Identification by full-coverage array CGH of human DNA copy number increases relative to chimpanzee and gorilla

Gary M. Wilson,¹ Stephane Flibotte,¹ Perseus I. Missirlis,¹ Marco A. Marra,¹ Steven Jones,¹ Kevin Thornton,² Andrew G. Clark,² and Robert A. Holt^{1,3}

¹Canada's Michael Smith Genome Sciences Centre, Vancouver, BC, Canada V5Z 4S6; ²Cornell University, Ithaca, New York 14853, USA

Duplication of chromosomal segments and associated genes is thought to be a primary mechanism for generating evolutionary novelty. By comparative genome hybridization using a full-coverage (tiling) human BAC array with 79-kb resolution, we have identified 63 chromosomal segments, ranging in size from 0.65 to 1.3 Mb, that have inferred copy number increases in human relative to chimpanzee. These segments span 192 Ensembl genes, including 82 gene duplicates (41 reciprocal best BLAST matches). Synonymous and nonsynonymous substitution rates across these pairs provide evidence for general conservation of the amino acid sequence, consistent with the maintenance of function of both copies, and one case of putative positive selection for an uncharacterized gene. Surprisingly, the core histone genes *H2A*, *H2B*, *H3*, and *H4* have been duplicated in the human lineage since our split with chimpanzee. The observation of increased copy number of a human cluster of core histone genes suggests that altered dosage, even of highly constrained genes, may be an important evolutionary mechanism.

[Supplemental material is available online at www.genome.org.]

Gene duplication has long been considered a primary mechanism of adaptive evolution (Ohno 1970). In theory, newly duplicated genes are redundant, and relaxed functional constraints allow acquisition of sequence changes in support of new functions and expression patterns. The importance of gene duplication in human evolution is supported by numerous studies that have documented DNA copy number differences between human and nonhuman primates. These studies have used diverse molecular approaches based on karyotyping (Yunis et al. 1980), physical clone maps (Fujiyama et al. 2002), partial coverage genomic arrays (Locke et al. 2003), cDNA arrays (Fortna et al. 2004), end-sequence profiling (Newman et al. 2005), draft genome sequence (Chimpanzee Sequencing and Analysis Consortium 2005), and high-quality sequence from chimp Chromosome 21 (orthologous to human Chromosome 22) (Watanabe et al. 2004). Careful alignment of accurately finished genome sequences from human and other primates promises to reveal DNA copy number differences at the highest possible resolution. Finished sequence for human is available, but the whole-genome shotgun (WGS) assembly for chimpanzee remains in draft form. The WGS approach is of proven value for rapidly and economically generating a full genome sequence, but it is clear that current approaches to assembling WGS data sets can underrepresent recently duplicated genome segments (She et al. 2004). Thus, to achieve a complete genome survey of DNA copy number in human versus chimpanzee, we have performed comparative genome hybridization (CGH) using the first full-coverage bacterial artificial chromosome (BAC) array of the human genome, which consists of 32,855 overlapping clones providing ~79 kb average resolution.

³Corresponding author.

E-mail rholt@bcgsc.ca; fax (604) 877-6085. Article published online ahead of print. Article and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.4456006. This is a direct approach and yields the greatest coverage at the highest level of resolution thus far achieved for comparison of these three species.

Results

Using full-coverage BAC array CGH, we executed a three-phased approach to identify segments of human genomic DNA that have likely been acquired since divergence from the common ancestor we share with chimpanzee. First, two samples of human genomic DNA (gDNA), one pooled from seven unrelated males and the other pooled from four unrelated females, were cohybridized to identify and exclude nodes on the array that gave aberrant ratios in a human-only comparison. Pooled DNAs were used in order to minimize the number of hybridization experiments, and to favor the detection of fixed rather than polymorphic copy number differences. Of the 31,842 mapped autosomal clones on the array, 212 showed aberrant ratios (>1.5 H-spread; see Methods) in the human-human comparison, and were excluded from further analysis. Next, we hybridized the human test DNA sample pooled from seven human males to a reference DNA sample comprised of DNA pooled from three unrelated male chimpanzees (Coriell Institute, Repository numbers NAO3448, NAO3450, NAO3452) (Fig. 1). These hybridizations were repeated under dye reversal, and a total of 1319 clones (855 increases, 464 decreases) were identified that consistently showed ratios that exceeded threshold in both dye orientations. As an added measure of stringency, we retained clones only if (1) they were confirmed by an equivalent copy number aberration in at least one additional overlapping clone, or (2) their location in the human reference genome sequence (NCBI_34) is supported by both their restriction digest pattern and BAC end sequence placement (Krzywinski et al. 2004). Under these criteria, a total of 585 clones gave elevated ratios in human relative to chimpanzee.



Figure 1. Human chimpanzee DNA copy number ratio determined by full-coverage BAC array CGH.

We used gorilla as an outgroup to determine the most likely ancestral copy number state. By parsimony, human chromosomal segments showing an increased copy number ratio relative to both chimpanzee and gorilla most likely represent insertions specific to the human lineage. This is true regardless of whether the human genomic region containing the given segment is more similar to chimpanzee or gorilla. Note, however, that there are further caveats to the parsimony approach that must be considered. While the widely accepted species tree of hominoids places human and chimpanzee as a clade, with gorilla as an outgroup, there are regions of the genome that are incongruent with the species tree. For regions of the genome consistent with a human-gorilla clade, the assignment of a copy number increase to human is unaffected, that is, parsimony still favors a single event in the human branch over two independent events in the chimp and gorilla branches. However, a study by Chen and Li (2001) reported that 12 of 53 surveyed loci were more consistent with a chimp-gorilla clade than a chimp-human clade. For these regions only, parsimony is blind to whether there is a copy number increase in the human branch, versus a decrease in the branch leading from the human/chimp/gorilla node to the chimp/gorilla node. Thus, regarding the ancestral copy number state, our hybridization results will be ambiguous for ~20% of the genome. In some regards, orangutan may be a more suitable outgroup, since the ratio of unresolved ancestral polymorphism to divergence is much lower because of the longer divergence time. However, a potential drawback in using orangutan as an outgroup in these experiments is that the arrays are spotted with human genomic clones, and hybridization becomes less reliable when more distant species are evaluated.

Thus, we proceeded with hybridization of the pooled human male test DNA sample to reference DNA from a single female gorilla (Coriell Institute, Repository number NGO5251). We decided to use human male test DNA rather than female test DNA for consistency with previous experiments. Because we had fewer chimpanzee and gorilla samples than human samples, there is some possibility that sites that are polymorphic in chimpanzee and gorilla have impacted our analysis. The fact that we

restrict analysis to genomic segments where chimpanzee and gorilla copy number agree, relative to human, minimizes this impact. Of the 585 clones that had an elevated ratio in human relative to chimpanzee, 235 also gave elevated ratios relative to gorilla and therefore likely represent human-specific copy number increases. Presumably, the subset of clones that did not show elevated ratios relative to gorilla represent copy number decreases in chimpanzee relative to the ancestral state. Again, as an added measure of stringency, clones have been retained in the set of 235 only if they are confirmed by an equivalent copy number aberration in at least one additional overlapping clone, or their location in the human reference genome sequence (NCBI_34) is supported by both their restriction digest pattern and BAC-end sequence placement. These 231 clones collapse into 55 contiguous chromosomal segments (43 with multiple clones, and 12 singletons) with minimum, maximum, and average segment lengths of 65,252 bp, 1,133,633 bp, and 308,959 bp, respectively, and a cumulative genome footprint of 16,992,728 bp.

Separately, we evaluated ratios of clones located on the X and Y chromosomes. We identified a total of eight X chromosome (ChrX) and 28 Y chromosome (ChrY) clones that met the criteria of concordant dye-flip ratios and an equivalent copy number difference in at least one overlapping clone. Sex chromosome ratios from the female gorilla sample are not directly comparable to those from the male chimpanzee and human reference samples, thus for sex chromosome differences we are unable to infer human increase rather than chimp decrease. However, evaluation of duplicate segments within the human reference genome sequence (below) supports the notion that these are copy number increases in the human lineage. These eight ChrX clones and 28 ChrY clones collapse into two ChrX contigs and six ChrY contigs covering 415,787 bp and 1,190,263 bp, respectively, bringing the cumulative genomic footprint of all segments (autosomal plus sex chromosome) to 18,598,778 bp (Table 1). These segments are the basis of further analysis. While loss of genetic material on the human lineage is of considerable interest, here we consider only observed copy number increases. This is because copy number increases, as opposed to losses, can

