Structural Variation in Great Ape Genomes

Jan A Aerts, *Wellcome Trust Sanger Institute, Hinxton, United Kingdom*

Chris Tyler-Smith, *Wellcome Trust Sanger Institute, Hinxton, United Kingdom*

Advanced article

Article Contents

Online posting date: 15th December 2009

Structural variation within and between Great Ape genomes affects more nucleotides than single-base variation, yet its extent and phenotypic consequences are much less well understood. The most-studied structural variants are copy number variations (CNVs) which can be generated by several different mechanisms including non-allelic homologous recombination, non-homologous end-joining and deoxyribonucleic acid (DNA) replication-related fork stalling and template switching. CNVs are closely related to segmental duplications (SDs): SDs can stimulate the formation of CNVs and themselves started out as CNVs, but became fixed in a species. Structural variation can be neutral but has also influenced our phenotypic evolution, for example our susceptibility to disease and our ability to digest certain types of food. Our understanding of the extent of structural variation is increasing rapidly, but it will be much more difficult to understand its phenotypic consequences.

Introduction

In the past few years, structural variation has been recognized as being responsible for much of the genetic variation within and between species. The Great Apes or hominidae are part of the primate order which also harbours lemurs, monkeys and lesser apes. The Great Apes (orangutans, gorillas, chimpanzees, bonobos and humans) are a small group, but one of particular interest to us because they are our closest living relatives – to some, we are the third chimpanzee. The orangutan lineage split from our common ancestor about 15–21 million years ago (mya); gorilla and chimpanzee/bonobo are estimated to have split off 8–10 and 5–8 mya, respectively.

Has genomic structural variation contributed to the phenotypic differences between humans and other apes since their split? Although the answer to this question seems likely to be 'yes', data on the extent of structural variation are only just beginning to become available, and little is known about their phenotypic consequences, so here we review an emerging field where there are more questions than answers. **See also**: Hominids: Molecular Phylogenetics; Nucleotide Sequence Divergence between Humans and Chimpanzees

ELS subject area: Evolution and Diversity of Life

How to cite:
Aerts, Jan A; and, Tyler-Smith, Chris (December 2009) Structural Variation in Great Ape Genomes. In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester.
DOI: 10.1002/9780470015902.a0021768

What is Structural Variation?

The term 'structural variation' can be defined in several ways; here, we use it to refer to medium- to large-scale rearrangements of the genome, excluding small-scale rearrangements like indels, microsatellites or short tandem repeats (STRs) and minisatellites. At the most general level, structural variations can be divided into two groups: balanced and unbalanced variations (see **Figure 1**). Balanced variation includes those rearrangements where the deoxyribonucleic acid (DNA) is restructured but the total content is not changed. Examples include inversions and translocations, where a DNA segment has changed in orientation or moved to another part of the genome. Duplications and deletions, in contrast, are examples of unbalanced variation, and do change the total DNA content. Because they change the number of copies of DNA segments, unbalanced variations are also called copy number variations or CNVs. In addition to simple duplications and deletions, more complex structures classified as deletion + duplication events or multiallelic CNVs are also

