

Research article

Gene conversion homogenizes the CMT1A paralogous repeats

Matthew E Hurles

Address: McDonald Institute for Archaeological Research, University of Cambridge, Cambridge, UK

E-mail: meh32@cam.ac.uk

Published: 11 December 2001

Received: 18 October 2001

BMC Genomics 2001, 2:11

Accepted: 11 December 2001

This article is available from: <http://www.biomedcentral.com/1471-2164/2/11>

© 2001 Hurles; licensee BioMed Central Ltd. Verbatim copying and redistribution of this article are permitted in any medium for any non-commercial purpose, provided this notice is preserved along with the article's original URL. For commercial use, contact info@biomedcentral.com

Abstract

Background: Non-allelic homologous recombination between paralogous repeats is increasingly being recognized as a major mechanism causing both pathogenic microdeletions and duplications, and structural polymorphism in the human genome. It has recently been shown empirically that gene conversion can homogenize such repeats, resulting in longer stretches of absolute identity that may increase the rate of non-allelic homologous recombination.

Results: Here, a statistical test to detect gene conversion between pairs of non-coding sequences is presented. It is shown that the 24 kb Charcot-Marie-Tooth type 1A paralogous repeats (CMT1A-REPs) exhibit the imprint of gene conversion processes whilst control orthologous sequences do not. In addition, Monte Carlo simulations of the evolutionary divergence of the CMT1A-REPs, incorporating two alternative models for gene conversion, generate repeats that are statistically indistinguishable from the observed repeats. Bounds are placed on the rate of these conversion processes, with central values of 1.3×10^{-4} and 5.1×10^{-5} per generation for the alternative models.

Conclusions: This evidence presented here suggests that gene conversion may have played an important role in the evolution of the CMT1A-REP paralogous repeats. The rates of these processes are such that it is probable that homogenized CMT1A-REPs are polymorphic within modern populations. Gene conversion processes are similarly likely to play an important role in the evolution of other segmental duplications and may influence the rate of non-allelic homologous recombination between them.

Background

There is a rapidly growing literature describing pathogenic rearrangements caused by illegitimate recombination between non-allelic homologous sequences [1–3]. These paralogous repeats can cause inversions which disrupt coding sequences, and microdeletions and microduplications that result in pathogenic changes in the copy number of intervening genes.

The peripheral neuropathy Charcot-Marie-Tooth disease type 1A (CMT1A) is caused by a microduplication of a 1.5 Mb region on chromosome 17 that lies between 24 kb

paralogous direct repeats known as CMT1A-REPs [4]. The reciprocal product, a microdeletion, causes hereditary neuropathy with liability to pressure palsies (HNPP) [5]. Similarly the SMS-REP direct paralogous repeats on chromosome 17 cause both pathogenic microdeletions (causing Smith-Magenis Syndrome) and microduplications [6]. Inverted paralogous repeats are responsible for the inversion of a portion of the factor VIII gene that causes hemophilia [7].

Whilst the focus of such rearrangements typically falls on those that result in disease, recombination between paral-

ogous repeats also causes structural polymorphism within the human genome. There are a number of polymorphic inversions known to be mediated by inverted paralogous repeats, including paracentric inversions of Yp [8], Xq [9] and 8p [10]. One of the orientations of both the Yp and 8p polymorphic inversions has been shown to predispose to further, pathogenic, rearrangements. The Yp inversion polymorphism predisposes carriers to XY translocations between the now similarly oriented PRKX and PRKY genes [11].

A recent study has shown that non-allelic homologous recombination between two human endogenous retroviral (HERV) proviral sequences on Yq results in deletion of the AZFa locus, causing male infertility [12]. This study also demonstrated that gene conversion events between these paralogous repeats (~94% homology) have homogenized them at least twice during recent human evolution, and that the presence of homogenized tracts within these repeats is polymorphic in modern populations. These independent conversion events, detailed at the sequence level [12], increase the maximum length of absolute identity between the repeats by a factor of five. Given the primacy of the length of absolute identity in determining the rate of recombination it was hypothesized that the resulting homogenized repeats might predispose carriers to AZFa microdeletions. As such, these events may create homogenized repeat premutations (HRP) analogous to the pre-mutations of triplet repeat disorders and the chromosomal inversions detailed above. Consequently it is of interest to determine whether such homogenizing events occur at other paralogous repeats, and whether homogenized tracts within paralogous repeats are polymorphic in extant populations.

In principle, without knowing anything *a priori* about the rate of gene conversion, finding homogenized repeats in the general population might require a large scale comparative sequencing effort that would be complicated by the repeated nature of the template. Alternatively, techniques of sperm small-pool PCR (SP-PCR) could be used to detect *de novo* conversion events [13]; however, homogenization could occur at many possible sites within most paralogous repeats, making detection problematic. An alternative to these lab-based methods is to use available sequence data to explore whether gene conversion processes have played a role in the evolution of these repeats.

There are many analytical methods that could detect the imprint of gene conversion in accurate genomic sequences of entire REPs. In practice, the absence of both orthologous sequences and multiple human sequences reduces the options considerably. Gene conversion events can be expected to produce longer stretches of identity between paralogous sequences than could be expected by chance.