Start End Size (bb) Band 1 16,264,647 16,504,347 239,700 1936,13 1 16,374,958 103,782,587 267,629 1921.1 1 141,999,150 141,904,149 204,999 1q21.1 1 142,670,168 1446,179,403 309,247 1q21.1 2 87,7850,005 87,967,456 87,451 2011.2 2 87,880,005 87,967,456 87,451 2011.2 2 89,920,779 89,226,310 235,551 2011.2 2 89,290,779 89,226,330 235,551 2011.2 2 110,374,075 110,628,474 254,399 2013.3 2 112,279,849 112,493,396 200,547 2q13 2 112,279,849 112,493,396 200,547 2q13 3 20,913.3 4p16.3 4q13.2 117.2 4 13,423 177,154 163,616,362 5q12.2 5 109,955,192 70,448,	-				
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	Chromosome	Start	End	Size (kb)	Band
1 103,514,958 103,782,587 267,629 102,11 1 142,697,150 141,904,149 204,999 142,11 1 142,697,150 144,704,000 209,435 142,11 1 145,670,162 147,704,000 209,435 142,12 2 87,880,005 87,967,456 87,751 201,12 2 87,880,005 87,967,456 87,751 201,12 2 89,9657,739 90,009,157 153,418 201,12 2 91,079,860 91,274,418 198,635 201,12 2 91,079,860 91,274,418 198,635 201,12 2 91,079,860 91,274,418 198,635 201,12 2 91,079,860 91,274,418 198,6105 401,32 3 91,42,214 70,618,319 186,105 401,32 4 70,432,214 70,618,319 186,105 401,32 5 60,07,757 67,33,302 123,047 63,32,12 5	1	16,264,647	16,504,347	239,700	1p36.13
1141,699,150141,994,149204,9991621,11144,697,568142,632,783134,915142,1-1-12,11146,607,668144,6179,409509,247142,1-1-12,1287,255,53188,008,04175,2,5102p11,2287,255,53188,008,04175,2,5102p11,2287,285,53188,008,04175,2,5102p11,2289,200,77989,526,330235,5512p11,2289,200,77989,526,330235,5512p11,2210,374,075116,628,474254,3992q132112,229,249112,493,396200,5472q132112,229,249112,493,396200,5472q13474,32,27470,77199163,7314p13,3560,907,75769,733,382727,6255q13,2560,907,75769,733,382727,6255q13,2560,907,75769,733,840200,3462q11,1-2774,931,07172,131,909200,8487q11,23774,931,94961,035,460200,3467q11,23774,503,81474,734,450230,6467q11,23774,503,81474,734,450230,6467q11,23774,931,94874,734,450230,6467q11,23774,944,0440,749,379165,9739p12940,174,40440,749,379165,9739p12941,113,34441,55,670245,3369p12<	1	103 514 958	103 782 587	267 629	1n21 1
1 142,697,868 142,832,783 134,915 142,11-1q12 1 146,800,368 147,090,803 290,435 1q21,1-1q12 2 87,880,005 87,967,456 87,451 2p11.2 2 87,880,005 87,967,456 87,451 2p11.2 2 87,880,005 87,967,456 87,451 2p11.2 2 89,290,779 89,326,337 135,518 2p11.2 2 89,290,779 89,326,337 135,518 2p11.2 2 91,079,860 91,272,418 198,588 2p11.2 2 110,374,075 110,628,474 254,399 2q13 2 112,292,849 112,493,396 200,547 2q13 4 13,423 177,154 165,711 4p16.3 4 13,423 20,33,080 138,082 5p14.1 5 26,149,382 20,33,080 136,082 5p14.1 5 26,149,382 270,33,080 136,082 5p14.1 5 26,1	1	141 699 150	141 904 149	204 999	1021.1
1 145,670,162 146,179,409 509,247 1q21,1-lq21,2 1 146,800,368 147,0790,803 200,4453 1q1,2 2 87,255,531 88,006,041 752,510 2p11,2 2 87,855,531 88,006,041 752,510 2p11,2 2 89,146,792 89,265,612 138,820 2p11,2 2 89,285,739 90,009,157 133,418 2p11,2 2 89,285,739 90,009,157 133,418 2p11,2 2 112,222,849 112,493,396 200,547 2q11-1-q21,2 4 13,423 177,154 165,105 4q13,2 5 60,007,757 69,753,382 72,7625 5q13,2 5 60,955,192 70,346,029 329,837 5q13,2 6 100,704,322 170,494,708 170,464 6q2,2 7 64,392,229 64,33,018 140,799 71,31,2 7 74,391,701 72,131,909 200,838 7q11,21 7	1	142 697 868	142 832 783	134 915	1021.1
1 14,6,800,386 17,095,803 290,445 140,12 2 82,785,531 82,000,01 75,2510 2p11.2 2 87,880,005 87,967,456 87,451 2p11.2 2 89,200,779 89,526,330 235,551 2p11.2 2 89,200,779 89,526,330 235,551 2p11.2 2 89,200,779 89,526,330 235,551 2p11.2 2 91,079,860 91,278,418 198,538 2p11.2 2 110,252,840 112,2490,765 230,103 2q21.1-2q2.2 4 13,423 177,154 163,731 4p16.3 5 66,007,757 69,735,382 727,625 5q13.2 5 69,951,92 70,348,029 392,887 5q13.2 6 9,951,92 70,348,029 392,877 5q13.2 7 60,853,494 61,053,840 200,346 7q1.1 7 73,953,814 73,73,438 140,789 7q1.1 7 73,803	1	145 670 162	146 170 400	500 247	1921.1
1 140,000,001 200,430 120,12 2 367,000,001 87,000,001 200,430 120,12 2 880,005 87,000,011 138,820 2011.2 2 89,166,792 89,255,612 138,820 2011.2 2 89,855,739 90,009,157 153,418 2011.2 2 91,079,860 91,278,418 198,558 2011.2 2 110,374,075 110,628,474 254,399 2013 2 112,229,249 112,493,396 200,647 2013 2 112,229,249 112,493,396 200,647 2013 2 112,292,849 112,493,396 200,647 2013 2 112,292,849 112,493,396 200,647 2013 2 112,292,849 112,493,396 200,647 2013 2 12,292,847 20,313,800 188,698 5013,2 5 12,286,922 170,346,029 392,837 5013,2 6 170,704,322 1	1	143,070,102	140,179,409	200,425	1,21.2
2 87,253,511 88,008,041 72,210 211.2 2 87,800,053 87,967,455 87,411 2011.2 2 89,900,779 89,526,513 123,511 2011.2 2 89,955,739 90,000,157 153,418 2011.2 2 91,079,860 91,278,418 198,558 2011.2 2 110,374,075 110,628,474 254,399 2013 2 112,257,9662 113,2809,765 230,103 2021.1-221.2 4 70,432,214 70,618,319 186,103 4015.2 5 26,149,382 26,133,080 183,669 59,14.1 5 26,149,382 26,333,080 183,669 59,14.1 6 170,704,322 170,894,763 190,441 60,27 7 64,353,014 64,353,014 203,466 70,11.21 7 73,503,814 73,73,44,80 206,636 70,11.21 7 73,503,814 73,73,44,80 206,636 70,11.23 7 </td <td>1</td> <td>140,800,368</td> <td>147,090,803</td> <td>290,435</td> <td>1921.2</td>	1	140,800,368	147,090,803	290,435	1921.2
2 87,880,005 87,967,456 87,431 2p11.2 2 89,146,792 89,2285,612 138,820 2p11.2 2 89,106,779 89,526,337 235,511 2p11.2 2 110,374,605 110,628,66 91,278,18 198,158 2p11.2 2 112,295,849 112,499,396 200,547 2p13 2 112,295,849 112,499,396 200,547 2p13 4 134,232 177,154 163,731 4p16.3 4 70,432,214 70,618,319 186,105 4p13.2 5 69,007,757 69,735,382 727,625 5p13.2 5 69,905,752 70,448,029 392,837 5p13.2 5 69,905,752 70,348,029 392,843 5p13.2 7 70,432,217,0494,763 190,441 6q22,2 7q14 7 73,391,085 74,0499,099 178,014 7q11.23 7 73,391,085 74,0499,099 175,014 7q11.23	2	87,255,531	88,008,041	/52,510	2p11.2
2 89,146,792 89,286,312 138,820 2p11.2 2 89,855,739 90,009,157 153,418 2p11.2 2 91,079,860 91,278,418 198,538 2p11.2 2 110,374,075 1116,483,474 254,399 2q13 2 112,272,849 112,493,395 200,547 2q13 2 112,272,849 112,493,392 200,547 2q13 2 112,377 2q13 2q11.2 2q13 5 69,007,757 69,735,382 272,625 5q13.2 5 112,286,922 113,064,958 178,036 5q22.2 7 60,035,494 61,035,840 200,346 7q11.1 7 71,931,071 72,131,909 200,838 7q11.23 7 73,308,814	2	87,880,005	87,967,456	87,451	2p11.2
2 89,290,779 89,526,330 235,551 2p11.2 2 99,635,739 99,009,157 153,418 2p11.2 2 10,079,860 91,278,418 198,558 2p11.2 2 110,274,075 110,628,474 234,399 2q13 2 112,292,849 111,493,365 200,547 2q13 4 74,422 177,65 200,547 2q13 5 26,6149,382 26,333,080 188,698 5p14.1 5 26,97,757 69,735,382 272,625 5q13.2 5 69,907,757 69,735,842 272,625 5q13.2 6 170,704,322 170,894,763 190,441 6q22 7 64,352,429 61,033,840 200,348 7q11.21 7 73,918,81 72,734,699 200,638 7q11.23 7 73,918,81 74,734,699 200,638 7q11.23 7 73,918,81 74,734,699 70,654 7q14 7 73,918,81	2	89,146,792	89,285,612	138,820	2p11.2
2 89,655,739 90,009,157 153,418 2p11.2 2 91,079,860 91,278,418 198,558 2p11.2 2 112,379,662 110,678,674 254,359 2p13 2 112,292,849 112,493,366 200,547 2p13 2 112,292,849 112,493,366 200,547 2p13 4 71,422,214 70,618,319 186,103 4p16.3 4 71,422,214 70,618,319 186,103 4p16.3 5 20,007,57 20,735 5p13.2 5p13.2 5 112,286,922 170,840,298 178,036 5q12.2 7 60,835,494 61,035,840 200,346 7q11.1 7 71,931,071 72,131,909 208,636 7q11.23 7 73,303,814 73,724,626 65,522 7q34 7 73,303,814 73,724,626 65,522 7q34 7 73,303,814 73,724,626 65,525 7q34 7 142,656,837 <td>2</td> <td>89,290,779</td> <td>89,526,330</td> <td>235,551</td> <td>2p11.2</td>	2	89,290,779	89,526,330	235,551	2p11.2
2 91,079,860 91,278,418 198,558 2p11.2 2 112,292,849 112,493,396 200,547 2q13 2 132,579,662 132,899,765 230,103 2q21.1-2q21.2 4 70,432,214 70,618,319 186,105 4q13.2 5 66,907,757 69,733,382 727,625 5q13.2 5 69,907,757 69,735,382 727,625 5q13.2 6 170,704,322 170,894,783 178,036 5q22.2 7 64,939,107 74,319,909 200,888 7q11.23 7 73,503,814 74,734,450 230,636 7q11.23 7 73,503,814 74,734,450 230,636 7q11.23 7 73,503,814 74,734,450 230,636 7q11.23 7 74,940,853 74,089,099 175,014 7q13.3 9 38,905,428 39,384,898 479,470 913.1 9 40,744,04 40,340,379 165,975 912 9	2	89,855,739	90,009,157	153,418	2p11.2
2 110.374/075 110.628/474 264.399 213 2 112,292,849 112,493,396 200,547 2013 2 132,579,662 132,809,765 230,103 2221.1-221.2 4 13,423 177,154 163,731 4p16.3 4 70,432,214 70,618,319 186,105 4p13.2 5 66,007,757 69,735,382 727,625 5q13.2 5 69,955,192 70,348,029 392,837 5q13.2 6 170,704,322 170,894,763 190,441 6q27 7 60,835,494 61,035,840 200,346 7q11.21 7 71,931,071 72,131,909 200,838 7q11.23 7 73,503,814 73,74,903 240,635 7q14.23 7 73,903,814 73,74,903 240,635 7q14.23 7 73,903,814 74,72,803 442,723,60 65,125 7q34 7 143,184,261 144,72,73,603 441,27 7q34	2	91.079.860	91.278.418	198.558	2p11.2
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	2	110,374,075	110,628,474	254 399	2g13
132.579.662 132.800,765 2010 $2221.1-221.2$ 4 13423 177.154 163.731 4016.5 4 70.432.214 70.618.319 186.105 4013.2 5 26.149.382 22.6333.080 183.698 5p14.1 5 66.077.57 607.35.382 727.623 5013.2 6 170.704.322 170.894,763 190.441 6627 7 60.835.494 61.035.840 200.346 7011.21 7 73.503.814 73.73.940 200.638 7011.21 7 73.503.814 73.73.491 200.638 7011.23 7 73.503.814 73.73.903 212.430 7043.2 7 73.503.814 73.73.903 212.430 7014.2 7 73.503.814 73.73.903 212.430 7043.2 7 74.26.68.374 142.72.56.67 531.204 7034.2 7 74.26.58.374 142.72.36.66 531.204 7034.2 7 74.26.58.374 <td>2</td> <td>112 292 849</td> <td>112 493 396</td> <td>200 547</td> <td>2013</td>	2	112 292 849	112 493 396	200 547	2013
