

# How homologous recombination generates a mutable genome

Matthew Hurles\*

Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge, CB10 1SA, UK

\*Correspondence to: Tel: +44 (0)1223 495377; Fax: +44 (0)1223 494919; E-mail: meh@sanger.ac.uk

Date received (in revised form): 7th July 2005

## Abstract

Recombination and mutation have traditionally been regarded as independent evolutionary processes: the latter generates variation, which the former reshuffles. Recent studies, however, have suggested that allelic recombination influences the underlying mutation rate, as high mutation rates are inferred in regions of high recombination. Furthermore, recombination between duplicated sequences introduces structural variation into the human genome and facilitates the formation of clustered gene families. Comparisons of whole-genome sequences reveal the expansion of gene family clusters to be an important mode of genome evolution. The negative aspect of this genomic dynamism is the contribution of these rearrangements to genetic diseases.

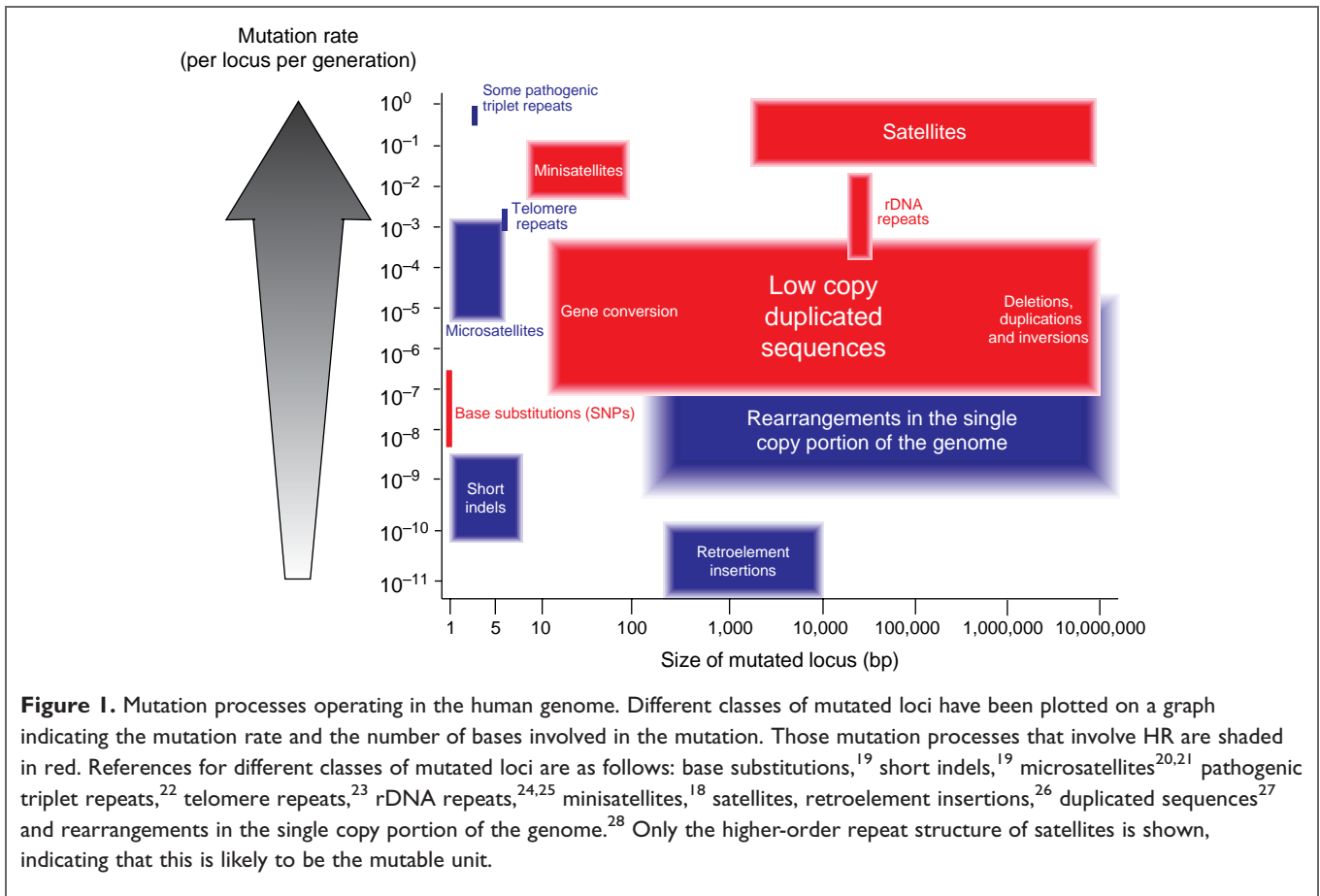
**Keywords:** *homologous recombination, gene conversion, rearrangement*

## Introduction

Homologous recombination (HR) is one of the fundamental mechanisms of DNA processing which, in various guises, is found in all phyla of life.<sup>1,2</sup> HR is capable of playing several distinct roles within an individual organism. In sexually reproducing species, meiotic HR is a carefully regulated process that occurs at a defined stage of differentiation in specific cell types. By contrast, in the same species, HR also operates as a major mechanism of DNA repair in all cell types at all times. HR has been clearly co-opted for different functions throughout its deep evolutionary history. Similarly, HR has been exploited as a laboratory tool for, among other applications, genetic engineering in model organisms.<sup>3</sup> The role of HR in DNA repair,<sup>4</sup> somatic mutation<sup>5</sup> and chromosomal engineering has been reviewed elsewhere; this paper will focus on meiotic HR and the recent studies demonstrating its impact on the mutability of mammalian genomes.