Sawyer devised a permutation test that determines whether variant sites between aligned coding sequences are distributed randomly along their length [14]. Here, modifications to Sawyer's statistical test are devised for non-coding sequences, in other words taking account of the increased mutability of CpGs, and confirm that gene conversion has been operating on the CMT1A-REPs. Evolutionary simulations of the divergence of these paralogous repeats under different models are compared to the observed sequences to determine the rates and mechanisms of the gene conversion processes that are likely to have operated during their diversification from initially identical duplicates.

Results

Full sequences of the two CMT1A-REPs have previously been determined [15]. The CMT1A-REPs are ~99% non-coding [16]. These sequences were aligned and sites varying between the copies identified. 257 base substitution variants were identified within a 23758 bp alignment. Stretches of absolute identity within the alignment that lie between variant sites are known as 'fragments'. Sawyer's gene conversion test, when applied to pairs of sequences, measures the sum of the squared total lengths of these fragments (SSUF); Sawyer originally qualified such fragments as being 'uncondensed'. If gene conversion has been operating, fragment lengths should be more varied and SSUF would be larger than if variant sites were distributed at random along the sequence. Therefore the same number of variant sites is then randomly permuted along the sequence many times and the SSUF for each replication calculated, and compared to the observed value. For levels of divergence of ~1% Sawyer's test has been shown to have at least equivalent power to a number of other methods for detecting gene conversion [17].

Sawyer's test was adapted here to take account of the increased mutability of CpG dinucleotides, to avoid false positives from non-random distributions of CpGs throughout an alignment (for details see Materials and Methods). Table 1 shows the results obtained by applying this 'Sawyer-CpG' test to the aligned CMT1A-REP sequences and comparing it to a control alignment of similarly sized orthologous sequences from a non-coding region of chromosome 12 in humans and chimps. The paralogous repeats show evidence of gene conversion whereas the orthologous sequences do not. This indicates that the pattern of identity between CMT1A repeats does not result from some as yet uncharacterized source of rate heterogeneity within hominoid sequence evolution. CMT1A-REP sequences were also drawn from recently finished BAC sequences and aligned, and again these showed a SSUF value significantly larger than the random permutations of the Sawyer-CpG test ($p = 0.011$), thus indicating

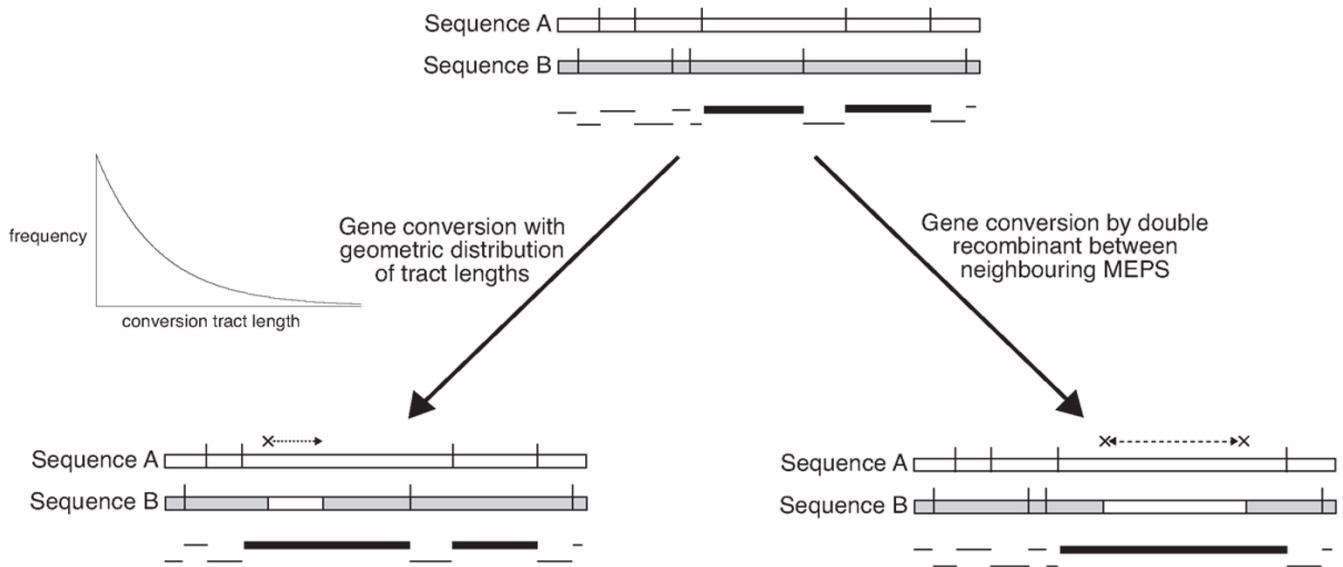


Figure 1
Schematic diagram of two models of gene conversion. Two paralogous sequences are aligned and variant sites between them are shown as vertical lines extending above each sequence. Beneath the alignments are shown the length of the fragments between each variant site. Fragments long enough to be considered a MEPS are thicker than those that are not. Dashed lines with arrows indicate the extent of gene conversion, with the initiation point(s) shown by an 'X'. Note the greater variation in fragment length after gene conversion.