2 12,2,323 12,27,154 12,373 12,41,1-3421 4 70,452,214 70,161,319 186,105 44,13.2 5 26,149,382 26,333,080 186,105 44,13.2 5 69,907,757 667,353,382 727,625 59,13.2 5 112,286,622 113,064,958 176,036 59,22.2 6 170,704,322 170,894,763 190,441 69,27 7 64,392,229 64,533,018 140,789 70,11.1 7 73,914,085 74,089,099 175,014 70,11.23 7 73,914,085 74,089,099 175,014 70,11.23 7 142,658,374 142,723,626 65,252 70,34 8 47,000,811 47,258,017 257,206 80,11.1 9 38,905,428 39,348,488 479,470 913.1 9 40,174,404 40,340,379 165,975 9112 9 41,431,201 41,355,828 114,627 911.1.2 9	2	132 570 662	132,800,765	230,103	$2q_{13}$
4 12,423 17/154 162,131 4016.3 4 70,432,214 70,618,319 186,105 4013.2 5 26,149,382 26,333,080 188,608 5014.1 5 69,007,757 69,753,882 727,625 5013.2 5 112,886,922 113,064,928 176,036 5022.2 6 170,704,322 170,894,763 190,441 6q27.1 7 66,333,049 200,346 701.11 6 7 71,311,071 72,131,980 200,858 701.23 7 142,659,324 142,723,625 65,220 70,345 7 142,659,324 142,723,625 65,220 70,345 7 142,659,324 142,723,625 65,220 70,345 8 47,006,14 43,594,067 232,207 70,345 9 39,756,773 39,976,673 39,206 70,912 9 40,475,504 40,503,04 41,453,05 912 9 41,457,781	2	132,379,002	132,009,703	230,103	2421.1-2421.2
4 70,432,214 70,018,319 185,098 40,152 5 26,143,382 26,333,080 183,698 51,41 5 69,055,192 70,348,029 322,837 53,132 5 112,886,922 113,064,958 170,064,322 22,22 6 170,704,322 170,894,763 190,441 6q,27 7 66,383,494 61,035,840 200,346 70,11.21 7 73,503,814 73,734,450 230,636 70,11.23 7 73,503,814 73,734,626 65,252 70,344 7 143,134,261 143,465,465 331,204 70,351 9 38,065,423 39,384,898 479,470 913.1 9 39,765,473 39,978,303 212,830 912 9 40,175,004 40,350,37 56,975 912 9 40,475,504 40,759,034 263,536 912 9 41,111,334 41,356,670 245,336 912 9 40,475,	4	13,423	1/7,154	103,/31	4010.3
5 26,149,382 26,333,080 183,698 5p14.1 5 69,007,757 69,735,382 727,625 5q13.2 5 112,886,922 173,046,029 392,837 5q13.2 6 170,704,322 170,894,763 190,441 6q27 7 66,385,494 61,035,840 200,346 7q11.1 7 64,392,229 64,533,018 140,789 7q11.21 7 73,914,085 74,089,099 175,014 7q11.23 7 73,914,085 74,089,099 175,014 7q11.23 7 142,658,374 142,723,626 65,252 7q34 8 47,000,811 47,258,017 25,206 8q11.1 9 38,905,428 39348,488 479,470 9p13.1 9 40,174,404 40,340,379 165,975 9p12 9 40,174,404 40,340,379 165,975 9p12 9 41,313,201 41,354,828 114,627 9p11.2 9 43,3	4	/0,432,214	70,618,319	186,105	4q13.2
5 66,007,757 66,735,382 727,625 Sq13.2 5 169,955,192 70,348,029 922,837 Sq13.2 5 112,886,922 113,064,958 178,036 Sq22.2 7 60,835,494 61,035,840 200,346 7q11.1 7 7,931,071 72,131,909 200,638 7q11.23 7 73,503,814 77,34,450 230,636 7q11.23 7 73,503,814 77,34,450 230,636 7q11.23 7 73,503,814 143,465,465 331,204 7q35 8 47,000,811 42,258,2017 257,206 8g11.1 9 39,765,473 39,978,303 212,830 9p12 9 40,174,404 40,340,379 165,975 9p12 9 40,174,55,04 40,350,077 500,914 9p13 9 41,431,201 41,356,670 245,336 9p12 9 41,431,201 41,356,670 245,336 9p12 9 43,787,	5	26,149,382	26,333,080	183,698	5p14.1
5 69,955,192 70,348,029 392,837 513.2 6 170,704,322 173,064,958 178,036 522.2 7 60,855,494 61,035,840 200,346 70,11.1 7 64,392,229 64,533,018 140,789 70,11.21 7 71,931,071 72,131,909 200,838 70,11.23 7 73,503,814 73,734,450 230,636 70,11.23 7 73,914,005 74,089,099 175,014 70,11.23 7 142,658,374 142,723,626 65,252 70,34 8 47,000,811 47,258,017 257,206 80,11.1 9 38,905,428 93,348,4898 479,470 9p,12 9 40,174,404 40,340,379 165,975 9p,12 9 40,474,404 40,340,379 165,975 9p,12 9 41,451,201 41,545,828 114,627 9p,12 9 41,451,201 43,56,770 260,914 90,12 9 <	5	69,007,757	69,735,382	727,625	5q13.2
5 112,886,922 113,064,958 178,036 5q22.2 6 170,704,322 170,894,763 190,441 6q27 7 66,835,494 61,035,840 200,346 7q11.1 7 7,931,071 72,131,909 200,838 7q11.23 7 73,503,814 72,734,450 200,636 7q11.23 7 142,658,374 142,723,626 65,252 7q34 8 47,000,811 47,258,017 257,206 8q11.1 9 38,905,428 39,344,808 479,470 9p13.1 9 40,174,404 40,340,379 165,575 9p12 9 40,474,404 40,340,379 165,575 9p12 9 41,413,201 41,545,828 114,627 9p12.2 9 41,413,201 41,545,828 144,627 9p11.2 9 41,545,828 144,627 9p11.2 9 9 45,274,163 46,315,077 560,914 9q13 9 45,216,018	5	69,955,192	70,348,029	392,837	5q13.2
6 170,704,322 170,894,763 190,441 66,27 7 66,335,494 61,035,840 200,346 70,11.1 7 71,931,071 72,131,909 200,838 70,11.23 7 73,503,814 73,734,450 200,636 70,11.23 7 73,503,814 73,734,450 200,636 70,11.23 7 143,658,374 142,756,26 65,252 70,34 7 143,134,261 143,465,465 331,204 70,35 8 47,000,811 47,258,6017 257,206 80,11.1 9 38,905,428 39,384,898 479,470 9p13.1 9 40,174,404 40,340,379 165,975 9p12 9 40,174,404 40,340,379 165,975 9p12 9 41,11,13,34 41,356,670 245,336 9p12 9 41,451,201 41,55,828 114,627 9p11.2 9 41,451,201 41,452,794 40,495,463 9p12.2 9	5	112.886.922	113,064,958	178,036	5g22.2
7 60 (335, 494 61 (35) (340 200 (346 7 7 64, 332, 229 64, 533, 018 140, 789 70, 11, 21 7 73, 503, 814 73, 73, 144, 90 200, 838 70, 11, 23 7 73, 503, 814 73, 73, 144, 90 200, 838 70, 11, 23 7 73, 514, 085 74, 099, 099 175, 014 70, 11, 23 7 142, 658, 374 142, 723, 626 65, 252 70, 344 8 47, 000, 811 47, 258, 017 257, 206 80, 11, 1 9 38, 905, 428 39, 394, 898 479, 470 913, 1 9 38, 905, 428 39, 394, 898 479, 470 913, 1 9 39, 765, 473 39, 978, 303 212, 830 912 9 40, 174, 404 40, 379 165, 675 9p12 9 41, 113, 34 41, 356, 670 245, 336 9p12 9 41, 312, 201 41, 545, 828 114, 627 9p11.2 9 43, 787, 741 44, 127, 010 339, 269 <td>6</td> <td>170,704,322</td> <td>170,894,763</td> <td>190,441</td> <td>6q27</td>	6	170,704,322	170,894,763	190,441	6q27
Comparison Comparison Comparison Comparison Comparison 7 71,931,071 72,131,909 200,838 7q11.21 7 73,303,814 72,731,909 200,838 7q11.23 7 73,303,814 72,731,4450 230,636 7q11.23 7 73,303,814 72,731,2626 65,252 7q34 7 143,134,261 143,465,465 331,204 7q35 8 47,000,811 47,258,6017 257,206 8q11.1 9 38,905,428 39,384,898 479,470 9p13.1 9 39,765,473 39,978,303 212,830 9p12 9 40,174,404 40,340,379 165,975 9p12 9 41,431,201 41,545,828 114,627 9p12 9 41,431,201 41,545,828 114,627 9p11.2 9 43,787,741 44,127,010 339,269 9p11.2 9 45,216,018 65,523,621 30,7603 9q13 9	7	60 835 494	61 035 840	200 346	7g11 1
7 04,32,223 04,335,016 140,765 24,11,21 7 77,931,071 72,313,909 200,636 74,11,23 7 73,914,085 74,089,099 175,014 74,11,23 7 142,658,374 142,723,626 65,252 74,34 8 47,000,811 47,258,017 257,206 84,11.1 9 38,905,428 39,348,898 479,470 9p13.1 9 39,765,473 39,978,303 212,830 9p12 9 40,495,504 40,759,034 263,530 9p12 9 40,495,504 40,759,034 263,530 9p12 9 41,431,201 41,545,828 114,627 9p11.2 9 41,431,201 41,545,828 114,627 9p11.2 9 43,577,81 42,021,347 463,466 9p11.2 9 45,574,163 65,1814,846 197,193 9q13 9 65,216,018 65,523,621 307,663 194,192,22 9	7	64 202 220	64 522 019	140 790	7q11.1 7q11.21
7 71,931,071 72,131,999 200,838 74,11,23 7 73,503,814 73,734,450 230,636 74,11,23 7 73,914,085 74,089,099 175,014 74,11,23 7 142,658,374 142,723,626 65,522 7q34 7 143,134,261 143,465,465 331,204 7q35 8 47,000,811 47,258,017 257,706 8g11.1 9 39,905,428 39,384,898 479,470 9p12 9 40,174,404 40,340,379 165,975 9p12 9 40,495,504 40,759,034 263,530 9p12 9 41,431,201 41,545,828 114,627 9p11.2 9 41,575,781 42,021,347 463,466 9p11.2 9 43,787,741 44,127,010 339,269 9p11.2 9 45,216,018 65,523,621 307,603 9q13 9 65,617,653 65,814,846 197,193 9q13-9q21.11 10 <t< td=""><td>7</td><td>71 021 071</td><td>72 121 000</td><td>140,707</td><td>7411.21</td></t<>	7	71 021 071	72 121 000	140,707	7411.21
7 73,914,085 74,74,450 240,656 74,11,23 7 142,658,374 142,723,626 65,252 74,34 7 143,134,261 143,465,465 331,204 735 8 47,000,811 47,258,017 257,206 80,11.1 9 38,905,428 39,384,898 479,470 9p,13.1 9 39,765,473 39,978,303 212,830 9p,12 9 40,174,404 40,340,379 165,975 9p,12 9 40,495,504 40,759,034 263,530 9p,12 9 41,431,201 41,545,828 114,627 9p,11.2 9 41,431,201 41,545,828 114,627 9p,11.2 9 43,577,881 42,021,347 463,466 9p,11.2 9 43,574,1 44,127,010 339,269 9p,11.2 9 65,617,653 65,814,846 197,193 90,13-90,21,11 10 46,138,077 560,914 90,13-92,21,11 10 46,315,077<	/	71,931,071	72,131,909	200,838	7011.23
7 73,914,085 74,089,099 175,014 7q11,23 7 142,658,374 142,723,626 65,252 7q34 7 143,134,261 143,465,465 331,204 7q35 8 47,000,0811 47,258,017 257,206 8q11.1 9 38,905,428 39,384,898 479,470 9p12 9 40,174,404 40,340,379 165,975 9p12 9 40,495,504 40,759,034 263,530 9p12 9 41,431,201 41,556,828 114,627 9p11.2 9 41,431,201 41,554,828 114,627 9p11.2 9 43,787,741 44,127,010 39,269 9p13.3 9 65,216,018 65,523,621 307,603 9q13 9 65,617,653 65,814,866 197,193 9q13-9q21.11 10 46,138,281 46,380,466 242,185 10q11.22 10 57,969,110 58,079,974 110,864 10q21.1 15 18,898,763 19,129,849 231,086 15q11.2 15 <	/	/3,503,814	/3,/34,450	230,636	/q11.23
7 142,658,374 142,723,626 65,252 7q34 7 143,134,261 143,465,465 31,204 7q35 8 47,000,811 47,258,017 257,206 8q11.1 9 38,905,428 39,384,898 479,470 9p13.1 9 39,765,473 39,978,303 212,830 9p12 9 40,174,404 40,340,379 165,975 9p12 9 40,495,504 40,759,034 263,530 9p12 9 41,431,201 41,545,828 114,627 9p11.2 9 41,431,201 41,545,828 114,627 9p11.2 9 43,787,741 44,17,010 339,269 9p13.2 9 65,574,163 64,135,077 560,914 9q13 9 65,216,018 65,523,621 307,603 9q13-921.11 10 46,138,281 46,380,466 242,185 10q11.22 11 14,130 18,405,573 335,572 14q11.2 15 18,887,63 19,128,449 231,086 15q11.2 15 20,081	7	73,914,085	74,089,099	175,014	7q11.23
7 143,134,261 143,465,465 331,204 7q35 8 47,000,811 47,258,017 257,206 8q11.1 9 38,905,428 39,384,898 479,470 9p13.1 9 39,765,473 39,978,303 212,830 9p12 9 40,174,404 40,340,379 165,975 9p12 9 40,495,504 40,379 245,336 9p12 9 41,111,334 41,356,670 245,336 9p12 9 41,431,201 41,454,828 114,627 9p1.2 9 43,577,741 44,127,010 39,269 9p1.2 9 65,216,018 65,523,621 307,603 9q13 9 65,617,653 65,814,846 197,193 9q13-9q21.11 10 46,382,81 46,380,466 242,185 10q11.22 10 45,873 19,129,849 231,086 15q11.2 15 19,301,328 19,116,381 615,053 15q11.2 15 20,051,804 22,6752,132 696,328 15q11.2 15 20,04834	7	142,658,374	142,723,626	65,252	7q34