Evolutionary geneticists have traditionally regarded mutation and recombination (along with selection and genetic drift) as relatively independent 'forces of evolution': while the former generates variation, the latter reshuffles existing variation into novel combinations. These ideas were formulated before DNA was identified as the molecule of inheritance,<sup>6-8</sup> however, and well before any understanding two of the molecular mechanisms of mutation was gained. Recent comparative analyses of whole-genome sequences<sup>9-12</sup> give a deeper appreciation of the distinct mutational mechanisms

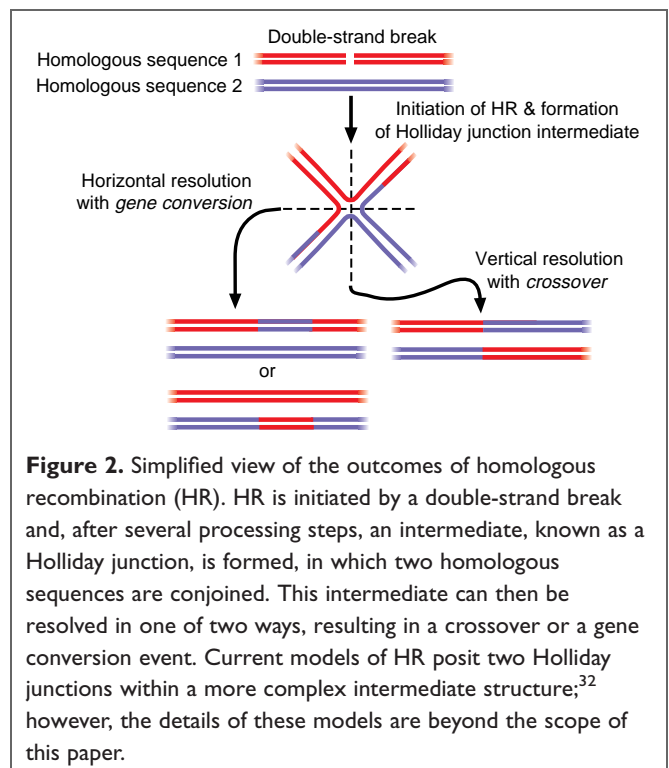
operating to shape genomes over evolutionary timescales. The mutability of any genome can be considered to be the summation of the effects of the distinct mutational mechanisms that operate in that genome. These impacts can be quantified in terms of the rate of each mutational mechanism, the number of bases involved in the resultant mutation and the number of susceptible sites within the genome. Figure 1 displays the major mutational mechanisms operating in mammalian genomes and demonstrates that both in their rate and the size of the resultant genomic alteration vary widely. The distinction between recombination and mutation described above becomes blurred by the involvement of HR in a number of these mutational processes. There is ongoing discussion about the relationship between allelic HR and single nucleotide polymorphism (SNP; see Hellmann *et al.*<sup>13</sup> and Nachman,<sup>14</sup> and section below, for example). Furthermore, HR between non-allelic (duplicated) sequences has been demonstrated to be an important mode of both pathogenic mutation<sup>15</sup> and genome evolution.<sup>16</sup> These duplicated sequences can exist in tandem arrays or in dispersed repeats.<sup>17</sup> HR is the predominant mutational mechanism operating in polymorphic tandem repeat arrays with repeat units longer than five base pairs (bp),<sup>18</sup> including minisatellites and ribosomal DNA arrays, whereas replication slippage operates on arrays of shorter repeat units. These two influences of HR on shaping genomic variation will be considered in turn, but first one should consider the mechanism of HR and the distribution of allelic HR throughout the human genome.



## Mechanism and genomic distribution of HR

The specifics of the multifarious protein–DNA interactions that underpin HR are beyond the scope of this paper (but have been reviewed by West<sup>1</sup>). It is sufficient to note that HR is initiated by a DNA double-strand break (DSB). This break is subsequently processed and then invades a homologous acceptor sequence.<sup>29</sup> After further processing, an intermediate is formed that can be resolved in one of two ways: a crossover results in the reciprocal splicing of the donor sequence to the acceptor sequence, whereas a gene conversion results in the non-reciprocal transfer of a short tract of sequence between the sequences (Figure 2). A crossover results in a change of phase of flanking markers on either side of the crossover, whereas a gene conversion is only observable if it encompasses a variant site between the homologous sequences. The ratio of these two outcomes is poorly characterised, although recent empirical<sup>30</sup> and statistical analyses<sup>21</sup> seem to indicate that gene conversion is the more frequent outcome.

The distribution of allelic HR throughout the human genome is likely to be extremely heterogeneous on the fine (kilobase [kb]) scale, although on the coarse scale (tens of



megabases), the broad pattern is one of 1.6-fold more recombination events in females than in males, depressed recombination near the centromere and increased recombination in subtelomeric regions.<sup>33</sup> On the fine scale, both empirical studies of recombination in sperm<sup>34</sup> and statistical analysis of patterns of variation in populations<sup>35,36</sup> indicate the widespread existence of hotspots of recombinatorial activity that can be orders of magnitude more active within an interval of 500 bp to 1 kb than in flanking ‘cold’ sequences. Population genetics theory suggests that these HR hotspots are likely to be short lived because the dynamics of the HR process are such that recombinogenic variants are doomed to be preferentially gene converted out of existence.<sup>37,38</sup> This prediction appears to have been bolstered by the recent observation that recombination hotspots are not conserved over the short evolutionary distance that separates humans and chimpanzees.<sup>39,40</sup> An absence of shared sequence motifs between known recombination hotspots<sup>35</sup> suggests that epigenetic mechanisms might be involved in the inheritance of recombinatorial activity at these locations.