Table 1: Comparison of paralogous CMT1A-REP sequence alignment with orthologous control sequences, together with results of the Sawyer-CpG permutation test for detecting the imprint of gene conversion. Significance at the 95% level is indicated in bold. Comparison of orthologous and paralogous sequence alignments

Sequences	Length (bp)	Variant sites (at CpGs)	CpG frequency	SSUF	P value
CMT1A-REPs	23758	257 (73)	0.0098	5220559	0.008
Chr.12 orthologs	23584	275 (62)	0.0067	4405300	0.058

that the results are not due to errors within the cosmid sequencing data.

It is possible that the presence of dispersed repeats (e.g. Alus and LINES) within, and not the paralogy between, CMT1A repeats drives these conversion processes. However, variant sites between the CMT1A-REPs are distributed equally between dispersed repeat and 'unique' sequences within the paralogous repeats (χ^2 test, 1 d.f., $p = 0.53$). Thus it is highly unlikely that the patterns of identity identified are due to conversion events from dispersed repeats that are repeated outside the CMT1A-REP sequences.

In order to investigate the dynamics of gene conversion, Monte-Carlo simulations of the evolution of CMT1A-

REPs were performed. All non-CpG sites were considered equally mutable, as were all CpG sites, although at a higher rate than non-CpG positions. As expected, running these simulations in the absence of gene conversion produced SSUF values that were significantly lower than that observed between CMT1A-REPs ($p < 0.01$).

Introducing gene conversion into these simulations requires a model for its mechanism. Studies of meiotic homologous recombination in mammals have indicated the importance of the length of absolute identity in determining the recombination rate. These studies have led to the concept of a minimal efficient processing segment (MEPS), thought to be roughly 200 bp in length [18]. Gene conversion within mammalian genomes has been

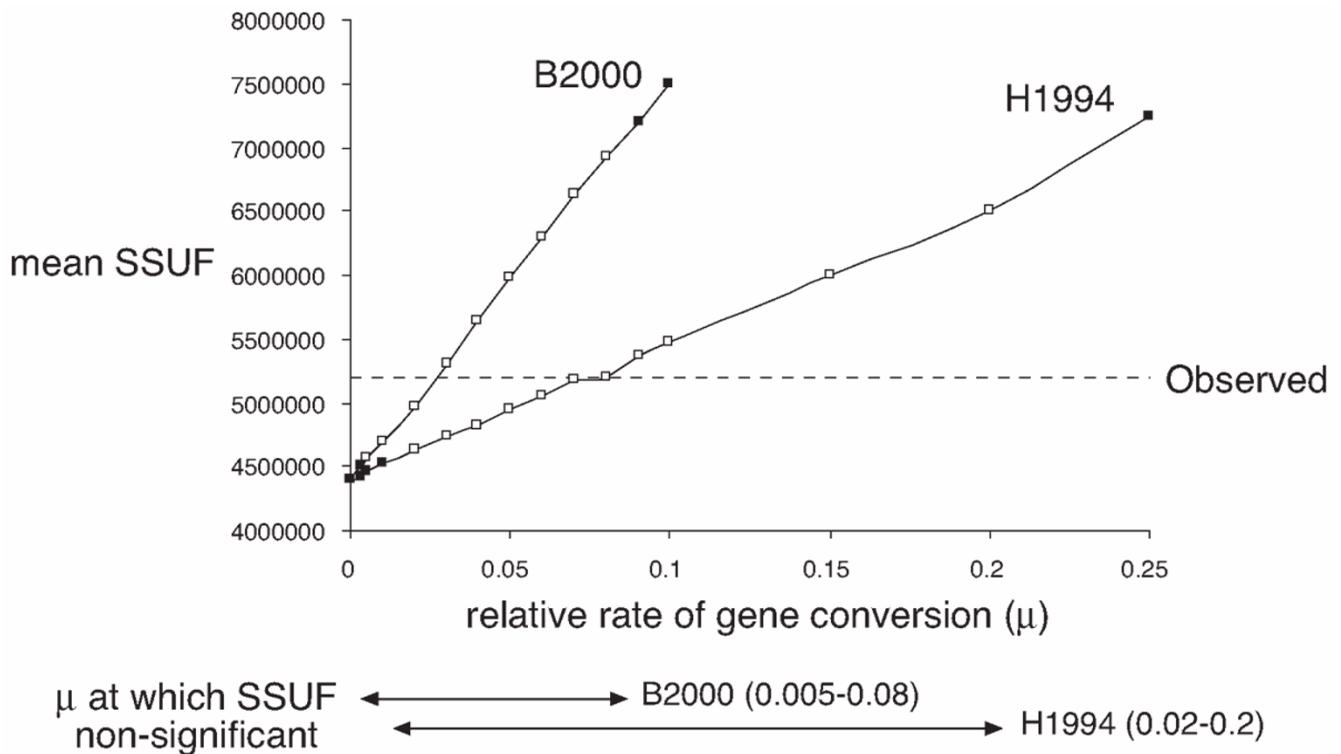


Figure 2
 Impact of simulated gene conversion on the sum of squared fragment lengths (SSUF) between aligned CMT1A-REPs. In order to obtain plausible ranges for rate of the gene conversion processes between CMT1A-REPs, the relative rate of gene conversion is varied under two models and is plotted against the mean of the sum of squared fragment lengths (SSUF) from 1000 simulated alignments. Significance is assessed against the SSUF from the observed alignment. Filled squares indicate rates at which simulated datasets are significantly different from the observed one, at the 95% level. The range of rates at which simulated datasets are not significantly different from the observed one is shown beneath the X-axis for both models.

identified on a locus-by-locus basis, rather than on a genome-wide scale, and little is known about its molecular mechanism. Whilst most theoretical considerations of gene conversion models have focused on a geometric distribution of gene conversion length derived from studies in *Drosophila* [19,20], the only empirical evidence for gene conversion between paralogous sequences to date does not fit this model [12]. This latter study supports a model of double recombinant events between sister chromatids, where the crossovers lie in MEPSs.

