Duplication and divergence in humans and chimpanzees

Stephen Wooding and Lynn B. Jorde*

Summary

It has become a truism that we humans are genetically about 99% identical to chimpanzees. The origins of this assertion are clear: among early studies of DNA sequences, nucleotide identity between humans and chimpanzees was found to average around 98.9%.[1] However, this figure is correct only with respect to regions of the genome that are shared between humans and chimpanzees. Often ignored are the many parts of their genomes that are not shared. Genomic rearrangements, including insertions, deletions, translocations and duplications, have long been recognized as potentially important sources of novel genomic material[2,3] and are known to account for major genomic differences between humans and chimpanzees.[4] Further, such changes have been implicated in a number of genetic disorders, such as DiGeorge, Angelman/Prader-Willi and Charcot-Marie-Tooth syndromes.[5] ***BioEssays*** **28:335–338, 2006.**

Many factors contribute to the location and frequency of genomic rearrangement; however, over the last few years, a class of elements called segmental duplications has emerged as being of particular importance.[6,7] SDs, or "low-copy repeats", are usually defined as copies of DNA that are 90–100% identical to one another and range from ~1 to > ~400 kb in size. Unlike *Alu* or L1 elements, segmental duplications do not form a single phylogenetic "family".[6,7] However, they are unified by the fact that they constitute regions of nearly identical sequence at different genomic locations, an attribute that can interfere with normal recombination and replication.[6,7] Accordingly, there has been substantial interest in mapping and characterizing these elements within the human genome to identify regions especially susceptible to rearrangement.[8–13] Comparisons of segmental duplications in humans and chimpanzees have been much anticipated because they might tell us about the origins of phenotypic difference between us and our closest relatives. These comparisons have been slower in coming, but hot on the heels of the publication of the chimpanzee genome,[14] Cheng et al.[15] have contributed an early installment of what will surely be an explosion in human–chimp comparative genomics.

Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT.
*Correspondence to: Lynn B. Jorde, Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT 84112.
E-mail: lbj@genetics.utah.edu
DOI 10.1002/bies.20385
Published online in Wiley InterScience (www.interscience.wiley.com).

Cheng et al.[15] compared whole-genome assemblies in humans and chimpanzees, along with whole-genome shotgun sequence data, to identify segmental duplications and deletions shared between, as well as unique to, humans and chimpanzees. Their approach allowed the identification of >90% of duplications greater than 20 kb in size. As predicted on the basis of smaller samples,[4] these data showed that chimpanzees and humans differ by approximately 4% when segmental duplications are considered in addition to nucleotide differences—a number that could climb much higher if smaller duplications or ubiquitiously repeated elements, such as transposons, are added to the equation. The old "99%" saw is clearly incorrect.

Cheng et al.'s comparisons revealed a total of approximately 80 Mb of segmentally duplicated DNA in humans and 65 Mb in chimpanzees. After taking genome size, depth of sequencing coverage and human–chimp differences in copy number into account, humans and chimpanzees were found to have similar levels of lineage-specific duplication: 1.35% and 1.33%, respectively. The similarity of these values suggests that chimpanzees, like humans, may have experienced an enrichment of segmental duplications relative to other taxa.[16–18] These lineage-specific changes are of considerable interest because they must have occurred after the evolutionary divergence of humans and chimpanzees, 5–10 million years ago. Thus, they are prime candidates in the search for genetic sources of phenotypic differences between the two species.

