daly, M.J., and Altshuler, D.M. (2006). Biases and reconciliation in estimates of linkage disequilibrium in the human genome. Am. J. Hum. Genet. *78*, 588–603.
61. Di Rienzo, A., Peterson, A.C., Garza, J.C., Valdes, A.M., Slatkin, M., and Freimer, N.B. (1994). Mutational processes of simple-sequence repeat loci in human populations. Proc. Natl. Acad. Sci. USA *91*, 3166–3170.
62. Ellegren, H. (2000). Heterogeneous mutation processes in human microsatellite DNA sequences. Nat. Genet. *24*, 400–402.
63. Huang, Q.Y., Xu, F.H., Shen, H., Deng, H.Y., Liu, Y.J., Liu, Y.Z., Li, J.L., Recker, R.R., and Deng, H.W. (2002). Mutation patterns at dinucleotide microsatellite loci in humans. Am. J. Hum. Genet. *70*, 625–634.
64. Xu, X., Peng, M., and Fang, Z. (2000). The direction of microsatellite mutations is dependent upon allele length. Nat. Genet. *24*, 396–399.
65. Amos, W., and Rubinstzein, D.C. (1996). Microsatellites are subject to directional evolution. Nat. Genet. *12*, 13–14.
66. Cooper, G., Burroughs, N.J., Rand, D.A., Rubinsztein, D.C., and Amos, W. (1999). Markov Chain Monte Carlo analysis of human Y-chromosome microsatellites provides evidence of biased mutation. Proc. Natl. Acad. Sci. USA *96*, 11916–11921.
67. Kayser, M., Roewer, L., Hedman, M., Henke, L., Henke, J., Brauer, S., Kruger, C., Krawczak, M., Nagy, M., Dobosz, T., et al. (2000). Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. Am. J. Hum. Genet. *66*, 1580–1588.
68. Brinkmann, B., Klintschar, M., Neuhuber, F., Huhne, J., and Rolf, B. (1998). Mutation rate in human microsatellites: Influence of the structure and length of the tandem repeat. Am. J. Hum. Genet. *62*, 1408–1415.
69. Dupuy, B.M., Stenersen, M., Egeland, T., and Olaisen, B. (2004). Y-chromosomal microsatellite mutation rates: Differences in mutation rate between and within loci. Hum. Mutat. *23*, 117–124.
70. Holtkemper, U., Rolf, B., Hohoff, C., Forster, P., and Brinkmannn, B. (2001). Mutation rates at two human Y-chromosomal microsatellite loci using small pool PCR techniques. Hum. Mol. Genet. *10*, 629–633.
71. Rosenberg, N.A., Pritchard, J.K., Weber, J.L., Cann, H.M., Kidd, K.K., Zhivotovsky, L.A., and Feldman, M.W. (2002). Genetic structure of human populations. Science *298*, 2381–2385.
72. Bataillon, T., Mailund, T., Thorlacius, S., Steingrimsson, E., Rafnar, T., Halldorsson, M.M., Calian, V., and Schierup, M.H. (2006). The effective size of the Icelandic population and the prospects for LD mapping: Inference from unphased microsatellite markers. Eur. J. Hum. Genet. *14*, 1044–1053.
73. Hill, W.G., and Robertson, A. (1968). Linkage disequilibrium in finite populations. Theor. Appl. Genet. *38*, 226–231.
74. Zapata, C. (2000). The D′ measure of overall gametic disequilibrium between pairs of multiallelic loci. Evolution Int. J. Org. Evolution *54*, 1809–1812.
75. Weber, J.L. (2006). Clinical applications of genome polymorphism scans. Biol Direct *1*, 16.
76. Tamiya, G., Shinya, M., Imanishi, T., Ikuta, T., Makino, S., Okamoto, K., Furugaki, K., Matsumoto, T., Mano, S., Ando, S., et al. (2005). Whole genome association study of rheumatoid arthritis using 27 039 microsatellites. Hum. Mol. Genet. *14*, 2305–2321.
77. Yatsu, K., Hirawa, N., Ogawa, M., Soma, M., Hata, A., Nakao, K., Ueshima, H., Ogihara, T., Tomoike, H., Kimura, A., et al. (2006). Genome-wide association mapping for essential hypertension with high-density microsatellite markers. J. Hypertens. *24*, 55–56.
78. Zondervan, K.T., and Cardon, L.R. (2004). The complex interplay among factors that influence allelic association. Nat. Rev. Genet. *5*, 89–100.
79. Cohen, J.C., Kiss, R.S., Pertsemlidis, A., Marcel, Y.L., McPherson, R., and Hobbs, H.H. (2004). Multiple rare Alleles contribute to low plasma levels of HDL cholesterol. Science *305*, 869–872.
80. Pritchard, J.K. (2001). Are rare variants responsible for susceptibility to complex diseases? Am. J. Hum. Genet. *69*, 124–137.
81. Romeo, S., Pennacchio, L.A., Fu, Y.X., Boerwinkle, E., Tybjaerg-Hansen, A., Hobbs, H.H., and Cohen, J.C. (2007). Population-based resequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. Nat. Genet. *39*, 513–516.
82. de Knijff, P. (2000). Messages through bottlenecks: On the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. Am. J. Hum. Genet. *67*, 1055–1061.
83. Hey, J., Won, Y.J., Sivasundar, A., Nielsen, R., and Markert, J.A. (2004). Using nuclear haplotypes with microsatellites to study gene flow between recently separated Cichlid species. Mol. Ecol. *13*, 909–919.
84. Ramakrishnan, U., and Mountain, J.L. (2004). Precision and accuracy of divergence time estimates from STR and SNPSTR variation. Mol. Biol. Evol. *21*, 1960–1971.
85. Tishkoff, S.A., Varkonyi, R., Cahinhinan, N., Abbes, S., Argyropoulos, G., Destro-Bisol, G., Drousiotou, A., Dangerfield, B., Lefranc, G., Loiselet, J., et al. (2001). Haplotype diversity and linkage disequilibrium at human G6PD: Recent origin of alleles that confer malarial resistance. Science *293*, 455–462.
86. Zegura, S.L., Karafet, T.M., Zhivotovsky, L.A., and Hammer, M.F. (2004). High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. Mol. Biol. Evol. *21*, 164–175.