8 47,000,811 47,288,017 257,206 8q11.1 9 38,905,428 39,384,898 479,470 9p13.1 9 39,765,473 39,978,303 212,830 9p12 9 40,174,404 40,340,379 165,975 9p12 9 40,495,504 40,759,034 263,530 9p12 9 41,411,201 41,545,828 114,627 9p11.2 9 41,557,881 42,021,347 463,466 9p11.2 9 63,574,163 64,135,077 560,914 9q13 9 65,216,018 65,523,621 307,603 9q13 9 65,216,018 65,823,621 307,603 9q13 9 65,216,018 65,814,846 197,193 9q13-9q21.11 10 46,138,281 46,380,466 242,185 10q11.22 10 57,969,110 58,079,974 110,866 15q11.2 15 18,989,763 19,129,849 231,086 15q11.2 15	7	143,134,261	143,465,465	331,204	7q35
9 38,905,428 39,384,898 479,470 913,1 9 39,765,473 39,978,303 212,830 912 9 40,174,404 40,340,379 165,975 9p12 9 40,495,504 40,379,034 263,530 9p12 9 41,111,334 41,356,670 245,336 9p12 9 41,431,201 41,545,828 114,627 9p11.2 9 41,57,781 42,021,347 463,466 9p11.2 9 45,216,018 65,523,621 307,603 9q13 9 65,217,653 65,814,846 197,193 9q13-9q21.11 10 46,138,281 46,380,466 242,185 10q11.22 110 46,138,281 46,380,466 242,185 10q11.22 114 18,070,001 18,405,573 35,572 14q11.2 115 19,301,328 19,916,381 615,053 15q11.2 115 20,613,690 21,041,101 227,411 15q11.2 115	8	47,000,811	47,258,017	257,206	8q11.1
9 39,765,473 39,978,303 212,830 9p12 9 40,174,404 40,340,379 165,975 9p12 9 40,495,504 40,759,034 263,530 9p12 9 41,411,1334 41,356,670 245,336 9p12 9 41,413,201 41,356,670 245,336 9p12 9 41,431,201 41,356,670 245,336 9p12 9 41,557,881 42,021,347 463,466 9p11.2 9 63,574,163 64,135,077 560,914 9q13 9 65,216,018 65,523,621 307,603 9q13 9 65,216,018 65,814,846 197,193 9q13-9q21.11 10 46,138,281 46,380,466 242,185 10q11.22 10 57,969,110 58,079,974 110,864 10q21.1 14 18,007,001 18,405,573 335,572 14q11.2 15 19,304 26,752,132 66,6328 15q11.2 15 20,81	9	38,905,428	39.384.898	479,470	9p13.1
9 40,174,404 40,340,379 165,075 9p12 9 40,495,504 40,759,034 263,530 9p12 9 41,111,334 41,356,670 245,336 9p12 9 41,411,201 41,545,828 114,627 9p11.2 9 41,557,881 42,021,347 463,466 9p11.2 9 43,787,741 44,127,010 339,269 9p11.2 9 63,574,163 64,135,077 560,914 9q13 9 65,617,653 65,814,846 197,193 9q13-9q21.11 10 46,138,281 46,380,466 242,185 10q11.22 10 57,969,110 58,079,974 110,864 10q21.1 14 18,070,001 18,405,573 335,572 14q11.2 15 18,898,763 19,129,849 231,086 15q11.2 15 20,813,690 21,041,101 227,441 15q11.2 15 20,055,804 26,752,132 696,328 15q11.2 16	9	39,765,473	39,978,303	212,830	9n12
9 40,495,504 40,759,034 263,530 9p12 9 41,431,201 41,548,828 114,627 9p11.2 9 41,431,201 41,548,828 114,627 9p11.2 9 41,557,881 42,021,347 463,466 9p11.2 9 43,787,741 44,127,010 339,269 9p11.2 9 63,574,163 64,135,077 560,914 9q13 9 65,216,018 65,523,621 307,603 9q13-9 9 65,617,653 65,814,846 197,193 9q13-9 10 46,138,281 46,380,466 242,185 10q11.22 10 57,969,110 58,079,974 110,864 10q21.1 14 18,070,001 18,405,573 335,572 14q11.2 15 19,301,328 19,916,381 615,053 15q11.2 15 20,813,690 21,041,101 227,411 15q11.2 15 20,813,690 21,041,101 243,780 15q11.2 15	9	40 174 404	40 340 379	165 975	9n12
9 41,111,334 41,35,670 245,336 912 9 41,431,201 41,545,828 114,627 9p11.2 9 41,557,881 42,021,347 463,466 9p11.2 9 43,787,741 44,127,010 339,269 9p11.2 9 63,574,163 64,135,077 560,914 9q13 9 65,216,018 65,523,621 307,603 9q13 9 65,617,653 65,814,846 197,193 9q13-9q21.11 10 46,138,281 46,380,466 242,185 10q11.22 10 56,617,653 65,814,846 197,193 9q13-9q21.11 10 56,617,653 65,814,846 197,193 9q13-9q21.11 10 58,079,974 110,864 10q21.1 14 18,070,001 18,405,573 335,572 14q11.2 15 19,301,328 19,916,381 615,053 15q11.2 15 20,813,690 21,041,101 227,411 15q11.2 15 20,655	0	10,17,1,101	10,510,575	263 530	9p12 9p12
9 41,11,354 41,350,670 243,330 9p12 9 41,431,201 41,545,828 114,627 9p11.2 9 43,787,741 44,127,010 339,269 9p11.2 9 63,574,163 64,135,077 560,914 9q13 9 65,216,018 65,523,621 307,603 9q13 9 65,617,653 65,814,846 197,193 9q13-9q21.11 10 46,138,281 46,380,466 242,185 10q11.22 10 57,969,110 58,079,974 110,864 10q21.1 14 18,070,001 18,405,573 335,572 14q11.2 15 19,301,328 19,916,381 615,053 15q11.2 15 20,013,690 21,041,101 227,411 15q11.2 15 20,04,834 22,248,614 243,780 15q11.2 16 <td>9</td> <td>41 111 224</td> <td>41 256 670</td> <td>203,330</td> <td>0p12</td>	9	41 111 224	41 256 670	203,330	0p12
9 41,431,201 41,545,828 114,627 9p11.2 9 41,557,881 42,021,347 463,466 9p11.2 9 43,787,741 44,127,010 339,269 9p11.2 9 63,574,163 64,135,077 560,914 9q13 9 65,216,018 65,523,621 307,603 9q13.9 9 65,617,653 65,814,846 197,193 9q13-9q21.11 10 46,138,281 46,380,466 242,185 10q21.1 10 46,138,281 46,304,466 242,185 10q21.1 14 18,070,001 18,405,573 335,572 14q11.2 15 18,898,763 19,129,849 231,086 15q11.2 15 19,301,328 19,916,381 615,053 15q11.2 15 20,813,690 21,041,101 227,411 15q12.1 15 20,04,834 22,248,614 243,780 15q11.2 15 20,043,755 21,948,510 196,151 16p1.2 16 21,002,359 21,198,510 196,151 16p1.2 1	9	41,111,554	41,530,070	243,330	9p12
9 41,557,881 42,021,347 465,466 9p11.2 9 43,5787,741 44,127,010 339,269 9p11.2 9 63,574,163 64,135,077 560,914 9q13 9 65,216,018 65,523,621 307,603 9q13 9 65,617,653 65,814,846 197,193 9q13-9q21.11 10 46,138,281 46,380,466 242,185 10q11.22 10 46,138,281 46,380,466 242,185 10q11.22 10 57,969,110 58,079,974 110,864 10q21.1 14 18,070,001 18,405,573 335,572 14q11.2 15 19,301,328 19,129,849 231,086 15q11.2 15 20,813,690 21,041,101 227,411 15q11.2 15 20,04,834 22,248,614 243,780 15q11.2 15 20,048,34 26,752,132 696,328 15q13.1 16 21,002,359 21,198,510 196,151 16p12.3 16 20,058,784 71,175,111 591,327 16q22.1-16q22.2	9	41,431,201	41,545,828	114,627	9p11.2
9 43,787,741 44,127,010 339,269 9p11.2 9 63,574,163 64,135,077 560,914 9q13 9 65,216,018 65,523,621 307,603 9q13 9 65,617,653 65,814,846 197,193 9q13-9q21.11 10 46,138,281 46,380,466 242,185 10q11.22 10 57,969,110 58,079,974 110,864 10q21.1 14 18,070,001 18,405,573 335,572 14q11.2 15 18,898,763 19,129,849 231,086 15q11.2 15 19,301,328 19,916,381 615,053 15q11.2 15 20,813,690 21,041,101 227,411 15q11.2 15 22,004,834 22,248,614 243,780 15q13.1 16 21,002,359 21,198,510 196,151 16q22.1-16q22.2 16 70,583,784 71,75,111 591,327 16q22.1-16q22.2 16 70,583,784 71,743,689 139,992 21q21.1 20 58,697,375 58,867,659 770,682 273,955 <td< td=""><td>9</td><td>41,557,881</td><td>42,021,347</td><td>463,466</td><td>9p11.2</td></td<>	9	41,557,881	42,021,347	463,466	9p11.2
9 63,574,163 64,135,077 560,914 9q13 9 65,216,018 65,523,621 307,603 9q13 9 65,617,653 65,814,846 197,193 9q13-9q21.11 10 46,138,281 46,380,466 242,185 10q11.22 10 57,969,110 58,079,974 110,864 10q21.1 14 18,070,001 18,405,573 335,572 14q11.2 15 18,898,763 19,129,849 231,086 15q11.2 15 19,301,328 19,916,381 615,053 15q11.2 15 20,813,690 21,041,101 227,411 15q11.2 15 22,004,834 22,248,614 243,780 15q11.2 15 26,055,804 26,752,132 696,328 15q13.1 16 21,002,359 21,198,510 19,6151 16p12.3 16 32,089,294 32,319,125 229,831 16p11.2 16 70,583,784 71,175,111 591,327 16q22.1-16q22.2	9	43,787,741	44,127,010	339,269	9p11.2
9 65,216,018 65,523,621 307,603 9q13 9 65,617,653 65,814,846 197,193 9q13-9q21.11 10 46,138,281 46,380,466 242,185 10q11.22 10 57,969,110 58,079,974 110,864 10q21.1 14 18,070,001 18,405,573 335,572 14q11.2 15 18,898,763 19,129,849 231,086 15q11.2 15 19,301,328 19,916,381 615,053 15q11.2 15 20,813,690 21,041,101 227,411 15q11.2 15 22,004,834 22,248,614 243,780 15q11.2 15 22,004,834 22,248,614 243,780 15q11.2 16 21,002,359 21,198,510 196,151 16p12.3 16 32,089,294 32,319,125 229,831 16p11.2 16 70,583,784 71,175,111 591,322 16q22.1-16q22.2 18 42,695,217 42,966,859 271,642 18q21.1 <	9	63,574,163	64,135,077	560,914	9q13
965,617,65365,814,846197,1939q13-9q21.111046,138,28146,380,466242,18510q11.221057,969,11058,079,974110,86410q21.11418,070,00118,405,573335,57214q11.21518,898,76319,129,849231,08615q11.21519,301,32819,916,381615,05315q11.21520,813,69021,041,101227,41115q11.21522,004,83422,248,614243,78015q11.21526,055,80426,752,132696,32815q13.11621,002,35921,198,510196,15116p12.31632,089,29432,319,125229,83116p11.21842,695,21742,668,59271,64218q21.12058,697,37558,867,659170,28420q13.32-20q13.332121,603,69721,743,689139,99221q21.12214,440,10315,014,058573,95522q11.1X88,109,77088,293,362183,592Xq21.31X88,109,77088,293,362183,592Xq21.31Y9,598,7909,717,440118,650YY21,03,62021,264,863161,243YY21,103,62021,264,863161,243YY25,055,85225,388,816292,964YY25,055,8927,637,569207,980Y	9	65,216,018	65,523,621	307,603	9q13
10 46,138,281 46,380,466 242,185 10q11.22 10 57,969,110 58,079,974 110,864 10q21.1 14 18,070,001 18,405,573 335,572 14q11.2 15 18,898,763 19,129,849 231,086 15q11.2 15 19,301,328 19,916,381 615,053 15q11.2 15 20,04,834 22,248,614 243,780 15q11.2 15 22,004,834 22,248,614 243,780 15q11.2 15 22,004,834 22,248,614 243,780 15q11.2 16 21,002,359 21,198,510 196,151 16p12.3 16 21,002,359 21,198,510 196,151 16p12.3 16 32,089,294 32,319,125 229,831 16p12.3 16 70,583,784 71,175,111 591,327 16q22.1-16q22.2 18 42,695,217 42,966,859 271,642 18q21.1 20 58,697,375 58,867,659 170,284 20q13.32-20q13.33 21 21,603,697 21,743,689 139,992 21q21.1	9	65,617,653	65,814,846	197,193	9q13–9q21.11
1057,969,11058,079,974110,86410q21.11418,070,00118,405,573335,57214q11.21518,898,76319,29,849231,08615q11.21519,301,32819,916,381615,05315q11.21520,813,69021,041,101227,41115q11.21522,004,83422,248,614243,78015q11.21526,055,80426,752,132696,32815q13.11621,002,35921,198,510196,15116p12.31632,089,29432,319,125229,83116p11.21670,583,78471,175,111591,32716q22.1-16q22.21842,695,21742,966,859271,64218q21.12058,697,37558,867,659170,28420q13.32-20q13.332121,603,69721,743,689139,99221q21.12214,440,10315,014,058573,95522q11.31X88,109,77088,293,362183,592Xq21.31Y9,598,7909,717,740118,650YY19,570,01019,871,468301,458YY19,570,01019,871,468301,458YY22,095,85225,88,816292,964YY22,095,85225,88,816292,964Y	10	46.138.281	46.380.466	242,185	10a11.22