Interestingly, although humans and chimpanzees showed similar overall rates of segmental duplication, the number of affected genes in each species differed. While 177 genes and partial genes showed evidence of duplication only in humans, just 94 appeared to be duplicated only in chimpanzees. This finding raises tantalizing questions about the relationship between segmental duplication and evolutionary adaptation. One hypothesis to explain the pattern would be that, as ancestral populations dispersed into new environments—for example, as our early ancestors moved from forest to savannah habitats, or as anatomically modern humans dispersed out of Africa—new segmental duplications were "accepted" into the human genome at high rates in much the same way that new amino acid substitutions are adaptively accepted into protein-coding sequences.[19] Thus, the human

excess may represent a genome-wide signature of positive natural selection. Some support for this suggestion is found in the overrepresentation of "environmental sensor" genes (such as surface antigens and drug-metabolizing enzymes) in segmentally duplicated regions in humans,[18] for these are the genes predicted to respond most quickly to environmental change.[19] A simple, testable prediction of this hypothesis would be that young segmental duplications are more common in non-African populations, which likely faced pressure to adapt to the temperate environments of Asia and Europe some 100,000 years ago.[19]

Previous comparisons between humans and more distantly related taxa raised questions about the mechanisms through which the human–chimp lineage became enriched with segmentally duplicated regions. High rates of duplication, systematic deletion of existing duplications in other taxa and conversion of ancient duplications have all been recognized as possibilities.[9,20,21] Here again, Cheng et al's careful comparison with chimpanzees provides unique insights. Among 17 duplications found in chimpanzees but not humans, and for which data from other apes were available, only 6 were also found in gorilla and orangutan and thus attributable to deletion in humans, as opposed to insertion in chimpanzees. Further, just 3% of regions showed high levels of divergence consistent with gene conversion. Thus, most of the difference between humans and chimpanzees at segmentally duplicated loci is attributable to a process of de novo duplication.

The finding that de novo duplication has likely been important in the divergence of humans and chimpanzees brings focus to a new question: what is facilitating these duplication processes? One possibility is suggested by the presence of a particularly ubiquitous repeated component unique to primate genomes: *Alu* elements. Earlier studies of the relationship between segmental duplications and *Alu* elements revealed that *Alu*s are found preferentially near the ends of duplications: although *Alu* elements account for only about 10% of the human genome, 27% of segmental duplications terminate within an *Alu* element—a highly significant enrichment.[12] Further, while 7,000 new *Alu* elements have been inserted in humans since we diverged from chimpanzees some six million years ago, only 2,300 *Alu*s have been inserted in the chimpanzee lineage. Perhaps differences in *Alu* retrotransposition, by contributing to nonallelic homologous recombination, can explain part of the difference between primates and other taxa, or even between humans and chimpanzees.

Genomic rearrangements can have a variety of important functional effects. The emergence of duplicate genes has long been recognized as a mechanism of diversification in gene families, which can evolve when duplicates—under selective pressures differing from those of the parent—adopt novel functions.[22] Duplicated regions can also result in changes in the timing, place and level of gene expression.[23] This phenomenon is observable in the data of Cheng et al.: 56% of the human-only duplications were associated with expression changes in humans, and 87% of these were in an upward direction. Chimpanzees showed a similar pattern, with 49% of chimpanzee-only duplications being associated with expression changes, 57% of these being in an upward direction. Structural rearrangements can also have striking effects on rates and patterns of evolution within specific genomic regions. For instance, inversions can serve as focal points for rapid evolutionary divergence because they suppress recombination (and thus recombination-mediated DNA repair). Segmental duplications are also associated with gene conversion events, reaching an apogee on the Y chromosome, where a 45% segmental duplication content is accompanied by extensive gene conversion.[21]

The diverse range of effects that can arise following segmental duplications and deletions is well illustrated by patterns of variation at *CYP2D6*, one of the earliest recognized examples.[24] *CYP2D6* encodes a cytochrome P450 enzyme involved in the metabolism of approximately 50% of known drugs, including widely prescribed medications such as selective serotonin reuptake inhibitors, beta-blockers and codeine, along with numerous natural toxins.[25] Dramatic inter-individual variation in CYP2D6 metabolism was recognized as early as 1975,[24] and subsequent studies identified a number of nucleotide substitutions associated with variable CYP2D6 activity.[26,27] However, a much more dramatic source of variation emerged with the discovery that *CYP2D6* is often found in tandemly repeated arrays, which can effectively increase CYP2D6 production and hence the rate of CYP2D6 metabolism.[28] This finding explained the so-called "ultrarapid" metabolizers of CYP2D6 substrates: individuals harboring multiple (sometimes more than 10) functional copies of the gene, as well as "poor" metabolizers, with none.[29,30] Thus, variability in CYP2D6 metabolism is a joint function of allelic variation, arising through nucleotide substitution and changes in copy number, arising through segmental duplication and deletion.