The number of recombination events per meiosis seems to vary significantly both between gametes and between healthy individuals.<sup>41</sup> Intriguingly, mothers with higher rates of recombination tend to have greater reproductive success,<sup>42</sup> which would suggest that selection on the dynamics of allelic HR is ongoing.

## Allelic HR and sequence diversity

There is, on average, higher sequence diversity in regions of higher recombination rate.<sup>14</sup> It has also been demonstrated that there is a similar correlation between recombination and divergence in comparisons between human and mouse<sup>43</sup> and human and chimpanzee.<sup>13</sup> This correlation need not be explained by a causative relationship between the two (ie HR being mutagenic), although it has been suggested that errors in the repair of DSBs that initiate HR could increase the mutation rate. The hypothesis that recombination is itself mutagenic contrasts with the observation that recombination hotspots are short-lived evolutionary phenomena. Although, genomic location (ie proximity to telomeres) is more evolutionarily stable and plays a role in patterning large-scale recombination activity. Patterns of fine-scale recombination rate in humans, however, may be a poor predictor of the recombinatorial landscape in which sequences have evolved on both human and mouse lineages since they last shared a common ancestor.

Selection has also been invoked to explain the relationship between recombination and diversity. The rationale is that positive (hitchhiking) or negative (background) selection at linked loci is expected to reduce diversity and that, by breaking down linkage, recombination can release neighbouring markers from this selection-induced reduction in

diversity.<sup>44,45</sup> It has been argued, however, that the correlation between recombination and divergence suggests that a neutral, rather than a selective, explanation is more likely. In addition, the relationship between recombination and diversity shows no correlation with gene density,<sup>43</sup> which further argues against selection.

It is possible that regions of higher recombination rate are likely to have an elevated mutation rate, not because recombination is itself mutagenic, but because the same genomic features elevate both. The best candidate appears to be GC content, which is positively correlated with recombination rate.<sup>46</sup> Indeed, HR is thought to be GC biased — that is, gene conversion will preferentially repair a mismatched base pair to a G-C rather than an A-T.<sup>47,48</sup> As a consequence, a high recombination rate might be expected to lead to a high GC content over time. Correspondingly, the approximately ten-fold greater mutability of the CpG dinucleotide<sup>19</sup> over other dinucleotides suggests that mutation should also be elevated in regions of high GC content. Once GC content is taken into account, much of the association between recombination and mutation at larger scales disappears, although a fine-scale correlation of diversity and recombination rate persists.<sup>13</sup>

## Non-allelic HR between duplicated sequences

Much of the recent interest in non-allelic HR has been caused by the observation from the Human Genome Project that almost half of the human genome is duplicated elsewhere in the genome.<sup>49–51</sup> Approximately 5–6 per cent of the human genome can be found in >1 kb blocks of >90 per cent sequence similarity to other locations in the genome (known as segmental duplications).<sup>50,51</sup> Furthermore, about 42 per cent of the genome is accounted for by families of dispersed repetitive elements — short interspersed nuclear elements (SINEs), long interspersed nuclear elements (LINEs) and human endogenous retroviruses (HERVs) —.<sup>9</sup> While these two classes of duplicated sequence are typically considered separately in evolutionary analyses, it is clear that non-allelic HR can occur within both sets of duplicated sequence.<sup>52,53</sup>