Consequently two different gene conversion models were introduced into the simulations individually. The first model assumes a random initiation of gene conversion along the sequence followed by branch migration that results in a geometric distribution of converted sequence, with a mean length of 353 bp [19], and is abbreviated as H1994. The second model is based on the empirical evidence described above [12] and converts sequence that lies between neighboring MEPSs, where the length of a MEPS is assumed to be greater than 200 bp of sequence

identity, and is abbreviated as B2000. These models are displayed schematically in figure 1. The rate of gene conversion relative to the rate of accumulation of base substitution variants is incorporated into the simulations via the parameter μ . If $\mu = 0.01$, one gene conversion event will occur, on average, per 100 base substitutions between the paralogous repeats.

Figure 2 shows the results of repeated simulations under the two models described above and using various values of μ . It can be seen that both gene conversion models are capable of generating simulated CMT1A-REP datasets that are not significantly different from the observed alignment. Upper and lower bounds on the relative rate of gene conversion can be determined for each of the two models. The range of conversion rates (μ) at which the B2000 model produces sequences that are not significantly different from the observed ones is 0.005–0.08, while for the H1994 model this range is 0.02–0.2.

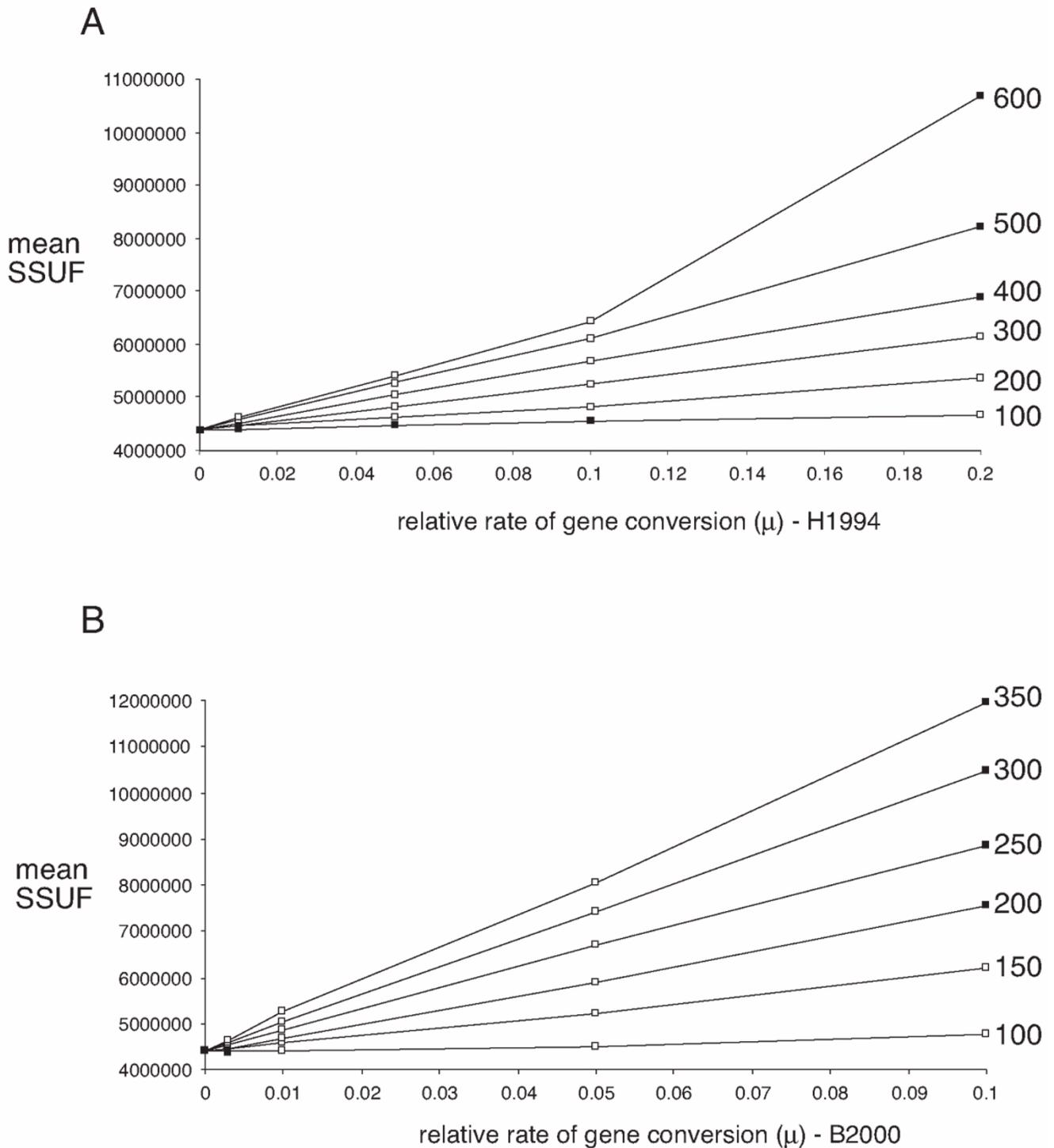


Figure 3