Segmental duplications can have phenotypic effects through more complex genomic rearrangements, as well. More than 25 different "contiguous gene syndromes" are caused by deletions of regions that are flanked by segmental duplications.[5] A well-known example involves a 4 Mb region on chromosome 7 that is consistently deleted in the imprinted Prader-Willi and Angelman syndromes and is flanked by highly similar 500 kb segmental duplications (Fig. 1A). These segmental duplications are thought to mediate nonallelic homologous recombination, leading both to deletions and duplications of the intervening DNA sequence. The latter are associated with some cases of autism.[31] Similarly, more than 90% of patients with William-Beuren syndrome have a common 1.6 Mb deletion that is flanked by segmental duplications.[32] An interesting variation on this theme is an inversion that alters the factor VIII gene on the X chromosome, accounting for about half

of severe hemophilia A cases. Here, a sequence found in intron 22 is duplicated and inverted upstream of the factor VIII gene (Fig. 1B). The tip of Xq can loop back on itself, allowing homologous recombination between the inverted repeats. A crossover within the region flanked by the repeats can then produce a large inversion. Because the presence of homologous X chromosomes in females tends to prevent looping, this mutation occurs mostly in male meiosis. With the recent identification of dozens of additional segmental duplication-flanked candidate regions,[8] many more examples of segmental duplication-associated diseases are likely to emerge.

Cheng et al.'s newly analyzed map of segmental duplications in the human and chimpanzee genomes is but the first step in understanding the role of these important regions in the dynamics of genome organization on a large scale. By broadening our perspective on the magnitude of differences between humans and chimpanzees, yet focusing our attention on a limited number of regions susceptible to rearrangement, the identification of these regions will doubtless fuel interest in these lesser-known portions of the genome in both evolutionary and biomedical circles. Perhaps most importantly, these findings suggest that human–chimp similarity is closer to 96% than 99%, and estimates are likely to drop further as smaller and more numerous segmental rearrangements are discovered. Measures of similarities and differences within and between human populations, which will yield a host of additional insights, remain in their earliest stages but seem likely to be headed in the same direction.[8,13,33]