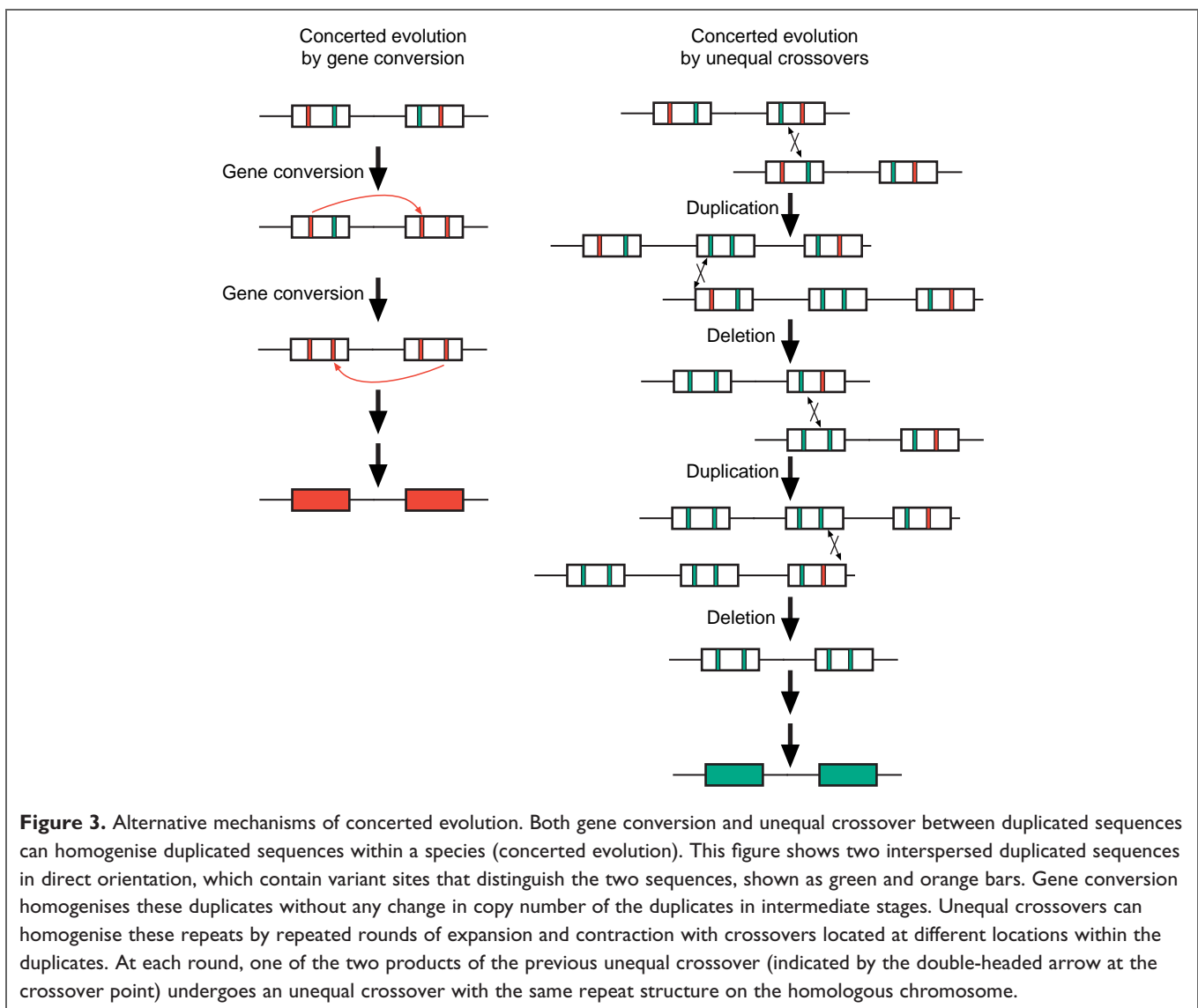
Multiple mechanisms account for the origins of these duplicated sequences. Families of dispersed repeats seem to populate the genome in bursts of infectious activity. Non-allelic HR between these small dispersed repeats can in turn lead to much larger segmental duplications.<sup>54</sup> It has been suggested that the rapid populating of an ancestral primate genome with *Alu* elements facilitated a recent burst of segmental duplication<sup>54</sup> approximately 40 million years ago. It also appears, however, that physical fragility of the DNA sequence additionally plays a role in generating segmental duplications.<sup>55</sup>

Non-allelic HR, similarly to allelic HR, can result in crossovers (sometimes known as unequal crossovers) and

gene conversion events and is characterised by the existence of hotspots of activity.<sup>56</sup> The duplicated substrates for non-allelic HR can be on the same or non-homologous chromosomes, although it appears that intrachromosomal interactions are much more frequent.<sup>57</sup> Crossovers promoted by non-allelic HR generate rearrangements; the precise structural change depends on the orientation of the duplicated sequences: repeats in direct orientation promote deletions and duplications, whereas repeats in inverted orientation sponsor inversions.<sup>15</sup> Copy number changes sponsored by non-allelic HR need not simply involve deletion or duplication of single copy sequence but may also include dramatic variation in the copy number of tandemly duplicated arrays; for example, an individual X chromosome may carry between one and nine copies of the X-linked opsin genes.<sup>58</sup> The prevalence of these tandem arrays in the human genome has yet to be

systematically characterised; however, genes known to exist in polymorphic arrays include those for amylase, alpha-defensins, beta-defensins, opsins, CYP2D6, TSPY, globins, rDNA and histones.

Concerted evolution describes the observation that duplicated sequences appear to be more closely related within species than they are to their orthologues in related species.<sup>59</sup> Concerted evolution can arise from both gene conversion and multiple rounds of unequal crossovers (Figure 3). Gene conversions transfer sequence between duplicated substrates, which can lead to the homogenisation of a family of repeats. Concerted evolutionary processes can thwart attempts to date duplication events that equate sequence similarity between two duplicated sequences to the time since duplication.<sup>60,61</sup> By homogenising duplicated sequences, gene conversion causes such analyses to underestimate the age of the



duplication event; many duplications are older than they might first appear. While concerted evolution was initially characterised in tandemly arrayed gene families (eg ribosomal DNA, globins, opsins), it has more recently been observed in interspersed duplications.<sup>62</sup>

As with allelic HR, it has been suggested influence that non-allelic HR sequence diversity and divergence in duplicated sequences. There have been conflicting reports of the direction of this influence. Analysis of the long, almost identical inverted repeats on Yq has indicated a significantly lower sequence divergence between humans and chimpanzees within the repeats than in flanking single copy sequence.<sup>63</sup> This difference was attributed to high levels of gene conversion operating between the inverted repeats repressing sequence divergence between orthologous sequences. By contrast, analyses of the duplicated HERV sequences that promote the *AZFa* deletion (also on Yq) have revealed elevated sequence divergence within known non-allelic HR hotspots<sup>64</sup> and increased sequence diversity flanking the hotspot.<sup>65</sup> Simulations of the gene conversion process suggest that elevated sequence diversity and orthologous divergence is to be expected when duplicated sequences are themselves slightly differentiated.<sup>64</sup> It remains to be seen whether these observations can be generalised to the entire genome, but the enrichment of apparent SNPs (in the dbSNP database of sequence variation) within segmental duplications<sup>66</sup> and the observation of the gene conversion process operating on other chromosomes is highly suggestive.<sup>67</sup>