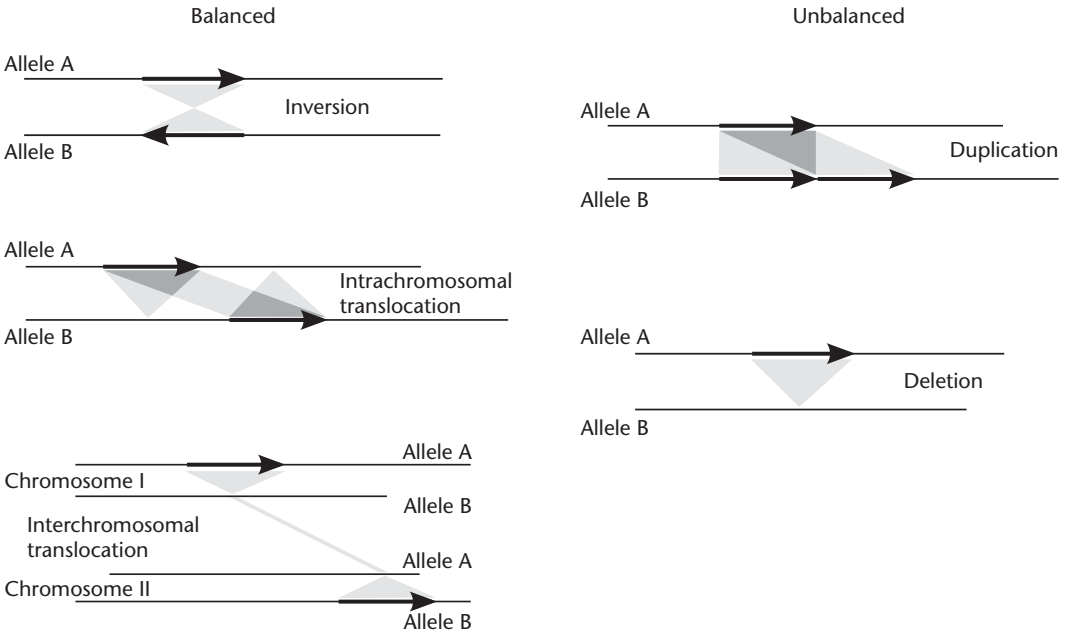

Figure 1 Types of structural variation. Each section shows two alleles that differ because of structural variation. The thin line represents the nonvariable DNA and the thick arrow the variable section. For the translocation, two chromosomes are shown; individuals with allele A have the variant locus inserted in chromosome I, whereas individuals with allele B have it inserted in chromosome II.

found. In total, CNVs affect an appreciable part of the human genome, with estimates going up to 12% (Redon *et al*., 2006).

Large chromosomal rearrangements exist between human and other apes, most notably the fusion of two ape chromosomes that gave rise to human chromosome 2, and nine pericentric inversions between human and chimpanzee. These rearrangements are detectable in metaphase chromosome preparations under the microscope and have been extensively described and discussed elsewhere (e.g. Kehrer-Sawatzki and Cooper, 2008a). Here we focus on submicroscopic rearrangements, in particular CNVs. This is because CNVs have been shown to be common and are easier to study than other kinds of submicroscopic structural variation, rather than because they have necessarily been more important in Great Ape evolution. **See also**: Chromosomal Rearrangements in Primates; Chromosomal Rearrangements in the Human and Chimpanzee Lineages; Chromosome Rearrangement Patterns in Mammalian Evolution; Copy Number Variation in the Human Genome

Mechanisms of Copy Number Variation

Three major mechanisms have been described to date that generate the CNV considered in this article: nonallelic homologous recombination (NAHR), nonhomologous end-joining (NHEJ) and the DNA replication-based fork stalling and template switching (FoSTeS). In addition to these, CNVs can also be generated by retrotransposition, although less frequently (Gu *et al*., 2008). **See also**: Structural Diversity of the Human Genome and Disease Susceptibility

Segmental duplications (SDs) – also known as low-copy repeats or LCRs – are genomic regions with more than one copy that are at least 1 kb long and have more than 90% sequence identity. Owing to this sequence similarity, these regions sometimes misalign during meiosis or mitosis leading to NAHR and consequently genomic rearrangements. NHEJ does not need this sequence identity and is one of the mechanisms used by eukaryotic cells to repair double-strand breaks. The FoSTeS mechanism is replication-based and involves detachment of one of the replicating strands from its replication fork and reattachment to a nearby fork. **See also**: Chromosome-specific Repeats (Low-copy Repeats)

These different mechanisms have important consequences for the type of structural variation they produce. As NAHR is associated with SDs, NAHR-mediated CNVs can recur and breakpoints consequently tend to be clustered; they are also, on average, larger. NHEJ-mediated structural variations, in contrast, are not associated with SDs so are more scattered, have unique origins and are smaller. Less is known about FoSTeS-mediated CNVs, but they may resemble NHEJ-mediated CNVs more than NAHR-mediated ones. The FoSTeS process can occur multiple times in series and can thereby generate very complicated rearrangements. Early studies of CNVs could only identify large ones, so led to the view that most originated by NAHR, but now that it is possible to identify

smaller CNVs, the importance of the other mechanisms is becoming better appreciated. For a more detailed description of these mechanisms, see Gu *et al*. (2008).