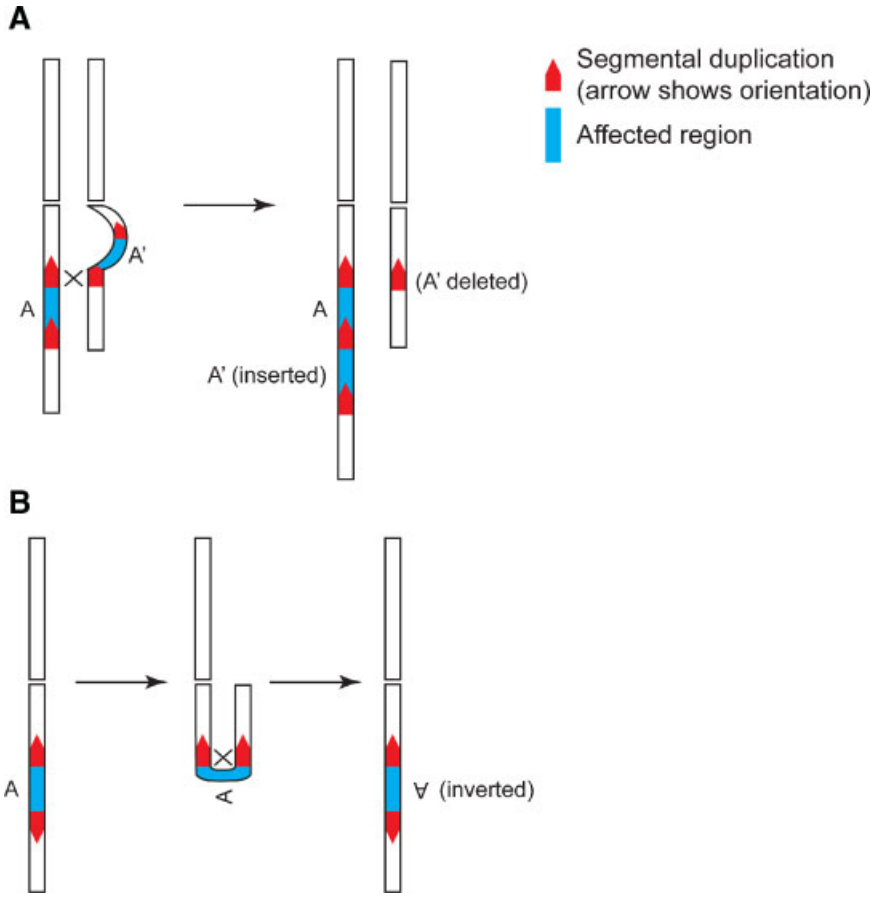

Figure 1. Examples of genomic rearrangement mediated by segmental duplications. In each panel, red arrows indicate segmental duplications and their directional orientation. Blue boxes indicate regions affected by rearrangement. **A:** Rearrangement underlying Prader-Willi and Angelman syndromes. Segmental duplications in different regions of the genome can cause misalignment, allowing nonallelic homologous recombination to occur. This can result in duplication of a region on one copy of the chromosome pair, but deletion on the homologue. **B:** Rearrangement underlying some cases of hemophilia A. Segmental duplications near each other, but in opposite orientations, can allow hairpins to form so that recombination of a chromosome with itself causes an inversion. Detailed information about this rearrangement is given by Lakich et al.[34]