1418,070,00118,405,573335,57214q11.21518,898,76319,129,849231,08615q11.21519,301,32819,916,381615,05315q11.21520,813,69021,041,101227,41115q11.21522,004,83422,248,614243,78015q11.21526,055,80426,752,132696,32815q13.11621,002,35921,198,510196,15116p12.31632,089,29432,319,125229,83116p11.21670,583,78471,175,111591,32716q22.1-16q22.21842,695,21742,966,859271,64218q21.12058,697,37558,867,659170,28420q13.32-20q13.332121,603,69721,743,689139,99221q21.12214,440,10315,014,058573,95522q11.1X87,762,34787,994,542232,195Xq21.31X88,109,77088,293,362183,592Xq21.31Y9,598,7909,717,440118,650YY19,570,01019,871,468301,458YY21,103,62021,264,863161,243YY25,095,85225,388,816292,964YY27,429,58927,637,569207,980Y	10	57,969,110	58,079,974	110,864	10g21_1
1516,05,0719,129,849231,08615q11.21519,301,32819,916,381615,05315q11.21520,813,69021,041,101227,41115q11.21522,004,83422,248,614243,78015q11.21526,055,80426,752,132696,32815q13.11621,002,35921,198,510196,15116p12.31632,089,29432,319,125229,83116p11.21670,583,78471,175,111591,32716q22.1-16q22.21842,695,21742,966,859271,64218q21.12058,697,37558,867,659170,28420q13.32-20q13.332121,603,69721,743,689139,99221q21.12214,440,10315,014,058573,95522q11.1X88,109,77088,293,362183,592Xq21.31X88,109,77088,293,362183,592Xq21.31Y9,598,7909,717,440118,650YY19,570,01019,871,468301,458YY25,095,85225,388,816292,964YY25,095,85225,388,816292,964YY25,095,85225,388,816292,964YY25,095,85225,388,816292,964YY25,095,85225,388,816292,964YY25,095,85225,388,816292,964YY25,095,85225,388,816292,964YY2	14	18 070 001	18 405 573	335 572	14011.2
1516,05017,12,04725,0501541121519,301,32819,916,381615,05315q11.21520,813,69021,041,101227,41115q11.21522,004,83422,248,614243,78015q13.11621,002,35921,198,510196,15116p12.31632,089,29432,319,125229,83116p11.21670,583,78471,175,111591,32716q22.1-16q22.21842,695,21742,966,859271,64218q21.12058,697,37558,867,659170,28420q13.32-20q13.332121,603,69721,743,689139,99221q21.12214,440,10315,014,058573,95522q11.1X87,762,34787,994,542232,195Xq21.31X88,109,77088,293,362183,592Xq21.31X89,09,8309,017,798107,968YY9,598,7909,717,440118,650YY21,03,62021,264,863161,243YY25,095,85225,388,816292,964YY25,095,85225,388,816292,964YY25,095,85225,388,816292,964YY25,095,85225,388,816292,964Y	15	18 898 763	19 1 29 849	231 086	15g11.2
1519,50,52019,710,581013,03513411.21520,813,69021,041,101227,41115q11.21522,004,83422,248,614243,78015q11.21526,055,80426,752,132696,32815q13.11621,002,35921,198,510196,15116p12.31632,089,29432,319,125229,83116q22.1-16q22.21670,583,78471,175,111591,32716q22.1-16q22.21842,695,21742,966,859271,64218q21.12058,697,37558,867,659170,28420q13.32-20q13.332121,603,69721,743,689139,99221q21.12214,440,10315,014,058573,95522q11.1X87,762,34782,93,362183,592Xq21.31X88,109,77088,293,362183,592Xq21.31Y9,598,7909,717,440118,650YY19,570,01019,871,468301,458YY21,103,62021,264,863161,243YY25,095,85225,388,816292,964YY25,095,85225,388,816292,964YY27,095,85225,388,816292,964YY27,042,58927,637.569207,980Y	15	10,000,000	10 016 201	615 053	15411.2
1520,813,69021,041,101227,41115q11.21522,004,83422,248,614243,78015q11.21526,055,80426,752,132696,32815q13.11621,002,35921,198,510196,15116p12.31632,089,29432,319,125229,83116p11.21670,583,78471,175,111591,32716q22.1-16q22.21842,695,21742,966,859271,64218q21.12058,697,37558,867,659170,28420q13.32-20q13.332121,603,69721,743,689139,99221q21.12214,440,10315,014,058573,95522q11.1X87,762,34787,994,542232,195Xq21.31X88,109,77088,293,362183,592Xq21.31Y9,598,7909,717,440118,650YY19,570,01019,871,468301,458YY21,103,62021,264,863161,243YY25,095,85225,388,816292,964YY25,095,85225,388,816292,964YY27,429,58927,637.569207,980Y	15	19,301,328	17,710,301	015,055	15411.2
1522,004,83422,248,614243,78015q11.21526,055,80426,752,132696,32815q13.11621,002,35921,198,510196,15116p12.31632,089,29432,319,125229,83116p11.21670,583,78471,175,111591,32716q22.1-16q22.21842,695,21742,966,859271,64218q21.12058,697,37558,867,659170,28420q13.32-20q13.332121,603,69721,743,689139,99221q21.12214,440,10315,014,058573,95522q11.1X87,762,34787,994,542232,195Xq21.31X88,109,77088,293,362183,592Xq21.31Y9,598,7909,717,440118,650YY19,570,01019,871,468301,458YY21,103,62021,264,863161,243YY25,095,85225,388,816292,964YY25,095,85225,388,816292,964YY27,429,58927,637.569207,980Y	15	20,813,690	21,041,101	227,411	15q11.2
1526,055,80426,752,132696,32815q13.11621,002,35921,198,510196,15116p12.31632,089,29432,319,125229,83116p11.21670,583,78471,175,111591,32716q22.1-16q22.21842,695,21742,966,859271,64218q21.12058,697,37558,867,659170,28420q13.32-20q13.332121,603,69721,743,689139,99221q21.12214,440,10315,014,058573,95522q11.1X87,762,34787,994,542232,195Xq21.31X88,109,77088,293,362183,592Xq21.31Y9,598,7909,717,440118,650YY19,570,01019,871,468301,458YY21,103,62021,264,863161,243YY25,095,85225,388,816292,964YY25,095,85225,388,816292,964YY27,429,58927,637.569207,980Y	15	22,004,834	22,248,614	243,780	15q11.2
1621,002,35921,198,510196,15116p12.31632,089,29432,319,125229,83116p11.21670,583,78471,175,111591,32716q22.1-16q22.21842,695,21742,966,859271,64218q21.12058,697,37558,867,659170,28420q13.32-20q13.332121,603,69721,743,689139,99221q21.12214,440,10315,014,058573,95522q11.1X87,762,34787,994,542232,195Xq21.31X88,109,77088,293,362183,592Xq21.31Y9,598,7909,717,440118,650YY19,570,01019,871,468301,458YY21,103,62021,264,863161,243YY25,095,85225,388,816292,964YY27,429,58927,637.569207,980Y	15	26,055,804	26,752,132	696,328	15q13.1
1632,089,29432,319,125229,83116p11.21670,583,78471,175,111591,32716q22.1-16q22.21842,695,21742,966,859271,64218q21.12058,697,37558,867,659170,28420q13.32-20q13.332121,603,69721,743,689139,99221q21.12214,440,10315,014,058573,95522q11.1X87,762,34787,994,542232,195Xq21.31X88,109,77088,293,362183,592Xq21.31Y9,598,7909,717,440118,650YY19,570,01019,871,468301,458YY21,103,62021,264,863161,243YY25,095,85225,388,816292,964YY27,429,58927,637.569207,980Y	16	21,002,359	21,198,510	196,151	16p12.3
1670,583,78471,175,111591,32716q22.1-16q22.21842,695,21742,966,859271,64218q21.12058,697,37558,867,659170,28420q13.32-20q13.332121,603,69721,743,689139,99221q21.12214,440,10315,014,058573,95522q11.1X87,762,34787,994,542232,195Xq21.31X88,109,77088,293,362183,592Xq21.31Y9,598,7909,717,440118,650YY19,570,01019,871,468301,458YY21,103,62021,264,863161,243YY25,095,85225,388,816292,964YY27,429,58927,637.569207,980Y	16	32,089,294	32,319,125	229,831	16p11.2
1842,695,21742,966,859271,64218q21.12058,697,37558,867,659170,28420q13.32-20q13.332121,603,69721,743,689139,99221q21.12214,440,10315,014,058573,95522q11.12214,440,10315,014,058573,95522q11.1X87,762,34787,994,542232,195Xq21.31X88,109,77088,293,362183,592Xq21.31Y8,909,8309,017,798107,968YY9,598,7909,717,440118,650YY19,570,01019,871,468301,458YY21,103,62021,264,863161,243YY25,095,85225,388,816292,964YY27,429,58927,637.569207,980Y	16	70,583,784	71,175,111	591,327	16q22.1–16q22.2
2058,697,37558,867,659170,28420q13.32-20q13.332121,603,69721,743,689139,99221q21.12214,440,10315,014,058573,95522q11.1X87,762,34787,994,542232,195Xq21.31X88,097,8009,017,798107,968YY9,598,7909,717,440118,650YY19,570,01019,871,468301,458YY21,103,62021,264,863161,243YY25,095,85225,388,816292,964YY27,429,58927,637.569207,980Y	18	42,695,217	42,966,859	271,642	18g21.1
2121,603,69721,743,689139,99221q21.12214,440,10315,014,058573,95522q11.1X87,762,34787,994,542232,195Xq21.31X88,109,77088,293,362183,592Xq21.31Y8,909,8309,017,798107,968YY9,598,7909,717,440118,650YY19,570,01019,871,468301,458YY21,03,62021,264,863161,243YY25,095,85225,388,816292,964YY27,429,58927,637.569207,980Y	20	58,697,375	58,867,659	170,284	20a13.32-20a13.33
22 14,440,103 15,014,058 573,955 22q1.1 X 87,762,347 87,994,542 232,195 Xq21.31 X 88,109,770 88,293,362 183,592 Xq21.31 Y 8,909,830 9,017,798 107,968 Y Y 9,598,790 9,717,440 118,650 Y Y 19,570,010 19,871,468 301,458 Y Y 21,103,620 21,264,863 161,243 Y Y 25,095,852 25,388,816 292,964 Y Y 27,429,589 27,637.569 207,980 Y	21	21 603 697	21 743 689	139 992	21a21 1
X 87,762,347 87,994,542 232,195 Xq1.31 X 88,109,770 88,293,362 183,592 Xq1.31 Y 8,909,830 9,017,798 107,968 Y Y 9,598,790 9,717,440 118,650 Y Y 19,570,010 19,871,468 301,458 Y Y 21,103,620 21,264,863 161,243 Y Y 25,095,852 25,388,816 292,964 Y Y 27,429,589 27,637.569 207,980 Y	 22	14 440 103	15 014 058	573 055	22a11 1
C 67,792,347 67,394,342 252,193 Xd(21.31) X 88,109,770 88,293,362 183,592 Xq(21.31) Y 8,909,830 9,017,798 107,968 Y Y 9,598,790 9,717,440 118,650 Y Y 19,570,010 19,871,468 301,458 Y Y 21,103,620 21,264,863 161,243 Y Y 25,095,852 25,388,816 292,964 Y Y 27,429,589 27,637.569 207,980 Y	Y	97 760 247	87 004 540	222 105	229111 Ya21 21
A 66,109,770 66,293,562 183,592 Xq21.31 Y 8,909,830 9,017,798 107,968 Y Y 9,598,790 9,717,440 118,650 Y Y 19,570,010 19,871,468 301,458 Y Y 21,103,620 21,264,863 161,243 Y Y 25,095,852 25,388,816 292,964 Y Y 27,429,589 27,637.569 207,980 Y	^ V	0/,/02,34/	01,774,042	232,193	AUZ 1.31
Y 8,909,830 9,017,798 107,968 Y Y 9,598,790 9,717,440 118,650 Y Y 19,570,010 19,871,468 301,458 Y Y 21,103,620 21,264,863 161,243 Y Y 25,095,852 25,388,816 292,964 Y Y 27,429,589 27,637.569 207,980 Y	^	88,109,//0	00,293,30Z	103,592	Aq21.31
Y 9,598,790 9,717,440 118,650 Y Y 19,570,010 19,871,468 301,458 Y Y 21,103,620 21,264,863 161,243 Y Y 25,095,852 25,388,816 292,964 Y Y 27,429,589 27,637.569 207,980 Y	Y	8,909,830	9,017,798	107,968	Ŷ
Y19,570,01019,871,468301,458YY21,103,62021,264,863161,243YY25,095,85225,388,816292,964YY27,429,58927,637.569207.980Y	Y	9,598,790	9,717,440	118,650	Y
Y21,103,62021,264,863161,243YY25,095,85225,388,816292,964YY27,429,58927,637.569207.980Y	Y	19,570,010	19,871,468	301,458	Y
Y 25,095,852 25,388,816 292,964 Y Y 27,429,589 27,637.569 207.980 Y	Y	21,103,620	21,264,863	161,243	Y
Y 27,429,589 27,637,569 207,980 Y	Y	25,095,852	25,388,816	292,964	Y
	Y	27,429.589	27,637.569	207.980	Y