Copy Number Variation and Segmental Duplications

Research by Newman *et al*. (2005) comparing human and chimpanzee indicated that 70–80% of the inversions and 40% of the insertions/deletions overlap with segmental duplications. It will therefore be appreciated that CNVs are closely related to segmental duplications. In addition to SDs acting as substrates for the formation of new CNVs, they themselves originate as CNVs that have become fixed in a population. Nevertheless, there is a distinction: SDs refer to loci within an individual haploid genome and do not have to be copy number variable; CNVs refer to locus differences between individuals and can vary between zero (i.e. deletion) and multiple copies. But the close relationship is important when considering structural variation in the Great Ape genomes. **See also**: Segmental Duplications and Their Role in the Evolution of the Human Genome

Sequence gains and losses generated by NAHR are highly enriched within the primate lineage. This contrasts with other types of genetic variation such as base changes, which actually evolve at a slower rate than in many other mammals because of the longer generation times of primates. In a comparison of rhesus macaque, orangutan, chimpanzee and human, Marques-Bonet *et al*. (2009) showed that approximately 73 Mb of sequence is duplicated in one or more of these species. It was suggested that a primate-specific burst of Alu repeat formation 35–40 mya predisposed the genome to Alu–Alu-mediated recombination leading to a major increase in NAHR and the creation of SDs (Bailey *et al*., 2003; Sharp and Eichler, 2006). The duplicated sequences did not, however, arise uniformly over the course of Great Ape evolution. Though 80% of the human segmental duplications arose after the divergence of the Great Ape lineage from the rest of the primates, the most significant burst of duplication activity was situated around the time of the divergence of gorilla from the human/chimpanzee ancestor (Marques-Bonet *et al*., 2009). Duplication activity then slowed down after the divergence of the human and chimpanzee lineages. **See also**: Transposable Element-driven Duplications during Hominoid Genome Evolution

CNVs, CNVRs, CNVEs and CNDs

Technologies for CNV discovery currently lack resolution in defining the exact base pair-resolution boundaries of CNVs in a sample. As many samples are often used to identify a CNV, the boundaries found for the same locus in different samples are often not exactly the same. As a result, we end up with a family of CNVs (one for each sample) at the same locus, but with slightly different boundaries. Much of the literature, therefore, refers to CNV regions or CNVRs (**Figure 2**), which are the unions of overlapping CNVs; if two CNVs partly overlap, they are considered one CNVR. In some cases – for example, the left CNVR in **Figure 2** – such CNVRs can be split into overlapping CNV events (CNVEs), which more probably reflect the underlying real duplication events. A common rule for splitting a CNVR into its constituent CNVEs is by grouping those CNVs that have a reciprocal overlap of at least 50%. Thus in the literature the reader will sometimes find CNVRs or CNVEs, although as our understanding increases, these terms should become superfluous. For convenience, we will distinguish between CNVs in general, which are variable within a species, and copy number differences (CNDs) which differ between species (**Figure 3**). A region can be both a CNV and CND.

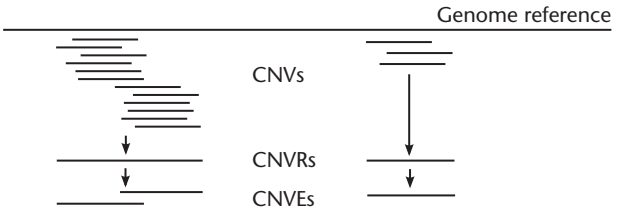

Figure 2 Copy number variations (CNVs), copy number variation regions (CNVRs) and CNV events (CNVEs). The extent of each CNV is determined by comparing two individuals, so that there may be a number of overlapping but not identical CNVs at the same locus. All overlapping CNVs are grouped into a single CNV region or CNVR. A CNVR can be split into two or more CNV events (CNVEs) by grouping its constituent CNVs according to a minimal reciprocal overlap (e.g. 50%).