 The effect of varying parameters of the gene conversion models on sum of squared fragment length (SSUF). a) The impact of varying the mean of the geometric distribution of conversion tract length for the H1994 model. b) The impact of varying the length of a MEPS on the B2000 model. Relative gene conversion rate is plotted against the mean SSUF value from 1000 simulated alignments. Numbers next to the lines indicate the size of the parameter in question, in base pairs. Significance is assessed against the SSUF from the observed alignment. Filled squares indicate rates at which simulated datasets are significantly different from the observed one, at the 95% level.

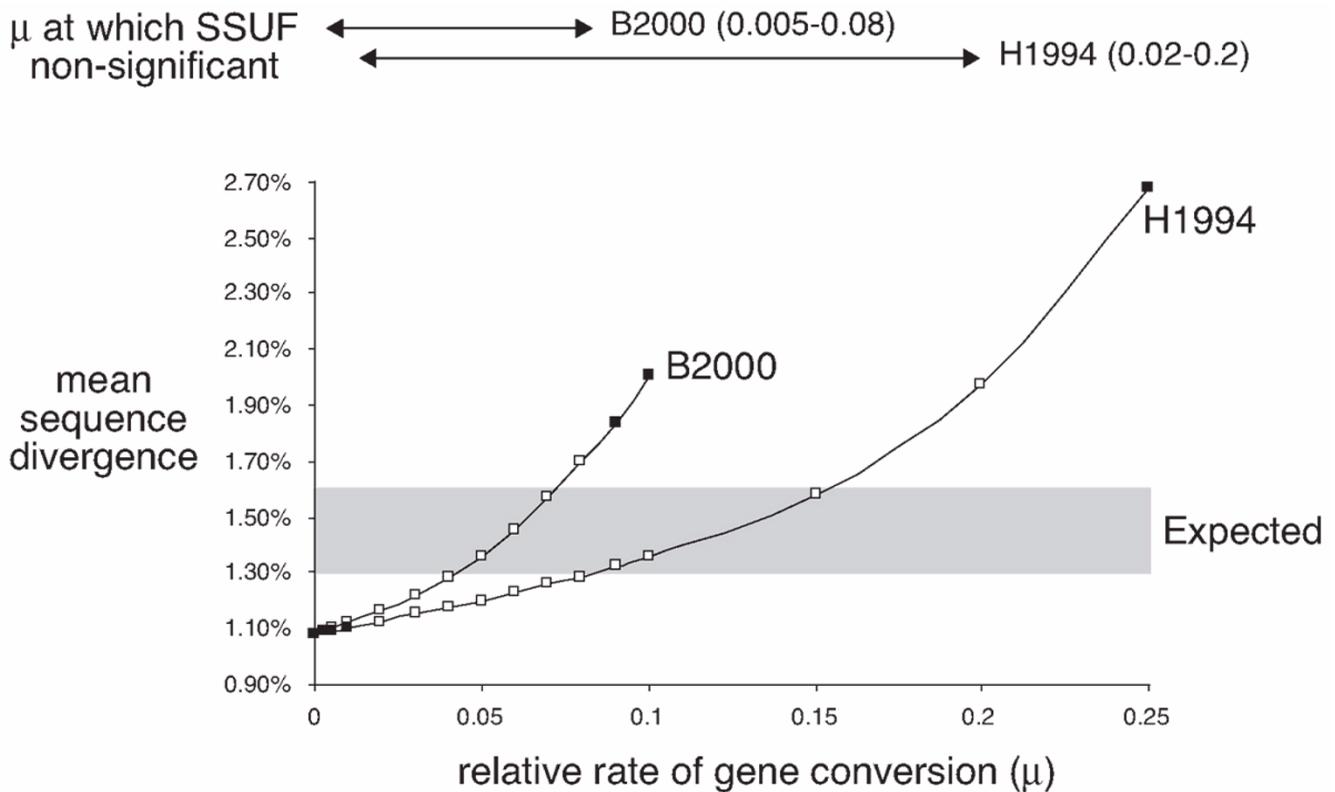


Figure 4
 Corrected sequence divergence of simulated CMT1A-REPs For each of the simulations represented in Figure 2 the mean number of variants that have arisen since the repeats began diverging is converted into a sequence divergence and plotted against the rate of gene conversion. The expected range for a locus that is shared between humans and chimps, but not gorillas is shown in grey. Each point represents an average over 1000 simulations. Significance is assessed against the SSUF from the observed alignment. Filled squares indicate rates at which simulated data sets are significantly different from the observed one, at the 95% level. The range of rates at which simulated datasets are not significantly different, in terms of SSUF, from the observed one is shown for both models above the chart.

Given our lack of knowledge of the mean conversion tract length in the human genome, the simulations (using the HI994 model) were repeated with this parameter varied within the range of 100–600 bp (Figure 3a). Similarly, for the B2000 model, the effect of varying the length of identity that constitutes a MEPS was varied within the range 100–350 bp (Figure 3b). These data show that as MEPS length and conversion tract length decrease, higher rates of gene conversion are required to explain the observed distribution of fragment sizes in the CMT1A-REP sequences.