 Table 1. Position in the human genome (NCBI build 34) of 63 DNA segments with increased copy number relative to chimpanzee and gorilla, as determined by full-coverage BAC array CGH

be readily validated through design of quantitative PCR experiments and through evaluation of signatures of duplication events in the reference human genome sequence as we describe below. Since it is expected that genomic segments recently gained in the human lineage have originated through duplication of existing sequence, we evaluated the degree of overlap between segments identified by CGH and segments identified by in silico

Table 2. Strict pi	aralog	ous gene pairs	(reciprocal t	best BLAST matches) p	resent on human DN	VA se	gments with i	increased cop	y number relative to	o chimp	anzee	and gor	illa	
Gene 1	Chr	Start	End	Description	Gene 2	Chr	Start	End	Description	$d_{\mathbb{N}}$	$d_{\rm S}$	$d_{\rm N}/d_{\rm S}$	P(1.0)	P(0.5)
ENSC00000186301	~~	16,353,203	16,360,133	Similar to macrophage stimulating 1 (hepatcyte growth factor-like)	ENSG0000186715	-	16,462,486	16,468,660	No description	0.233	0.843	0.276	0.000	0.000
ENSG0000051415	-	103,423,056	103,655,958	α-Amylase, salivary	ENSG00000174876		103,584,694	103,592,978	α-Amylase salivary	0.015	0.061	0.245	0.000	0.044
ENSC0000132043 ENSC0000182639 ENSC0000183598		146,102,223 147,000,080 147,001,408	146,214,718 147,000,460 147,001,818	No description Histone H2B.J H3 histone family, H3 histone family, member M	ENSC0000168614 ENSC0000184678 ENSC0000183702		145,923,370 147,074,392 147,040,763	145,958,684 147,074,772 147,042,418	No description Histone H2B.Q H3 histone, family 2; H3 histone family, member	0.268 0.004 0.000	0.260 0.188 0.418	1.032 0.021 0.001	0.694 0.001 0.000	0.000 0.004 0.000
ENSC00000183558 ENSC00000182217		147,030,455 147,047,334	147,030,847 147,049,286	Histone H2A.O Histone H4	ENSG00000183717 ENSG00000183941	(147,039,210 147,020,824	147,039,773 147,022,776	Histone H2A.O Histone H4	0.000	0.001	0.387	0.758 0.797	0.811 0.743
ENSC00000173756	7 7	89,423,308	89,423,634	NG9 precursor NG9 precursor IG k chain V-II region	ENSG00000182563	7 7	89,330,327	89,330,632	No description	0.092	0.237	0.388	0.014	0.500
ENSG00000131426	2	89,897,439	89,897,914	GM607 precursor IG k chain V-I region HK 101 precursor	ENSG00000163245	2	89,301,505	89,301,971	IG κ chain V-I region	0.027	0.029	0.903	0.888	0.403
ENSC00000172038	2	89,912,079	89,912,595	IG k chain V-III region	ENSG00000186192	2	89,058,824	89,248,127	IG k chain C region	0.045	0.098	0.459	0.116	0.858
ENSC00000163184 ENSC00000169606 ENSC00000154927 ENSC00000154927 ENSC00000177631 ENSC00000184671	00040	89,957,240 132,594,457 132,690,833 49,320 69,006,181	89,957,539 132,595,857 132,735,739 78,099 69,062,049	CLL precursor (Heumatol factor) IG k chain V-I region No description No description Baculoviral IAP repeat-containing	ENSG00000186856 ENSG00000172981 ENSG00000186825 ENSG0000182141 ENSG0000185284	22 22 5	89,062,541 14,635,352 132,763,505 15,009,448 69,659,709	89,242,332 14,636,482 132,764,092 15,026,562 69,678,715	No description No description No description No description No description	0.000 0.031 0.000 0.327 0.336	0.021 0.165 0.006 0.582 0.582	0.001 0.191 0.001 0.562 0.318	0.040 0.000 0.118 0.000 0.000	0.105 0.001 0.210 0.003 0.003
ENSC00000182078 ENSC00000183666 ENSC00000180027 ENSC00000174133 ENSC00000170135	N N N N N	69,234,613 69,498,165 70,249,240 70,271,857 73,622,573	69,234,750 69,537,080 70,289,288 70,276,207 73,679,914	protein 1 (neuronal apoptosis inhibitory protein) No description No description SMA3 protein No description Transcription Transcription Transcription	ENSC00000186932 ENSC00000185761 ENSC00000183761 ENSC00000185759 ENSC00000186361 ENSC00000174428	N N N N N	70,293,834 69,239,339 69,997,497 70,010,592 73,962,893	70,293,971 69,301,070 70,038,832 70,014,942 73,977,192	No description No description SMA3 protein No description Transcription factor	0.000 0.000 0.000 0.039 0.052	0.001 0.001 0.001 0.026 0.026	0.001 0.000 0.410 1.514 0.524	0.766 0.682 0.730 0.001 0.002	0.874 0.832 0.813 0.000 0.820
ENSG0000123965	~	73,718,985	73,733,868	Postmeiotic segregation increased 2-like 5	ENSC00000135165	~	71,762,053	75,683,468	GTF2IRD2 Zona pellucida sperm-binding protein 3 precursor	0.007	0.009	0.793	0.802	0.611
													(cor	ntinued)