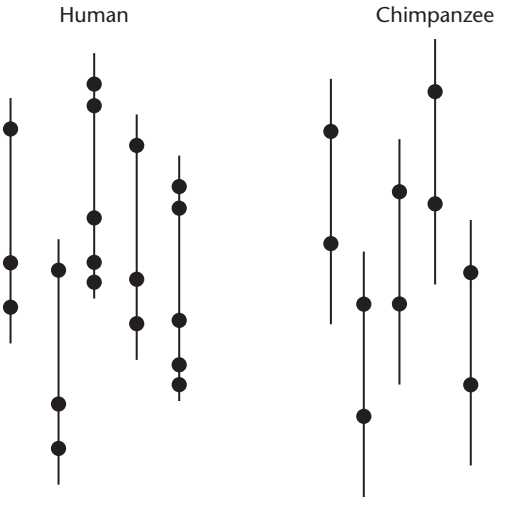

Figure 3 Copy number variation (CNV) compared with copy number difference (CND). Lines represent different copies of the same chromosome within the species; dots represent copies of a locus. In this figure, human chromosomes contain either 3 or 5 copies of the locus, resulting in CNV within the population. The chimpanzee chromosomes all carry two copies and so do not show variability within the species; the locus is therefore not a CNV. However, the copy number within chimpanzee (2) is different from the copy number in human (3 or 5) so the locus constitutes a CND between these species.

Consequences of Copy Number Variations

Both balanced and unbalanced structural variation can influence the phenotype of an individual and are likely to have been among the genetic changes underlying speciation. It has been shown that up to 18% of heritable gene expression variation can be explained by CNVs (Stranger *et al*., 2007). Moreover, Lupski (2007) reports that almost a third of all human genes exhibit a CNV in one or more primate species.

There are several ways in which CNVs can exert a phenotypic effect. First, and most obviously, gene-containing CNVs can duplicate (gain) or delete (lose) members of the genepool and therefore alter transcript levels in the cell. It might be expected that more copies would lead to more protein product, but additional mechanisms may complicate the consequences. Alternatively, increased copy number of a repressor element could lead to reduction in protein level. Copy number changes involving dosage-sensitive genes are a significant cause of disease in humans, referred to as genomic disorders. Note that structural variations do not have to be unbalanced to have an effect on phenotype. Balanced structural variation can disrupt and inactivate genes, fuse portions of different genes or alter regulation. **See also**: Gene Deletions in Evolution; Indels in the Evolution of the Human and Chimpanzee Genomes; Relevance of Copy Number Variation to Human Genetic Disease

On the evolutionary scale, gene duplication has been one of the driving forces of speciation by generating gene 'spare parts' that can potentially evolve to perform new or more specialized functions (Hurles, 2004). Three possible fates exist following gene duplication: the duplicated gene might be degraded by mutation; it might evolve a new function (neofunctionalization) or it might take over part of the function of the original copy (subfunctionalization) (Hurles, 2004). The importance of these gains and losses was highlighted in a gene-centric study by Fortna *et al*. (2004) who demonstrated that approximately 3% of our genes have undergone lineage-specific copy number changes among five hominoid species. In addition, Demuth *et al*. (2006) have shown that over longer evolutionary timescales more than half of the gene families are influenced by duplication or loss in at least one lineage of human, chimpanzee, mouse, rat or dog. **See also**: Human-specific Changes of Genome Structure

The high abundance of copy-number variations in the human lineage has led to a gene turnover rate (i.e. genes gained or lost), that is 2.5 times higher in humans than the average for mammalian species. Approximately 180 gene families show significant levels of expansion or contraction compared to other species. Several of these also show evidence for positive selection at the amino acid level, suggesting that selection has been one of the driving forces in the creation of CNDs (Hahn *et al*., 2007). In addition, the expansion of large gene families is not due to a small number of large SDs containing many related genes, but rather to many SDs each containing only one or a few genes (Hahn *et al*., 2007). This also supports a role for positive selection, rather than chance, in the formation of CNDs. Several studies have pointed towards a common set of gene families that are expanded in humans compared to other primate species. These include genes involved in inflammatory response, such as the immunoglobulin heavy chain variable region gene family.