The relative rate for gene conversion at CMT1A-REPs can be converted to an absolute *de novo* rate through a consideration of the rate of base substitution. The rate at which variants appear between the paralogs can be estimated as the genome average per nucleotide mutation rate multiplied by twice the length of the repeats (as variants can arise in either repeat). Thus the absolute conversion rate is

estimated by μ multiplied by this locus divergence rate. Using recent estimates of the human base substitution mutation rate [21], the central value between the bounds indicated in Figure 2 gives a gene conversion rate between the two CMT1A-REPs of 1.3×10^{-4} per generation for the HI994 model, and 5.1×10^{-5} per generation for the B2000 model.

It has previously been demonstrated that duplication of the CMT1A-REPs occurred after the divergence of gorillas from the human/chimp common ancestor [22]. Given such a history we should expect a level of sequence divergence of 1.3–1.6% [23]; in fact between these repeats we observe 1.1% sequence divergence (excluding indels). Indeed the alignment of orthologous sequences, which diverged more recently, exhibits greater sequence divergence despite having a lower frequency of CpGs. Figure 4 details the corrected, as opposed to the apparent, level of sequence divergence between repeats under the different

simulated scenarios – in other words the number of variant sites that have arisen during the simulated evolution rather than the number that appear in the alignment of extant sequences. The expected sequence divergence of 1.3–1.6% falls within the plausible bounds of μ , as determined by the SSUF measure previously, and encompasses the central values used above to calculate absolute rates.

Discussion

This is the first estimate of the rate of gene conversion between non-coding dispersed repeats, and thus it is difficult to compare with previous estimates for gene conversion processes. However, independent bounds on the rate of gene conversion can be provided by a recent study that detected the products of recombination between proximal and distal CMT1A-REP sequences in pools of sperm DNA [24]. This study detected products within a 1.4 kb region containing the 1 kb 'hotspot' in which breakpoints of 75% of all CMT1A-REP-sponsored duplications are found. These products could be expected to result from both single crossovers resulting in duplications and gene conversion events with one boundary within the hotspot region. The authors argued that the majority of such products resulted from single crossovers on the basis that the rate determined in small-pool PCR of $\sim 1.5 \times 10^{-5}$ was similar to estimates of CMT1A duplication frequencies based on patient prevalence ($0.98\text{--}6.5 \times 10^{-5}$) [24]. The uncertainty in these two estimates allows us to estimate bounds on the rate of gene conversion. The lower bound is clearly zero. The upper bound can be calculated by taking the highest possible SP-PCR rate given in this study (3.03×10^{-5}) and subtracting the lowest patient prevalence rate [24]. This gives an upper bound of 2.05×10^{-5} . It should be remembered that this figure refers solely to a single 1.4 kb interval within the 24 kb repeats. If we assume the gene conversion rate is constant across the repeat this gives a total rate of 3.5×10^{-4} per generation. Thus the estimates for gene conversion rates provided here fall well within the bounds defined by recent SP-PCR work on the same repeats. The assumption of a constant gene conversion rate across the repeat might not seem reliable; however, given the fact that the proximal and distal hotspots do not exhibit significantly higher levels of homology than the rest of the CMT1A-REPs it seems that gene conversion is unlikely to be as tightly localized to the hotspot as single crossovers.

It has been noted previously that using different models for gene conversion by unequal crossing over can result in rate estimates varying by a factor of 10 [25]. Using different models of gene conversion for the evolution of CMT1A-REPs is also likely to change the rate estimates, although they will not alter the requirement for considerable amounts of gene conversion in explaining the extant pattern of sequence diversity.

The probability of finding homogenized tracts within CMT1A-REPs in the general population is dependent on the numbers and diversity of the chromosomes screened and the degree to which selection acts on the homogenized repeats. Nevertheless, the *de novo* rate determined above is a lower bound on this probability. As sequenced extant CMT1A-REPs retain evidence of past gene conversion events it seems that homogenized repeats are not immediately removed from the gene pool and therefore the real probability of detecting homogenized repeats is likely to be considerably higher. Some hint of the true polymorphic nature of these repeats is revealed by a recent finding that over a third of individuals carry single 24 kb CMT1A-REPs that contain some single base variants that are proximal-specific and some that are distal-specific in the available full length sequences of these repeats [24].

Conclusions

Data presented here suggest that gene conversion plays a significant role in shaping the patterns of homology between autosomal paralogous repeats involved in pathogenic rearrangements. In addition, evidence has been presented that homogenized tracts within paralogous repeats are likely to be polymorphic within modern populations, polymorphisms which may well predispose carriers to pathogenic non-allelic homologous recombination. In light of these results, more effort should be made to ascertain diversity within the many known sets of paralogous repeats associated with pathogenic rearrangements. In addition, the potential predisposing effects of such homogenizations need to be tested, either by association studies, or directly by SP-PCR. Already it has been shown that there are significant variations in the rates of CMT1A duplications (as determined by SP-PCR) among different sperm donors [24].