## Evolutionary benefits of having a duplicated genome

Gene duplication and divergence has long been prophesied to be the major mechanism by which novel gene functions arise.<sup>68</sup> Once a gene has been duplicated, selective constraints are relaxed and there are several mechanisms by which the duplicates can diverge in function while fulfilling the role of the ancestral gene (reviewed by Hurles<sup>69</sup>). The widespread existence of gene families pays testament to the importance of gene duplication in evolution. Comparative whole-genome sequence analysis now gives a complete picture of how genomes adapt to novel environments. Comparisons between human and rodent and mammalian and avian genomes<sup>10–12</sup> have implied the importance of lineage-specific expansions of particular clustered gene families, although greater effort is required, on a locus by locus basis, to discount the role of neutral processes in the origins of these structures. The clustering of these genes implicates non-allelic HR, both in their origins and in their patterns of sequence evolution. These lineage-specific expansions are often of gene families involved in sensory perception, toxin metabolism, immune response and reproduction.<sup>70</sup> These gene functions are also observed in single copy genes that show evidence of

recent positive selection,<sup>71</sup> suggesting that these functions are among the most important for rapid adaptation to novel environments. Interestingly, many pathogens also exhibit clusters of genes involved in antigenic variation; it appears that non-allelic HR is an important mutational mechanism operating in the ongoing arms race between pathogens and immune systems. The relatively high mutation rate of non-allelic HR, compared with that of sequence evolution, is probably an important factor underlying this observation.

## Non-allelic HR and disease

As with all mutational processes, non-allelic HR generates variation that is subject to natural selection. While this variation can confer evolutionary benefits, as described above, it can also cause disease. Both unequal crossovers and gene conversions between duplicated sequences have pathogenic potential. A growing number of genetic diseases (Table 1) have been recognised to be caused by deletions and duplications of dosage-sensitive genes, inversions disrupting genic structures and gene conversions ablating normal gene function.<sup>15</sup> Disease-causing rearrangements have been identified within tandemly duplicated gene arrays, as well as between interspersed duplicates. While non-allelic HR is not the sole cause of structural variation in the human genome, HR between tandemly duplicated arrays, between segmental duplications and between dispersed repeats has been demonstrated to be a major cause of these mutations.

Perhaps the most common outcome of gene duplication is that one of the copies acquires mutations that may render it non-functional. Not only is the potential for evolving a novel function lost, but this pseudogene now contains a reservoir of mutations that can be gene converted into the remaining functional gene.<sup>72–74</sup> The prevalence of pseudogenes in the human genome suggests that many genes will have associated pseudogenes.<sup>75</sup>

While much attention has focused on the role of non-allelic HR in genetic disorders with Mendelian inheritance patterns, little effort has been devoted to investigating its role in the genetics of complex diseases. This is despite the longstanding existence of examples of the role of rearrangements in more complex phenotypes such as drug response<sup>76</sup> and resistance to infectious diseases such as malaria.<sup>77</sup> More recently, elevated copy number of a segmental duplication containing the gene *CCL3L1* has been shown to protect against HIV/AIDS.<sup>78</sup> In addition, a chromosomal inversion with a convoluted evolutionary history has been demonstrated to confer a selective advantage in recent generations of the Icelandic population.<sup>79</sup> The lack of studies investigating structural polymorphism and complex traits has perhaps been due to an underestimation of the degree of structural variation in the human genome. Recent studies demonstrate that there is much more large-scale copy number variation than was previously thought to exist and also

**Table 1.** Examples of diseases caused by non-allelic HR.

Disease	Inheritance	Type of NAHR	Size	OMIM
Velocardiofacial syndrome	Autosomal dominant	Deletion	3 Mb	192430
Haemophilia A	X-linked recessive	Inversion	500 kb	306700
Charcot–Marie–Tooth type IA	Autosomal dominant	Duplication	1.5 Mb	118220
AZF <sub>a</sub>	Y-linked	Deletion	800 kb	400042
Congenital adrenal hyperplasia	Autosomal recessive	Gene conversion	N/A	201910
Gaucher disease	Autosomal recessive	Gene conversion	N/A	230800
Hunter syndrome	X-linked recessive	Inversion	20 kb	309900
Prader–Willi Syndrome	Autosomal dominant	Deletion	3.5 Mb	176270
Alpha-thalassaemia	Autosomal recessive	Deletion	3.7 kb*	141800

NAHR, non-allelic homologous recombination; OMIM, database of online Mendelian inheritance in Man; Mb, megabases; kb, kilobases.

\*There are alpha-globin deletions of varying sizes, this is one of the smallest.

point towards methods that can redress this under-ascertainment in a systematic fashion.<sup>80–82</sup>

## Conclusions

The mutagenic potential of non-allelic HR was identified early in the history of molecular genetics, yet, due to the difficulty of experimentally interrogating duplicated sequences, a fuller appreciation of its evolutionary and pathogenic roles has had to await the publication of whole-genome sequences. Clearly, there are both costs and benefits to having a highly duplicated, and therefore mutable, genome.

Despite recent advances, non-allelic HR remains perhaps the most poorly characterised mutation process in the human genome. While the human genome reference sequence provides a reasonable understanding of where non-allelic HR is likely to occur, little is known about the rates of these processes and how they vary between individuals and over evolutionary time-scales. It is worth noting that variation in the frequencies of chromosomal rearrangements along different evolutionary lineages need not reflect the degree of duplication in an ancestral genome, but might result from specific demographic histories (eg population bottlenecks) that have transiently favoured the fixation of chromosomal rearrangements.

Given the role that non-allelic HR appears to have played in the rapid adaptation to novel environments observed within mammalian genome comparisons, it will be of great interest to investigate the genomic changes that must have accompanied the adaptation of different groups of humans to the wide range of environments that our species presently occupies. The recent work on the *CCL3L1*-containing segmental duplication described above illustrates how a reservoir

of structural variation allows a rapid response to the new selective environment posed by a novel human pathogen.<sup>78</sup>

## Acknowledgments

This work was funded by the Wellcome Trust. The author is grateful to Jim Lupski for his comments on an earlier version of the manuscript.