Locations of Copy Number Variations

CNVs are not equally distributed over the genome, but show a higher prevalence in the pericentromeric and subtelomeric regions. Pericentromeric regions, in particular, are highly variable within the primates – even at the cytogenetic level. A two-stage model has been proposed for the generation of these duplications (Sharp and Eichler, 2006). First, in a series of seeding events, one or more progenitor loci transpose together to a pericentromeric acceptor. This generates a mosaic block of duplicated segments derived from different loci. It is sometimes possible to identify one copy far removed from a centromere, which is most likely the progenitor locus. Second, inter- and intrachromosomal duplication occurs, so that these large blocks are duplicated near other centromeres. These events may be evolutionarily neutral, but a more interesting possibility is that some may have been advantageous and thus driven by positive selection. In contrast to the pericentromeric rearrangements, the CNV enrichment near telomeres is likely to be a consequence of normal recombination: as cross-overs occur between the ends of chromosomes, distal sequences are translocated between chromosomes (Xue and Tyler-Smith, 2008). It seems that pericentromeric and subtelomeric regions are highly permissive for inter-chromosomal rearrangements, whereas intrachromosomal duplications tend to take place in the interstitial regions of the chromosome (Kehrer-Sawatzki and Cooper, 2008b). In a phenomenon called duplication shadowing, existing segmental duplications often act as a birth place for new ones, leading to clusters containing duplications with a wide age range and known as duplication hubs or acceptor regions. **See also**: Structural Diversity of the Human Genome and Disease Susceptibility

Lineage-specific and Shared SDs and CNVs

As expected, lineage-specific SDs were created relatively recently and are highly copy number variable. One might expect that shared SDs, which originated earlier in evolution, would tend to be more often fixed than lineage-specific SDs. However, that appears not to be the case, and the number of fixed shared SDs is lower than expected (Marques-Bonet *et al*., 2009). As a result, a significant proportion of human and chimpanzee SDs are also CNVs

and about a fifth of all CNVs found in either is shared between them. Moreover, CNVs with a high minor allele frequency in one species are often also common in the other.

Orthologous loci can be variable in both human and chimpanzee and are called shared CNVs. Some of these (e.g. in the MHC region) may represent variation created before the divergence of human and chimpanzee, and are maintained through extreme and long-term balancing selection. This is, however, a rare phenomenon as neutral polymorphisms tend to be lost by genetic drift and few that were present in our common ancestor would still be shared today (Perry *et al*., 2006). Instead, shared CNVs often overlap with inherently unstable SDs and are often not identical-by-descent, but created by recurrent NAHR at those duplications. This has led to the independent formation of CNVs in both species after the two lineages split. Interestingly, the CNV content, like the single nucleotide polymorphism (SNP) diversity, seems to be higher in chimpanzees than in humans (Perry *et al*., 2006) and provides another source of genetic diversity within the chimpanzee species (Chimpanzee Sequencing and Analysis Consortium, 2005).

A recent survey of human and chimpanzee CNVs and CNDs by Perry *et al*. (2008) identified 70 autosomal CNVs per within-chimpanzee comparison and 80 per within-human comparison, resulting in 438 CNVRs in chimpanzee and 353 in human. Of these, 144 were shown to overlap (i.e. are shared CNVs). In addition, they found 355 autosomal CNDs, approximately 75% of which overlap with a CNVR in one or both species. The remaining 25% represent fixed regions and thus some of them might have played key roles in the speciation between human and chimpanzee.

It is estimated that approximately 70 Mb of sequence has been differentially duplicated between human and chimpanzee (Cheng *et al*., 2005). A significant proportion are shown to result in gene expression differences between the two species.

Lineage-specific hyperexpansions have been observed in many species, including gorilla (Babcock *et al*., 2007), chimpanzee (Cheng *et al*., 2005) and human (Sikela, 2006). A 36-kb region orthologous to the human chromosome 2 fusion locus is hyperexpanded in chimpanzee subtelomeres; whereas humans contain only four copies of this region, chimpanzee have more than 400, causing its genome to grow more than 14 Mb since its split from the human lineage.