Monte-Carlo simulations that incorporate non-allelic homologous recombination processes, and their comparison to accurate genomic sequences, should prove a useful method for disentangling the relative importance of various factors on the rate of non-allelic homologous recombination. Such factors include the length and orientation of repeats, distance between repeats, chromosomal location, chromatin structure and the presence of specific protein binding sites. Furthermore attempts to reconstruct the evolutionary history of low copy repeats (LCRs) and dispersed repeats [26–28] must take into account the confounding effect of gene conversion.

Materials and Methods

CMT1A sequence data were drawn from two sources. Cosmid sequences containing each CMT1A-REP [15] are under accession numbers HSU71217 and HSU71218. Finished BAC sequence for the proximal CMT1A-REP can be found under accession numbers AC005838 and

AC005324. Distal CMT1A-REP BAC sequences can be found under AC005224 and AC005389. Orthologous sequences used as controls, and containing no known genes or EST matches, were taken from finished BAC sequences from chromosome 12. Human sequences were from AC005294 and chimpanzee sequence from AC006582. All BAC sequences used here have expected error rates of less than 1 in 10000 bases.

Sequence alignments were performed using Se-AL version 2.0a7b [http://evolve.zoo.ox.ac.uk/] and are available from the author on request. Variant sites and CpG positions were determined using the dreg pattern matching program in the EMNU package at the HGMP-RC [http://www.hgmp.mrc.ac.uk/]. All CpG dinucleotides present in either sequence were considered to be present in the ancestral sequence. Dispersed repeats were identified using RepeatMasker [http://ftp.genome.washington.edu/cgi-bin/RepeatMasker].

In order to take account of the differential mutability of bases within coding sequences Sawyer's original test [14] assigns each base to one of four classes depending on the degeneracy of its codon. A permutation test then maintains the same number of variant sites in each class. Rate heterogeneity among sites in non-coding sequences does not depend on codon degeneracy. CpG mutational hotspots, however, play an important role [23,29]. In addition the rate of small insertion/deletion events (indels) is clearly related to slippage induced by the 'cryptic simplicity' of the surrounding sequence [30], in other words the degree of local repetition of small nucleotide motifs. However, whereas mutation at CpGs is well characterized, the dynamics of cryptic simplicity are very poorly understood, and consequently all indels were discarded from the alignment. Remaining sites were then assigned into two classes, those where a CpG was present in the ancestral sequence and those where it was not. These two classes of variant sites were then randomly permuted independently to maintain the same frequency of CpG variant sites, the positions of which are conditioned on the position of CpG sites in the ancestral sequence.

The Sawyer-CpG gene conversion test was performed using 10000 permutations. Each Monte Carlo simulation comprised 1000 replications that were halted once the number of extant variant sites equaled that found between the aligned sequences. The relative rate of CpG to non-CpG mutation was determined heuristically by selecting a rate at which the observed (73/257) value of the ratio of variant sites that were due to CpG mutation equaled the median of this value among 1000 simulations. The relative rate of CpG increased mutability that was adopted (14.8) was similar to that determined empirically [29]. P values were determined in a one-tailed fashion as the frac-

tion of replicates whose SSUF values are greater (if mean SSUF < observed SSUF) or less (if mean SSUF > observed SSUF) than that from the sequenced CMT1A-REPs.

All software and simulations were written in Interactive Data Language (IDL) version 5.4, an array-based higher order language. Analysis was performed on a Mac G4 Powerbook with 400 MHz processor. Source code and IDL executable files are available from the author on request.

Acknowledgements

I am grateful for the useful comments of Mark Jobling, Peter Forster and Chris Tyler-Smith on earlier versions of the manuscript. My work is supported by the McDonald Institute for Archaeological Research.