## References

- West, S.C. (2003), 'Molecular views of recombination proteins and their control', *Nat. Rev. Mol. Cell. Biol.* Vol. 4, pp. 435–445.
- Bhattacharyya, M.K., Norris, D.E. and Kumar, N. (2004), 'Molecular players of homologous recombination in protozoan parasites: Implications for generating antigenic variation', *Infect. Genet. Evol.* Vol. 4, pp. 91–98.
- Thomas, K.R. and Capecchi, M.R. (1986), 'Introduction of homologous DNA sequences into mammalian cells induces mutations in the cognate gene', *Nature* Vol. 324, pp. 34–38.
- Dudas, A. and Chovanec, M. (2004), 'DNA double-strand break repair by homologous recombination', *Mutat. Res.* Vol. 566, pp. 131–167.
- Bishop, A.J. and Schiestl, R.H. (2003), 'Role of homologous recombination in carcinogenesis', *Exp. Mol. Pathol.* Vol. 74, pp. 94–105.
- Wright, S. (1931), 'Evolution in Mendelian populations', *Genetics* Vol. 16, pp. 97–159.
- Haldane, J.B.S. (1932), 'The Causes of Evolution', Longmans and Green, London, UK.
- Fisher, R.A. (1930), 'The Genetical Theory of Natural Selection', Clarendon Press, Oxford, UK.
- IHGSC (2001), 'Initial sequencing and analysis of the human genome', *Nature* Vol. 409, pp. 860–921.
- ICGSC (2004), 'Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution', *Nature* Vol. 432, pp. 695–716.
- MGSC (2002), 'Initial sequencing and comparative analysis of the mouse genome', *Nature* Vol. 420, pp. 520–562.
- RGSPC (2004), 'Genome sequence of the Brown Norway rat yields insights into mammalian evolution', *Nature* Vol. 428, pp. 493–521.