Similarly, a lineage-specific hyperexpansion is found in gorilla involving the segmental duplication LCR22. This duplicated locus is orthologous to human region 22q11.2. Nonallelic homologous recombination in this locus is associated with the diGeorge syndrome (Babcock *et al*., 2007). The fact that such a region is extremely duplicated in another species might give clues as to the understanding of this syndrome, revealing copy number-tolerant segments.

It has to be noted that this field of research is still young and there has been a bias towards the discovery and investigation of large CNVs/CNDs. Estimates for the number of CNVs and CNDs within and between species differ between studies and many more await discovery (see also Locke *et al*., 2003; Frazer *et al*., 2003; Wilson *et al*., 2006).

Examples

Here we will consider some specific examples of genes that are under the influence of CNVs and/or CNDs.

AMY1 and diet

Many plant species store their energy reserves as the polysaccharide starch, which is an important source of carbohydrates for several primates. The enzyme amylase – produced in saliva and pancreas – initiates digestion of this polymer into disaccharide sugars.

The salivary amylase gene *AMY1* is present in different copy numbers between and within different primate species. Although chimpanzees have only two copies per diploid genome, human copy number ranges from 2 to approximately 14. The average human has roughly three times as many *AMY1* gene copies as chimpanzees (Perry *et al*., 2007). Although bonobo does have a higher *AMY1* copy number than chimpanzee, the coding sequence of these extra copies is disrupted, rendering these genes nonfunctional.

Interestingly, *AMY1* CNDs are associated with the typical diets of humans and chimpanzees. The chimpanzee is a frugivorous animal, taking most of its sugars from fruit and ingesting little starch compared to most human populations. The human species however exhibits a variety of traditional diets, ranging, for example, from African rainforest hunter-gatherers with a low-starch diet to agriculturalists and hunter-gatherers who rely on tubers and roots and therefore have a high-starch diet. These different diets are reflected in the *AMY1* gene copy number: higher mean copy numbers are present in those populations that need to digest more starch. This is, however, not the complete story. Humans with a low starch diet still have higher copy number than chimpanzees, perhaps reflecting a more complex diet in extinct human ancestors.

Furthermore, the Old World monkey macaque has an even higher number of *AMY1* copies than human. This might be related to the fact that, uniquely in this subfamily of primates, food is stored in cheek pouches where salivary digestion takes place (Perry *et al*., 2007). These observations – together with enzyme level measurements – indicate that *AMY1* copy number is largely consistent with a history of diet-related selection pressures. Selection for increased expression has led to increases in copy number rather than upregulation of a single gene by other mechanisms.

CCL3L1 and HIV

CCL3L1 is one of the most-studied examples of a gene whose natural population variation in copy number may

have significant phenotypic consequences. It encodes a chemokine that plays an important role in leukocyte localization during infection and has come to the forefront of genetic research because one if its receptors, CCR5, acts as a cofactor for human immunodeficiency virus (HIV).

Chemokines are a family of low-molecular-weight polypeptides that signal through transmembrane receptors. Approximately 50 chemokine genes are known at present, which are divided into 4 families. The beta-family (which includes *CCL3L1* and *CCL4L1*) is mostly located at 17q11.2-17q12 but contains some genes on chromosomes 2, 7, 9 and 16. Both *CCL3L1* and *CCL4L1* are duplicated and exhibit variability in their copy number.

The amount of CCR5 receptor available on the leukocyte membrane greatly influences docking of HIV. Individuals who are homozygous for a *CCR5* deletion allele and consequently lack the cell-surface expression of *CCR5* are highly resistant to infection by HIV. CCL3L1 also binds to CCR5, inhibiting docking of HIV to the CCR5 receptor, although it is not clear yet what the underlying mechanism is. Binding of CCL3L1 to CCR5 might cause allosteric hindrance of HIV binding; it might cause the CCR5 receptor to internalize into the leukocyte; or the amount of CCL3L1 might have an effect on the localization of leukocytes.