176 Genome Research www.genome.org

Gene 1	Chr	Start	End	Description	Gene 2	Chr	Start	End	Description	ď	ds	$d_{\rm N}/d_{\rm S}$	P(1.0)	P(0.5)
ENSC00000165178 ENSC00000181881 ENSC00000170341	ァァァ	73,984,535 143,339,071 143,406,977	73,999,836 143,443,973 143,407,891	No description No description Seven transmembrane	ENSC00000182487 ENSC00000181926 ENSC00000170356	ファフ	72,046,778 143,383,314 143,320,745	72,062,076 143,389,573 143,321,671	No description No description Seven transmembrane	0.003 0.000 0.000	0.023 0.000 0.004	0.137 0.858 0.001	0.007 0.991 0.104	0.072 0.832 0.189
ENSC0000106714	6	39,063,817	39,278,466	helix receptor Contactin associated protein-like 3 precursor (cell recognition	ENSG00000154529	6	39,704,516	39,839,734	helix receptor No description	0.009	0.020	0.422	0.229	0.809
ENSC00000147926 ENSC00000179828 ENSC00000170165 ENSC00000170161 ENSC00000176299	6 6 6 6 <u>7</u>	39,908,370 40,483,601 43,977,955 63,660,398 18,238,610	39,912,422 40,501,134 43,988,544 63,661,000 18,239,551	molecule CASPR3) No description No description No description No description Seven transmembrane	ENSG0000182355 ENSG0000182153 ENSG0000176057 ENSG0000186383 ENSG0000186383	9 9 15	39,345,699 41,776,397 43,907,570 41,506,025 19,866,208	39,351,955 41,793,931 43,918,159 41,506,582 19,867,149	No description No description No description No description Seven Transmembrane	0.009 0.000 0.015 0.015	0.006 0.003 0.003 0.055 0.028	1.560 0.001 0.001 0.269 1.034	0.295 0.154 0.147 0.091 0.942	0.005 0.261 0.251 0.420 0.097
ENSC00000176294	14	18,285,736	18,286,659	helix receptor Seven transmembrane	ENSG00000183706	15	19,880,105	19,881,055	helix receptor Seven transmembrane	0.077	0.191	0.402	0.000	0.356
ENSC00000175733 ENSC00000128731	15 15	21,016,715 25,958,549	21,019,583 26,169,609	helix receptor No description HECT domain and	ENSG00000184095 ENSG00000140181	15 15	19,057,902 20,829,990	19,060,769 20,883,368	helix receptor No description No description	0.083 0.020	0.091 0.034	0.913 0.599	0.756 0.017	0.025 0.391
EN SG00000153684 EN SG00000169861	15 16	26,233,151 32,106,704	26,236,568 32,113,996	No description No description	ENSC00000183629 ENSC00000182414	15 16	26,495,524 32,009,310	26,498,941 32,099,878	No description IG heavy chain V-III	0.000 0.305	0.001 0.533	0.001 0.572	0.590 0.036	0.671 0.611
ENSG00000183677	18	42,811,361	42,813,622	RNA polymerase II transcription factor SIII subunit A2	ENSG00000183791	18	42,806,560	42,808,200	region RNA polymerase II transcription factor SIII subunit	0.070	0.085	0.825	0.381	0.018
ENSG00000130538	22	14,823,378	14,824,325	(elongin A2) No description	ENSG00000186445	14	18,171,223	18,172,203	A3 (elongin A3) Seven transmembrane	0.008	0.000		0.066	0.016
ENSG0000185912	\succ	8,898,452	8,960,986	Testis-specific V_encoded protein	ENSG00000187194	≻	9,000,204	9,001,613	helix receptor No description	0.052	0.126	0.412	0.089	0.699
ENSG00000131007	≻	19,638,276	19,645,782	Transcript Y 9 protein	ENSG00000131009	≻	19,788,761	19,796,267	Transcript Y 9	0.000	0.002	0.001	0.821	0.941
EN SG0000169953	≻	19,788,522	19,830,805	Heat shock transcription factor, Y-linked	ENSC00000172468	≻	19,603,741	19,646,021	Heat shock transcription factor, Y-linked	0.000	0.002	0.001	0.601	0.633
Gene identifiers, coc compared to a null	rdinat 1 _N /d _s (es, and descrip of 1.0 or 0.5, r	otions are from respectively.	ו the Ensembl database	(v. 27.35a) and are ba	ased o	on the NCBI_3.	4 genome bui	ld. <i>P</i> -values for likeliho	ood ratio	os are g	liven for	observed	l d _N /ds

Genome Research 177

Table 2. Continued

analysis (BLAST matches >1 kb long with >90% identity, as described in Krzywinski et al. 2004) of the reference human genome sequence (NCBI_34). Whereas coverage of the genome by in silico predicted segmental duplications is 5.2%, in silico predicted duplications covered 73.4% of the sequence represented in the 63 CGH-identified chromosomal segments, more than an order of magnitude enrichment. This is consistent with the notion that copy number gains as identified by CGH have arisen through segmental duplication. While the sequence identity between in silico defined duplicates can suggest an approximate duplication date, only phylogenetic analysis allows a time interval to be directly estimated for the duplication event. Of the total 8499 in silico defined duplicated segments, 176 (2.1%) intersect with the portion of the genome that appears from CGH to be have been gained specifically in the human lineage, suggesting that these duplications occurred in the past 4 to 6 million years (the divergence time of human and chimpanzee). It is important to note, however, that this is a minimal estimate for the proportion of recent duplicates, since segments smaller than the resolution of the BAC array (~79 kb) will not have been detected by CGH.

A total of 192 non-pseudogene Ensembl genes were detected on the 63 duplicated segments. If these segments arose through segmental duplication, we would expect representation from paralogous genes within this set. The coding sequences of these genes were compared by reciprocal BLAST analysis (expect-value cutoff = 10^{-10}), which identified 41 strict paralogous gene pairs (82 genes total) (Table 2). For these genes, pairwise synonymous and nonsynonymous substitution rates were estimated for the aligned sequence using the codon substitution models of Yang and Nielsen (2000) as implemented in PAML (phylogenetic analysis by maximum likelihood) (Yang 1997). The mean d_N/d_S ratio of the gene set is 0.433, the median ratio is 0.388, and the highest observed ratio is 1.56. Likelihood ratio tests were applied to test the null hypothesis of equal rates of substitution (H₀: $d_{\rm N} = d_{\rm S}$) versus the alternative hypothesis (H_A: $d_{\rm N} \neq d_{\rm S}$). For 15 of the 41 genes, the null hypothesis had a nominal P-value <5%, and in all but one of these significant cases, $d_N/d_S < 1$, implying that purifying selection had been removing deleterious amino acid variants, and that both copies of the gene must have retained function for at least a fraction of the time subsequent to the gene duplication. An extreme case of protein conservation since the time of gene duplication would be if $d_N/d_s = 0$ (i.e., no amino acid changes were tolerated), and in this case it might be appropriate to compare the null hypothesis to the alternative with $d_N/d_S < 0.5$. This null hypothesis arises because we allow for one homolog to be strictly conserved $(d_N/d_S = 0)$ and the other to be strictly neutral $(d_N/d_S = 1)$, thus a somewhat more liberal test of one of the pair showing excess divergence considers as a null hypothesis the average of $d_N/d_S = 0.5$ (Thornton and Long 2002). This test identified as significant three genes that had $d_N/d_S > 0.5$ (Table 2), but in no case have these genes been functionally characterized.

We selected two duplicated loci (*AMY1A* and *CNTNAP3*) for evaluation by an independent method (real-time quantitative PCR; Taqman). These loci were sequenced in our human, chimpanzee, and gorilla samples, and then primer/probe sets were designed to regions of sequence that were perfectly conserved between duplicates and among species. Of note, a third human amylase family member (*AMY2B*) was present on the duplicated segment that contained the two copies of *AMY1A*; thus the amylase primer probe sets were designed to a region of exact sequence identity among all three amylases. Results from these PCR assays (Fig. 2) verify increased copy number of these loci in human versus chimpanzee and gorilla.

Discussion

Using full-coverage BAC array CGH, we have identified 63 genomic segments with an increased hybridization ratio in human versus chimpanzee. Because these segments also show an increased hybridization ratio in human versus gorilla, the most parsimonious explanation is that these CGH-defined segments have been duplicated very recently in human evolutionary history, subsequent to our divergence from chimpanzee. This interpretation is supported by the high representation within these segments of in silico defined human segmental duplications, and the verification by real-time quantitative PCR of copy number differences at selected loci. However, the formal possibility remains that some subset of these CGH-defined segments has been independently lost in both chimpanzee and gorilla, rather than gained in humans. Owing to high sequence similarity among these three closely related primates and the substantial length of the CGH BAC probes (~200 kb), it is exceedingly unlikely that sequence divergence is responsible for any observed differences. It must be considered that a portion of the genome does not represent the species tree but, rather, supports a chimp-gorilla clade over a chimp-human clade. For copy number differences in this portion of the genome, which remains to be accurately mapped, parsimony is not effective in assigning the ancestral copy number state. However, the fact that we have relied on gorilla as an outgroup should not have a significant impact on the results of the present study because we evoke parsimony only in the first data-filtering step of our analysis. Subsequent analysis is strictly focused on paralogous gene pairs within genomic segments with copy number alteration. Tandem duplication is a signature of DNA copy number increase, and provides a level of internal validation to our analysis. Furthermore, where we have done quantitative gene dosage analysis for further verification of human copy number gains (Fig. 2), the data have supported this interpretation.

Interestingly, we observe a substantial DNA copy number increase at chromosome 2q13 in human. This is the site of the telomeric fusion between chimpanzee chromosomes 12 and 13 in the human/chimp common ancestor that resulted in human



Figure 2. Relative copy number of the *AMY1A* and *CNTNAP3* loci in three unrelated chimpanzees and a single gorilla compared to a single pooled sample of human gDNA.

chromosome 2 (Yunis et al. 1980). Furthermore, the expansions of human chromatin adjacent to the centromeres of chromosomes 1, 9, and 16 that were noted by Yunis et al. (1980) were also observed in the present study (Table 1). Fortna et al. (2004) previously used cDNA arrays for comparative analysis of hominoid genomes, and found a total of 25 relatively large DNA segments (11.1 Mb average length) that appeared to be duplicated in the human lineage. Of the 63 human copy number gains we measured in the present study, 30 segments have coordinates that agree with this previous study. The remaining 33 segments appear to be novel findings from our complete coverage array.