13. Hellmann, I., Ebersberger, I., Ptak, S.E. *et al.* (2003), 'A neutral explanation for the correlation of diversity with recombination rates in humans', *Am. J. Hum. Genet.* Vol. 72, pp. 1527–1535.
14. Nachman, M.W. (2001), 'Single nucleotide polymorphisms and recombination rate in humans', *Trends Genet.* Vol. 17, pp. 481–485.
15. Stankiewicz, P. and Lupski, J.R. (2002), 'Genome architecture, rearrangements and genomic disorders', *Trends Genet.* Vol. 18, pp. 74–82.
16. Eichler, E.E. and Sankoff, D. (2003), 'Structural dynamics of eukaryotic chromosome evolution', *Science* Vol. 301, pp. 793–797.
17. Achaz, G., Netter, P. and Coissac, E. (2001), 'Study of intrachromosomal duplications among the eukaryote genomes', *Mol. Biol. Evol.* Vol. 18, pp. 2280–2288.
18. Jeffreys, A.J., Barber, R., Bois, P. *et al.* (1999), 'Human minisatellites, repeat DNA instability and meiotic recombination', *Electrophoresis* Vol. 20, pp. 1665–1675.
19. Nachman, M.W. and Crowell, S.L. (2000), 'Estimate of the mutation rate per nucleotide in humans', *Genetics* Vol. 156, pp. 297–304.
20. Xu, X., Peng, M. and Fang, Z. (2000), 'The direction of microsatellite mutations is dependent upon allele length', *Nat. Genet.* Vol. 24, pp. 396–399.
21. Kayser, M., Roewer, L., Hedman, M. *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.* Vol. 66, pp. 1580–1588.
22. Leeflang, E.P., Tavare, S., Marjoram, P. *et al.* (1999), 'Analysis of germline mutation spectra at the Huntington's disease locus supports a mitotic mutation mechanism', *Hum. Mol. Genet.* Vol. 8, pp. 173–183.
23. Coleman, J., Baird, D.M. and Royle, N.J. (1999), 'The plasticity of human telomeres demonstrated by a hypervariable telomere repeat array that is located on some copies of 16p and 16q', *Hum. Mol. Genet.* Vol. 8, pp. 1637–1646.
24. Nei, M. (1987), 'Molecular Evolutionary Genetics', Columbia University Press, New York, NY.
25. Strachan, T., Webb, D. and Dover, G.A. (1985), 'Transition stages of molecular drive in multiple-copy DNA families in *Drosophila*', *Embo J.* Vol. 4, pp. 1701–1708.
26. Deininger, P.L. and Batzer, M.A. (1999), 'Alu repeats and human disease', *Mol. Genet. Metab.* Vol. 67, pp. 183–193.
27. Shaffer, L.G. and Lupski, J.R. (2000), 'Molecular mechanisms for constitutional chromosomal rearrangements in humans', *Annu. Rev. Genet.* Vol. 34, pp. 297–329.
28. Kurahashi, H. and Emanuel, B.S. (2001), 'Unexpectedly high rate of *de novo* constitutional t (11;22) translocations in sperm from normal males', *Nat. Genet.* Vol. 29, pp. 139–140.
29. Szostak, J.W., Orr-Weaver, T.L., Rothstein, R.J. and Stahl, F.W. (1983), 'The double-strand-break repair model for recombination', *Cell* Vol. 33, pp. 25–35.
30. Jeffreys, A.J. and May, C.A. (2004), 'Intense and highly localized gene conversion activity in human meiotic crossover hot spots', *Nat. Genet.* Vol. 36, pp. 151–156.
31. Frisse, L., Hudson, R.R., Bartoszewicz, A. *et al.* (2001), 'Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels', *Am. J. Hum. Genet.* Vol. 69, pp. 831–843.
32. Heyer, W.D., Ehmsen, K.T. and Solinger, J.A. (2003), 'Holliday junctions in the eukaryotic nucleus: Resolution in sight?', *Trends Biochem. Sci.* Vol. 28, pp. 548–557.
33. Kong, A., Gudbjartsson, D.F., Sainz, J. *et al.* (2002), 'A high-resolution recombination map of the human genome', *Nat. Genet.* Vol. 31, pp. 241–247.
34. Jeffreys, A.J., Kauppi, L. and Neumann, R. (2001), 'Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex', *Nat. Genet.* Vol. 29, pp. 217–222.
35. McVean, G.A., Myers, S.R., Hunt, S. *et al.* (2004), 'The fine-scale structure of recombination rate variation in the human genome', *Science* Vol. 304, pp. 581–584.
36. Crawford, D.C., Bhangale, T., Li, N. *et al.* (2004), 'Evidence for substantial fine-scale variation in recombination rates across the human genome', *Nat. Genet.* Vol. 36, pp. 700–706.
37. Boulton, A., Myers, R.S. and Redfield, R.J. (1997), 'The hotspot conversion paradox and the evolution of meiotic recombination', *Proc. Natl. Acad. Sci. USA* Vol. 94, pp. 8058–8063.
38. Jeffreys, A.J. and Neumann, R. (2002), 'Reciprocal crossover asymmetry and meiotic drive in a human recombination hot spot', *Nat. Genet.* Vol. 31, pp. 267–271.
39. Winckler, W., Myers, S.R., Richter, D.J. *et al.* (2005), 'Comparison of fine-scale recombination rates in humans and chimpanzees', *Science* Vol. 308, pp. 107–111.
40. Ptak, S.E., Hinds, D.A., Koehler, K. *et al.* (2005), 'Fine-scale recombination patterns differ between chimpanzees and humans', *Nat. Genet.* Vol. 37, pp. 429–434.
41. Lynn, A., Koehler, K.E., Judis, L. *et al.* (2002), 'Covariation of synaptonemal complex length and mammalian meiotic exchange rates', *Science* Vol. 296, pp. 2222–2225.
42. Kong, A., Barnard, J., Gudbjartsson, D.F. *et al.* (2004), 'Recombination rate and reproductive success in humans', *Nat. Genet.* Vol. 36, pp. 1203–1206.
43. Lercher, M.J. and Hurst, L.D. (2002), 'Human SNP variability and mutation rate are higher in regions of high recombination', *Trends Genet.* Vol. 18, pp. 337–340.
44. Kaplan, N.L., Hudson, R.R. and Langley, C.H. (1989), 'The "hitchhiking effect" revisited', *Genetics* Vol. 123, pp. 887–899.
45. Charlesworth, B., Morgan, M.T. and Charlesworth, D. (1993), 'The effect of deleterious mutations on neutral molecular variation', *Genetics* Vol. 134, pp. 1289–1303.
46. Fullerton, S.M., Bernardo Carvalho, A. and Clark, A.G. (2001), 'Local rates of recombination are positively correlated with GC content in the human genome', *Mol. Biol. Evol.* Vol. 18, pp. 1139–1142.
47. Marais, G. (2003), 'Biased gene conversion: Implications for genome and sex evolution', *Trends Genet.* Vol. 19, pp. 330–338.
48. Eyre-Walker, A. (1993), 'Recombination and mammalian genome evolution', *Proc. R. Soc. Lond. B Biol. Sci.* Vol. 252, pp. 237–243.
49. IHGSC (2004), 'Finishing the euchromatic sequence of the human genome', *Nature* Vol. 431, pp. 931–945.
50. She, X., Jiang, Z., Clark, R.A. *et al.* (2004), 'Shotgun sequence assembly and recent segmental duplications within the human genome', *Nature* Vol. 431, pp. 927–930.
51. Bailey, J.A., Gu, Z., Clark, R.A. *et al.* (2002), 'Recent segmental duplications in the human genome', *Science* Vol. 297, pp. 1003–1007.
52. Roy, A.M., Carroll, M.L., Nguyen, S.V. *et al.* (2000), 'Potential gene conversion and source genes for recently integrated *Alu* elements', *Genome Res.* Vol. 10, pp. 1485–1495.
53. Fredman, D., White, S.J., Potter, S. *et al.* (2004), 'Complex SNP-related sequence variation in segmental genome duplications', *Nat. Genet.* Vol. 36, pp. 861–866.
54. Bailey, J.A., Liu, G. and Eichler, E.E. (2003), 'An *Alu* transposition model for the origin and expansion of human segmental duplications', *Am. J. Hum. Genet.* Vol. 73, pp. 823–834.
55. Zhou, Y. and Mishra, B. (2005), 'Quantifying the mechanisms for segmental duplications in mammalian genomes by statistical analysis and modeling', *Proc. Natl. Acad. Sci. USA* Vol. 102, pp. 4051–4056.
56. Lupski, J.R. (2004), 'Hotspots of homologous recombination in the human genome: Not all homologous sequences are equal', *Genome Biol.* Vol. 5, p. 242.
57. Murti, J.R., Bumbulis, M. and Schimenti, J.C. (1994), 'Gene conversion between unlinked sequences in the germline of mice', *Genetics* Vol. 137, pp. 837–843.
58. Neitz, M. and Neitz, J. (1995), 'Numbers and ratios of visual pigment genes for normal red-green color vision', *Science* Vol. 267, pp. 1013–1016.
59. Zimmer, E.A., Martin, S.L., Beverley, S.M. *et al.* (1980), 'Rapid duplication and loss of genes coding for the alpha chains of hemoglobin', *Proc. Natl. Acad. Sci. USA* Vol. 77, pp. 2158–2162.