References

1. Goodman M, Grossman LI, Wildman DE. 2005. Moving primate genomics beyond the chimpanzee genome. Trends Genet 21:511–517.
2. Bridges CB. 1935. Salivary chromosome maps. Journal of Heredity 26:60–64.
3. Muller HJ. 1936. Bar duplication. Science 83:528–530.
4. Britten RJ. 2002. Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels. Proc Natl Acad Sci USA 99:13633–13635.
5. Stankiewicz P, Lupski JR. 2002. Genome architecture, rearrangements and genomic disorders. Trends Genet 18:74–82.
6. Emanuel BS, Shaikh TH. 2001. Segmental duplications: an 'expanding' role in genomic instability and disease. Nat Rev Genet 2: 791–800.
7. Samonte RV, Eichler EE. 2002. Segmental duplications and the evolution of the primate genome. Nat Rev Genet 3:65–72.
8. Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, et al. 2005. Segmental duplications and copy-number variation in the human genome. Am J Hum Genet 77:78–88.
9. Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, et al. 2002. Recent segmental duplications in the human genome. Science 297:1003–1007.
10. Zhang L, Lu HH, Chung WY, Yang J, Li WH. 2005. Patterns of segmental duplication in the human genome. Mol Biol Evol 22:135–141.
11. Bailey JA, Yavor AM, Massa HF, Trask BJ, Eichler EE. 2001. Segmental duplications: organization and impact within the current human genome project assembly. Genome Res 11:1005–1017.
12. Bailey JA, Liu G, Eichler EE. 2003. An Alu transposition model for the origin and expansion of human segmental duplications. Am J Hum Genet 73:823–834.
13. Fortna A, Kim Y, MacLaren E, Marshall K, Hahn G, et al. 2004. Lineage-specific gene duplication and loss in human and great ape evolution. PLoS Biol 2:E207.
14. The Chimpanzee Sequencing and Analysis Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437:69–87.
15. Cheng Z, Ventura M, She X, Khaitovich P, Graves T, et al. 2005. A genome-wide comparison of recent chimpanzee and human segmental duplications. Nature 437:88–93.
16. Bailey JA, Church DM, Ventura M, Rocchi M, Eichler EE. 2004. Analysis of segmental duplications and genome assembly in the mouse. Genome Res 14:789–801.
17. Marques AC, Dupanloup I, Vinckenbosch N, Reymond A, Kaessmann H. 2005. Emergence of young human genes after a burst of retroposition in primates. PLoS Biol 3:e357.
18. Tuzun E, Bailey JA, Eichler EE. 2004. Recent segmental duplications in the working draft assembly of the brown Norway rat. Genome Res 14:493–506.
19. Bamshad M, Wooding SP. 2003. Signatures of natural selection in the human genome. Nat Rev Genet 4:99–111.
20. International Human Genome Sequencing Consortium. 2004. Finishing the euchromatic sequence of the human genome. Nature 431:931–945.
21. Rozen S, Skaletsky H, Marszalek JD, Minx PJ, Cordum HS, et al. 2003. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. Nature 423:873–876.
22. Ohno S. 1970. Evolution by gene duplication. New York, NY: Springer-Verlag. p. 160.
23. Li WH, Yang J, Gu X. 2005. Expression divergence between duplicate genes. Trends Genet 21:602–607.
24. Mahgoub A, Idle JR, Dring LG, Lancaster R, Smith RL. 1977. Polymorphic hydroxylation of Debrisoquine in man. Lancet 2:584–586.
25. Ingelman-Sundberg M. 2005. Genetic polymorphisms of cytochrome P450 2D6 (CYP2D6): clinical consequences, evolutionary aspects and functional diversity. Pharmacogenomics J 5:6–13.
26. Bradford LD. 2002. CYP2D6 allele frequency in European Caucasians, Asians, Africans and their descendants. Pharmacogenomics 3:229–243.
27. Skoda RC, Gonzalez FJ, Demierre A, Meyer UA. 1988. Two mutant alleles of the human cytochrome P-450db1 gene (P450C2D1) associated with genetically deficient metabolism of debrisoquine and other drugs. Proc Natl Acad Sci USA 85:5240–5243.
28. Ingelman-Sundberg M. 1999. Duplication, multiduplication, and amplification of genes encoding drug-metabolizing enzymes: evolutionary, toxicological, and clinical pharmacological aspects. Drug Metab Rev 31:449–459.
29. Dalen P, Dahl ML, Ruiz ML, Nordin J, Bertilsson L. 1998. 10-Hydroxylation of nortriptyline in white persons with 0, 1, 2, 3, and 13 functional CYP2D6 genes. Clin Pharmacol Ther 63:444–452.
30. Johansson I, Lundqvist E, Bertilsson L, Dahl ML, Sjoqvist F, et al. 1993. Inherited amplification of an active gene in the cytochrome P450 CYP2D locus as a cause of ultrarapid metabolism of debrisoquine. Proc Natl Acad Sci USA 90:11825–11829.
31. Folstein SE, Rosen-Sheidley B. 2001. Genetics of autism: complex aetiology for a heterogeneous disorder. Nat Rev Genet 2:943–955.
32. Bayes M, Magano LF, Rivera N, Flores R, Perez Jurado LA. 2003. Mutational mechanisms of Williams-Beuren syndrome deletions. Am J Hum Genet 73:131–151.
33. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, et al. 2004. Large-scale copy number polymorphism in the human genome. Science 305:525–528.
34. Lakich D, Kazazian HH Jr, Antonarakis SE, Gitschier J. 1993. Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A. Nat Genet 5:236–241.