References

1. Lupski J: **Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits.** *Trends Genet* 1998, **14**:417-422
2. Emanuel BS: **Segmental duplications: an expanding role in genomic instability and disease.** *Nature Reviews Genetics* 2001, **2**:791-800
3. Eichler EE: **Recent duplication, domain accretion and the dynamic mutation of the human genome.** *Trends Genet* 2001, **17**:661-669
4. Pentao L, Wise CA, Chinault AC, Patel PI, Lupski JR: **Charcot-Marie-Tooth Type-1a Duplication Appears to Arise from Recombination at Repeat Sequences Flanking the 1.5 Mb Monomer Unit.** *Nature Genetics* 1992, **2**:292-300
5. Chance PF, Abbas N, Lensch MW, Pentao L, Roa BB, Patel PI, Lupski JR: **Autosomal-Dominant Neuropathies Result from Reciprocal DNA Duplication/Deletion of a Region on Chromosome-17.** *Human Molecular Genetics* 1994, **3**:223-228
6. Potocki L, Chen K-S, Park S-S, Osterholm DE, Withers MA, Kimonis V, Summers AM, Meschino WS, Anyane-Yeboah K, Kashork CD, Shaffer LS, Lupski JR: **Molecular mechanism for duplication 17p11.2 - the homologous recombination reciprocal of the Smith-Magenis microdeletion.** *Nat. Genet* 2000, **24**:84-87
7. Lakich D, Kazazian HH, Antonarakis SE, Gitschier J: **Inversions Disrupting the Factor-VIII Gene Are a Common-Cause of Severe Hemophilia-A.** *Nature Genetics* 1993, **5**:236-241
8. Tilford CA, Kuroda-Kawaguchi T, Skaletsky H, Rozen S, Brown LG, Rosenberg M, McPherson JD, Wylie K, Sekhon M, Kucaba TA, Waterston RH, Page DC: **A physical map of the human Y chromosome.** *Nature* 2001, **409**:943-945
9. Small K, Iber J, Warren ST: **Emerin deletion reveals a common X-chromosome inversion mediated by inverted repeats.** *Nature Genetics* 1997, **16**:96-99
10. Giglio S, Broman KW, Matsumoto N, Calvari V, Gimelli G, Neumann T, Ohashi H, Voullaire L, Larizza D, Giorda R, Weber JL, Ledbetter DH, Zuffardi O: **Olfactory receptor-gene clusters, genomic-inversion polymorphisms, and common chromosome rearrangements.** *American Journal of Human Genetics* 2001, **68**:874-883
11. Jobling MA, Williams G, Schiebel K, Pandya A, McElreavey K, Salas L, Rappold GA, Affara NA, Tyler-Smith C: **A selective difference between human Y-chromosomal DNA haplotypes.** *Curr. Biol* 1998, **8**:1391-1394
12. Blanco P, Shlumukova M, Sargent CA, Jobling MA, Affara N, Hurles ME: **Divergent outcomes of intrachromosomal recombination on the human Y chromosome: male infertility and recurrent polymorphism.** *Journal of Medical Genetics* 2000, **37**:752-758
13. Hollies CR, Monckton DG, Jeffreys AJ: **Attempts to detect retrotransposition and de novo deletion of Alus and other dispersed repeats at specific loci in the human genome.** *European Journal of Human Genetics* 2001, **9**:143-146
14. Sawyer S: **Statistical tests for detecting gene conversion.** *Molecular Biology and Evolution* 1989, **6**:526-538
15. Reiter LT, Murakami T, Koeuth T, Gibbs RA, Lupski JR: **The human COX10 gene is disrupted during homologous recombination**

- between the 24 kb proximal and distal CMT1A-REPs. *Human Molecular Genetics* 1997, **6**:1595-1603
16. Inoue K, Dewar K, Katsanis N, Reiter LT, Lander ES, Devon KL, Wyman DW, Lupski JR, Birren B: **The 1.4-Mb CMT1A duplication/HNPP deletion genomic region reveals unique genome architectural features and provides insights into the recent evolution of new genes.** *Genome Research* 2001, **11**:1018-1033
 17. Wiuf C, Christensen T, Hein J: **A simulation study of the reliability of recombination detection methods.** *Molecular Biology and Evolution* 2001, **18**:1929-1939
 18. Lukacsovich T, Waldman AS: **Suppression of intrachromosomal gene conversion in mammalian cells by small degrees of sequence divergence.** *Genetics* 1999, **151**:1559-1568
 19. Hilliker AJ, Harauz G, Reaume AG, Gray M, dark SH, Chovnick A: **Meiotic Gene Conversion Tract Length Distribution within the Rosy Locus of Drosophila-Melanogaster.** *Genetics* 1994, **137**:1019-1024
 20. Wiuf C, Hein J: **The coalescent with gene conversion.** *Genetics* 2000, **155**:451-462
 21. Halushka MK, Fan JB, Bentley K, Hsie L, Shen NP, Weder A, Cooper R, Lipshutz R, Chakravarti A: **Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis.** *Nature Genetics* 1999, **22**:239-247
 22. Kiyosawa H, Chance PF: **Primate origin of the CMT1A-REP repeat and analysis of a putative transposon-associated recombinational hotspot.** *Human Molecular Genetics* 1996, **5**:745-753
 23. Chen FC, Li WH: **Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees.** *American Journal of Human Genetics* 2001, **68**:444-456
 24. Han LL, Keller MP, Navidi W, Chance PF, Arnheim N: **Unequal exchange at the Charcot-Marie-Tooth disease type 1A recombination hot-spot is not elevated above the genome average rate.** *Human Molecular Genetics* 2000, **9**:1881-1889
 25. Nei M: *Molecular Evolutionary Genetics.* New York: Columbia University Press; 1987
 26. Johnson WE, Coffin JM: **Constructing primate phylogenies from ancient retrovirus sequences.** *Proceedings of the National Academy of Sciences of the United States of America* 1999, **96**:10254-10260
 27. Johnson ME, Viggiano L, Bailey JA, Abdul-Rauf M, Goodwin G, Rocchi M, Eichler EE: **Positive selection of a gene family during the emergence of humans and African apes.** *Nature* 2001, **413**:514-519
 28. Roy AM, Carroll ML, Nguyen SV, Salem AH, Oldridge M, Wilkie AOM, Batzer MA, Deininger PL: **Potential gene conversion and source genes for recently integrated Alu elements.** *Genome Research* 2000, **10**:1485-1495
 29. Nachman MW, Crowell SL: **Estimate of the mutation rate per nucleotide in humans.** *Genetics* 2000, **156**:297-304
 30. Tautz D, Trick M, Dover G: **Cryptic simplicity in DNA is a major source of genetic variation.** *Nature* 1986, **322**:652-656

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMedcentral will be the most significant development for disseminating the results of biomedical research in our lifetime."

Paul Nurse, Director-General, Imperial Cancer Research Fund

Publish with **BMC** and your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours - you keep the copyright



Submit your manuscript here:

<http://www.biomedcentral.com/manuscript/>

editorial@biomedcentral.com