60. Stankiewicz, P., Shaw, C.J., Withers, M. *et al.* (2004), 'Serial segmental duplications during primate evolution result in complex human genome architecture', *Genome Res.* Vol. 14, pp. 2209–2220.
61. Hurles, M.E. (2001), 'Gene conversion homogenizes the CMT1A paralogous repeats', *BMC Genomics* Vol. 2, p. 11.
62. Mefford, H.C., Linnardopoulou, E., Coil, D. *et al.* (2001), 'Comparative sequencing of a multicopy subtelomeric region containing olfactory receptor genes reveals multiple interactions between non-homologous chromosomes', *Hum. Mol. Genet.* Vol. 10, pp. 2363–2372.
63. Rozen, S., Skaletsky, H., Marszalek, J.D. *et al.* (2003), 'Abundant gene conversion between arms of palindromes in human and ape Y chromosomes', *Nature* Vol. 423, pp. 873–876.
64. Hurles, M.E., Willey, D., Matthews, L. and Hussain, S.S. (2004), 'Origins of chromosomal rearrangement hotspots in the human genome: Evidence from the AZFa deletion hotspots', *Genome Biol.* Vol. 5, p. R55.
65. Bosch, E., Hurles, M.E., Navarro, A. and Jobling, M.A. (2004), 'Dynamics of a human inter-paralog gene conversion hotspot', *Genome Res.* Vol. 14, pp. 835–844.
66. Estivill, X., Cheung, J., Pujana, M.A. *et al.* (2002), 'Chromosomal regions containing high-density and ambiguously mapped putative single nucleotide polymorphisms (SNPs) correlate with segmental duplications in the human genome', *Hum. Mol. Genet.* Vol. 11, pp. 1987–1995.
67. Bagnall, R.D., Ayres, K.L., Green, P.M. and Giannelli, F. (2005), 'Gene conversion and evolution of Xq28 duplicons involved in recurring inversions causing severe hemophilia A', *Genome Res.* Vol. 15, pp. 214–223.
68. Ohno, S. (1970), 'Evolution by Gene Duplication', George Allen & Unwin, London, UK.
69. Hurles, M. (2004), 'Gene duplication: The genomic trade in spare parts', *PLoS Biol.* Vol. 2, p. E206.
70. Emes, R.D., Goodstadt, L., Winter, E.E. and Ponting, C.P. (2003), 'Comparison of the genomes of human and mouse lays the foundation of genome zoology', *Hum. Mol. Genet.* Vol. 12, pp. 701–709.
71. Clark, A.G., Glanowski, S., Nielsen, R. *et al.* (2003), 'Inferring nonneutral evolution from human–chimp–mouse orthologous gene trios', *Science* Vol. 302, pp. 1960–1963.
72. Teich, N., Nemoda, Z., Kohler, H. *et al.* (2005), 'Gene conversion between functional trypsinogen genes PRSS1 and PRSS2 associated with chronic pancreatitis in a six-year-old girl', *Hum. Mutat.* Vol. 25, pp. 343–347.
73. Tayebi, N., Stubblefield, B.K., Park, J.K. *et al.* (2003), 'Reciprocal and nonreciprocal recombination at the glucocerebrosidase gene region: Implications for complexity in Gaucher disease', *Am. J. Hum. Genet.* Vol. 72, pp. 519–534.
74. Collier, S., Tassabehji, M., Sinnott, P. and Strachan, T. (1993), 'A *de novo* pathological point mutation at the 21-hydroxylase locus: Implications for gene conversion in the human genome', *Nat. Genet.* Vol. 3, pp. 260–265.
75. Torrents, D., Suyama, M., Zdobnov, E. and Bork, P. (2003), 'A genome-wide survey of human pseudogenes', *Genome Res.* Vol. 13, pp. 2559–2567.
76. Ingelman-Sundberg, M. (2005), 'Genetic polymorphisms of cytochrome P450 2D6 (CYP2D6): Clinical consequences, evolutionary aspects and functional diversity', *Pharmacogenomics J.* Vol. 5, pp. 6–13.
77. Flint, J., Hil, A.V., Bowden, D.K. *et al.* (1986), 'High frequencies of alpha-thalassaemia are the result of natural selection by malaria', *Nature* Vol. 321, pp. 744–750.
78. Gonzalez, E., Kulkarni, H., Bolivar, H. *et al.* (2005), 'The influence of *CCL3L1* gene-containing segmental duplications on HIV-1/AIDS susceptibility', *Science* Vol. 307, pp. 1434–1440.
79. Stefansson, H., Helgason, A., Thorleifsson, G. *et al.* (2005), 'A common inversion under selection in Europeans', *Nat. Genet.* Vol. 37, pp. 129–137.
80. Tuzun, E., Sharp, A.J., Bailey, J.A. *et al.* (2005), 'Fine-scale structural variation of the human genome', *Nat. Genet.* Vol. 37, pp. 727–732.
81. Sebat, J., Lakshmi, B., Troge, J. *et al.* (2004), 'Large-scale copy number polymorphism in the human genome', *Science* Vol. 305, pp. 525–528.
82. Iafrate, A.J., Feuk, L., Rivera, M.N. *et al.* (2004), 'Detection of large-scale variation in the human genome', *Nat. Genet.* Vol. 36, pp. 949–951.