Humans can have between 0 and approximately 10 copies of the *CCL3L1* gene. A lower number of copies is associated with a higher susceptibility to HIV infection, but interestingly the exact number of copies is not the best indicator. It is rather the relative copy number compared to the population average that influences susceptibility (Gonzalez *et al*., 2005). Human populations differ significantly in their *CCL3L1* copy number: the median copy number in Africans is six whereas that in Europeans and Asians is only three. With each unit of increase in *CCL3L1* copy number above that median, there is a dose-dependent, step-wise decrease in HIV susceptibility and risk of progressing rapidly to acquired immunodeficiency syndrome (AIDS). In each population, copy numbers higher than the median provide protection against HIV. This phenomenon is not only observed in humans, but also in other primates. Two distinct populations of rhesus macaque – with an Indian or a Chinese origin – also exhibit very different within-species median copy numbers of 9 and 12, respectively. As in humans, it is the copy number relative to the population median that is an indicator of susceptibility to *simian immunodeficiency virus* or SIV and its rate of progression to AIDS (Degenhardt *et al*., 2009).

However, recent work by Perry *et al*. (2008) contradicts some of these findings and, while confirming CNV within humans, suggests that the *CCL3L1* locus is not copy number variable within chimpanzees and that there is no human–chimpanzee CND. Different methodologies for CNV detection were used in each of these studies; the region is structurally complex and further research will be necessary to understand the *CCL3L1* locus and its relationship to HIV susceptibility more completely.

Inflammatory-response genes

As part of a genome-wide study of CNVs and CNDs in chimpanzees and humans, Perry *et al*. (2008) investigated the functional categories (gene ontology) of genes that are part of these variations. Strikingly, they found a set of CNDs that involve inflammatory-response genes and are all fixed losses in chimpanzee compared to human. As a result, the *APOL1* (apolipoprotein L1), *APOL4*, *CARD18* (caspase recruitment domain family, member 18), *IL1F7* and *IL1F8* (interleukin 1 family, member 8) genes are completely missing from the chimpanzee genome. The specific functional role of APOL4 is unclear, but APOL1 has been proposed to be the lytic factor responsible for resistance against the *Trypanosoma brucei* parasite, which causes sleeping sickness in human. The *CARD18* and *IL1F7* genes are part of a pathway for inflammatory response involving CASP1 (caspase 1). The absence of these genes from the chimpanzee genome suggests that the response of chimpanzees to parasites and inflammation may differ significantly from that in humans. This intriguing possibility is consistent with the general rapid evolution of immune-related genes.

Future Directions

The importance of structural variation, and CNVs in particular, in healthy individuals has only become fully appreciated in recent years, and research into CNDs between primate species and their biological consequences is only in its infancy. The resolution obtained when investigating CNVs has until recently been very coarse, and the underlying event (i.e. duplication in one individual or deletion in another) that led to a CNV could often not be determined. Fortunately, new approaches to CNV discovery, involving high-resolution oligonucleotide arrays and whole-genome resequencing, are changing this which will have a major impact on this field of research. Looking at more individuals – both humans and nonhuman primates – will also increase the power for CNV detection. Current efforts in CNV research are very much focussed on the detection and description of CNV events. The next step will be to link that information to disease as well as to phenotypical and/or cultural differences between humans and our closest relatives, the other Great Apes. **See also**: Copy Number Variation in the Human Genome; Segmental Duplications and Their Role in the Evolution of the Human Genome

Acknowledgment

Our work is supported by The Wellcome Trust.

References

Babcock M, Yatsenko S, Hopkins J *et al*. (2007) Hominoid lineage specific amplification of low-copy repeats on 22q11.2 (LCR22s)

associated with velo-cardo-facial/digeorge syndrome. *Human Molecular Genetics* **16**: 2560–2571.

Bailey JA, Liu G and Eichler EE (2003) An Alu transposition model for the origin and expansion of human segmental duplications. *American Journal of Human Genetics* **73**: 823–834.

Cheng Z, Ventura M, She X *et al*. (2005) A genome-wide comparison of recent chimpanzee and human segmental duplications. *Nature* **437**: 88–93.

Chimpanzee Sequencing and Analysis Consortium (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**: 69–87.

Degenhardt JD, de Candia P, Chabot A *et al*. (2009) Copy number variation of CCL3-like genes affects rate of progression to simian-AIDS in Rhesus macaques (*Macaca mulatta*). *PLoS Genetics* **5**: e1000346.

Demuth JP, Bie TD, Stajich JE, Cristianini N and Hahn MW (2006) The evolution of mammalian gene families. *PLoS ONE* **1**: e85.

Fortna A, Kim Y, MacLaren E *et al*. (2004) Lineage-specific gene duplication and loss in human and Great Ape evolution. *PLoS Biology* **2**: 937–954.

Frazer KA, Chen X, Hinds DA *et al*. (2003) Genomic DNA insertions and deletions occur frequently between humans and nonhuman primates. *Genome Research* **13**: 341–346.

Gonzalez E, Kulkarni H, Bolivar H *et al*. (2005) The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* **307**: 1434–1440.

Gu W, Zhang F and Lupski JR (2008) Mechanisms for human genomic rearrangements. *PathoGenetics* **1**: 4.

Hahn MW, Demuth JP and Han SG (2007) Accelerated rate of gene gain and loss in primates. *Genetics* **177**: 1941–1949.

Hurles M (2004) Gene duplication: the genomic trade in spare parts. *PLoS Biology* **2**: 900–904.

Kehrer-Sawatzki H and Cooper DN (2008a) Chromosomal rearrangements in the human and chimpanzee lineages. *Encyclopedia of Life Sciences* DOI:10.1002/9780470015902.a0020738.

Kehrer-Sawatzki H and Cooper DN (2008b) Molecular mechanisms of chromosomal rearrangement during primate evolution. *Chromosome Research* **16**: 41–45.

Locke DP, Segraves R, Carbone L *et al*. (2003) Large-scale variation among human and great ape genomes determined by array comparative genomic hybridization. *Genome Research* **13**: 347–357.

Lupski JR (2007) An evolution revolution provides further revelation. *BioEssays* **29**: 1182–1184.

Marques-Bonet T, Kidd JM, Ventura M *et al*. (2009) A burst of segmental duplications in the genome of the African great ape ancestor. *Nature* **457**: 877–881.

Newman TL, Tuzun E, Morrison VA *et al*. (2005) A genome-wide survey of structural variation between human and chimpanzee. *Genome Reseach* **15**: 1344–1356.

Perry GH, Dominy NJ, Claw KG *et al*. (2007) Diet and the evolution of human amylase gene copy number variation. *Nature Genetics* **39**: 1256–1260.

Perry GH, Tchinda J, McGrath SD *et al*. (2006) Hotspots for copy number variation in chimpanzees and humans. *Proceedings of the National Academy of Sciences of the USA* **103**: 8006–8011.

Perry GH, Yang F, Marques-Bonet T *et al*. (2008) Copy number variation and evolution in humans and chimpanzees. *Genome Research* **18**: 1698–1710.

Redon R, Ishikawa S, Fitch KR *et al*. (2006) Global variation in copy number in the human genome. *Nature* **444**: 444–454.

Sharp A and Eichler EE (2006) Segmental duplications. In: Lupski JR and Stankiewicz P (eds) *Genomic Disorders: The Genomic Basis of Disease*, pp. 73–88. Totowa, NJ: Humana Press.

Sikela JM (2006) The jewels of our genome: the search for the genomic changes underlying the evolutionarily unique capacities of the human brain. *PLoS Genetics* **2**: e80.

Stranger BE, Forrest MS, Dunning M *et al*. (2007) Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**: 848–853.

Wilson GM, Flibotte S, Missirlis PI *et al*. (2006) Identification by full-coverage array CGH of human DNA copy number increases relative to chimpanzee and gorilla. *Genome Research* **16**: 173–181.

Xue Y and Tyler-Smith C (2008) Segmental duplications and their role in the evolution of the human genome. *Encyclopedia of Life Sciences* DOI:10.1002/9780470015902.a0020838.

Further Reading

Lupski JR and Stankiewicz P (eds) (2006) *Genomic Disorders: The Genomic Basis of Disease*. Totowa, NJ: Humana Press.

Caldecott J and Miles L (eds) (2005) *World Atlas of Great Apes and their Conservation*. Berkeley: University of California Press.

Ohno S (1970) *Evolution by gene duplication*. New York: Springer.