We analyzed the gene content of the 63 chromosomal segments with increased copy number in human. Genes within these regions were subjected to reciprocal BLAST analysis to find duplicated copies. Among the 41 high-confidence paralogous gene pairs we detected, the most highly represented gene family is immunoglobulin (IGK) genes, with five paralogous pairs. This is consistent with earlier whole-genome comparative analysis of, for example, fly and mosquito (Christophides et al. 2002; Hill et al. 2002; Zdobnov et al. 2002) or rat and mouse (Gibbs et al. 2004), where immune-related gene families have been found to be prone to expansion. The second most highly represented gene family is histones, with four paralogous pairs. In particular, the present study highlights human-specific duplications of members of the core histone minor cluster at chromosome 1q21 (Table 2). Histone octamers (comprised of two proteins each from core histone multimember families H2A, H2B, H3, and H4) form the core of the nucleosome, and are among the most evolutionarily conserved of all proteins, with substantial sequence conservation between humans and organisms as divergent as sea urchin (Grunstein et al. 1976). It is possible, given the relatively strict structural requirements of the nucleosome and exceptional sequence conservation across species, that variation in gene dosage becomes a viable alternative to sequence variation for adapting histone expression and function in accordance with selective pressures. While it is not clear what selective advantage might be conferred by many of the human duplicated genes highlighted in this study, of primary interest are genes with a potential role in nervous system development. An interesting candidate in this regard is ENSG00000106714 (Contactin-associated protein, CNTNAP3) and its uncharacterized paralog ENSG00000154529. CNTNAP3 is a member of the neurexin family of cell recognition molecules. Neurexins and their membrane-bound ligands (neuroligins) are thought to mediate interactions between neurons, including synapse formation. CNTNAP3, a member of the NCP subgroup of neurexins, is expressed throughout the human brain and is important in ion channel localization and neuron-glial interactions (Spiegel et al. 2002). Should further investigation of CNTNAP3 and its paralog verify non-pseudogene status, it will be important to evaluate what role these genes might have in synaptic function.

Under neutral evolution, coding mutations will be fixed at the same rate as silent mutations, giving a d_N/d_S ratio of 1. The median d_N/d_S ratio observed in our gene set was 0.388, which is consistent with net purifying selection acting on these recently duplicated genes. This observation is consistent with previous reports of reduced d_N/d_S ratios between paralogous genes in *Drosophila* (Thornton and Long 2002) and *Arabidopsis* (Zhang et al. 2002), and is indicative of continued function of both gene copies subsequent to the duplication event. While several genes in our set had d_N/d_S ratios >1, and a strict application of the PAML test of H₀: $d_N/d_S = 1$ versus H_A: $d_N/d_S > 1$ identified only one case of positive selection, a more liberal test that uses $H_0: d_N/d_S = 1/2$ (and an alternative hypothesis with d_N/d_S as a free parameter) identifies an additional two unannotated genes with weak support for significant positive selection. This latter test might, instead, be considered a test for constraint, however, as it assumes that the original functional gene copy tolerates zero nonsynonymous changes.

The d_N/d_S ratios reported here are average ratios for the aligned length of each protein pair. Identification and sequencing of the strict orthologs of these genes in chimpanzee and additional primates will allow evaluation of synonymous and nonsynonymous substitution rates in a site-specific and lineage-specific manner and will likely yield further insight into human adaptive evolution. Further exploration of genes and noncoding functional sequences within the boundaries of these variable segments will be helpful for elucidating the genetic basis of human-specific traits.

Methods

Comparative genomic hybridization

Hybridizations were done using the whole-genome SMRT array (Ishkanian et al. 2004), which consists of amplified MseI fragments from 32,433 tiled Human BACs (Krzywinski et al. 2004) spotted in triplicate on two aldehyde-coated slides, and gives an effective resolving power of 79 kb. Test and reference DNAs were digested with MseI, labeled by random priming with the fluorescent nucleotide analogs Cy5-dCTP and Cy3-dCTP, and purified using Sephadex G50. For each hybridization experiment, the test and reference DNAs were combined and denatured, and repetitive sequences were blocked by coincubation with denatured human Cot-1 DNA (Invitrogen). Repeat-blocked DNA was then hybridized at 45°C for 48 to 72 h. After hybridization, the slides were washed at 45°C for 15 min in 80% DIGEasy hybridization solution (Roche Scientific), $2 \times$ SSC (pH 7), followed by three washes of 5 min at room temperature with $0.1 \times$ SSC (pH 7), 0.1% SDS, four rinses of 30 sec each in $0.1 \times$ SSC (pH 7) at room temperature, a brief rinse with deionized, distilled water, and then dried. Imaging was done using a Packard Biosciences Scan Array Express instrument.

Microarray analysis

An open-source software package, called MIA, was developed and implemented to extract intensity ratios from the raw images. The analysis of CGH array images is divided into several broad steps: (1) addressing, which consists of finding the location of subarrays and individual tiles containing one and only one spot; (2) segmentation, which consists of identifying the pixels belonging to the spot within each tile; (3) extraction of spot and background intensities; and (4) normalization of data. These steps are not specific to CGH experiments, and the software can therefore be useful for any two-color microarray work such as the analysis of gene expression with spotted arrays. Similar to the approach of Yang et al. (2001), the addressing and segmentation steps are performed on a combined 8-bit image obtained by a square-root transformation, but the intensities are extracted from the original raw 16-bit images. Several mathematical techniques were applied in order to obtain a completely automated addressing procedure. The average spacing between spots is deduced by analyzing the Fourier transform of one-dimensional spectra projected in both the horizontal and vertical directions. The average spot size is obtained by granulometry, more precisely, by studying the effect of successive morphological openings with structural elements of increasing size (Soille 2003). Once the subarrays are located, individual tiles are first positioned on a regular twodimensional square lattice, then an optional optimization can be performed for each tile. The Seeded Region Growing algorithm (Adams and Bischof 1994) was used in the segmentation process of each tile. A square seed (by default of size two by two pixels) associated with the spot is positioned at the center of intensity of the tile. The pixels on the edge of the tile serve as background seeds unless their intensity falls within the top 10% of that tile, and the pixels with intensities above that threshold form the seeds for artifacts. Therefore, following the application of the Seeded Region Growing algorithm, each pixel within the tile is assigned as being part of the spot, part of the background, or part of an image artifact. There are no geometrical constraints applied to the shape of the spot except that it is obviously constrained to the limits of the tile as determined at the addressing stage. Spot intensities are extracted from the raw 16-bit images. The spot pixels as determined in the segmentation phase are averaged to extract the spot intensity. Similar to the work of Yang et al. (2002), the background intensity is obtained by probing a morphological opening obtained with a large structural element (default of 2.5 times the average spot spacing). Intensity ratios were normalized with the help of the robust LOESS regression on the so-called M-A plot, where $M = \log_2(I_1/I_2)$ and $A = \log_2[(I_1 \times I_2)^{1/2}]$, with I_1 and I_2 being the background subtracted intensities of the spot in the two images.

We selected individual thresholds for each array using the same type of calculations used to compute box-and-whisker plots (Tukey 1977). In a box-and-whisker plot, the difference between the upper (H2) and lower (H1) hinges is called the H-spread. The definition of hinge is similar to that of quartile and therefore H-spread is similar to the interquartile range. The upper (lower) whisker is located 1.5 \times the H-spread above (below) the upper (lower) hinge, and values outside the whiskers are considered extreme. For each individual array, the thresholds for DNA copy number aberrations were set at the whisker levels. A MySQL relational database called CGHdb was built and populated with all ratio and ancillary data. Results from a human female (test) versus human male (reference) hybridization allowed estimation of true-positive, false-negative, and false-positive rates based on sex chromosome copy number. A total of 1134 of the 1430 Xchromosome clones had increased copy number, and 157 of the 196 Y chromosome clones had decreased copy number in the female test DNA sample. These values give a true-positive rate of 79.2% for increases and 80.1% for decreases, with corresponding false-negative rates of 20.8% for increases and 19.9% for decreases. Of the 30,216 autosomal clones, 665 (2.2%) showed increased copy number, and 592 (2.0%) showed decreased copy number. It is not possible to determine if these autosomal copy number differences reflect true positives or false positives.

Quantitative PCR

Primer/probe sets for two separate test genes (*AMY1A* and *CNTNAP3*) were designed using the repeat masked reference human genome sequence (NCBI_34; April 2003 release; http:// genome.ucsc.edu/) (Table 1). Primers and probes were selected in regions of exact match between the test gene and its top paralog, as defined below (Identification of Paralogs). The *SNAP25* gene, previously established to be single copy in the human genome (Bailey et al. 2002), was selected as a reference gene. The target sequences for test and reference primers and probes were verified in our human, chimp, and gorilla samples by direct sequencing. To determine relative copy number, 10 ng of genomic DNA was assayed in triplicate in 20-µL reactions containing 1× final con-

centration TaqMan Universal Master Mix (ABI part number 4304437), 200 nM each primer and probe, and 10 ng of template DNA. Each experiment was performed using a 384-well optical PCR plate and the ABI 7900HT machine with default cycling conditions. Copy number of the test locus in chimp or gorilla versus human was defined as $2^{-\Delta\Delta C_T}$, where ΔC_T is the difference in threshold cycles for the test and reference loci.

Identification of paralogs

We sought to identify paralogous genes within the 55 recently duplicated human genomic regions. cDNA sequences for 206 Ensembl genes from these regions were collected with the Ensembl API v.27_35a. Using BioPerl v1.4 (Stajich et al. 2002), all sequences were transformed into peptides. Each gene's best partner within this set was determined by using the BLAST algorithm (Altschul et al. 1997) (*E*-value 10^{-10} , ungapped alignments) against the peptide library. Once a gene's best partner was identified, the BLAST algorithm was rerun with the best hit as the query sequence. If, in the second iteration, the best BLAST hit was the original query sequence, then this gene pair was deemed paralogous. High-scoring segment pairs (HSPs) from the aligned regions were extracted in-frame in nucleotide form for subsequent analysis of substitution rates.

Acknowledgments

We thank Wan Lam for fabrication of arrays, and Evan Eichler for in silico predictions of segmental duplications in the human reference genome sequence. We thank the Canadian Institutes of Health Research and the BC Cancer Agency for funding this study. R.A.H. is a Michael Smith Foundation for Health Research scholar. A.G.C. is supported by NIH grant HG003229.

References

- Adams, R. and Bischof, L. 1994. Seeded region growing. *IEEE Trans. Pattern Anal. Machine Intell.* **16:** 641–647.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389–3402.
- Bailey, J.A., Gu, Z., Clark, R.A., Reinert, K., Samonte, R.V., Schwartz, S., Adams, M.D., Myers, E.W., Li, P.W., and Eichler, E.E. 2002. Recent segmental duplications in the human genome. *Science* 297: 1003–1007.
- Chen, F.-C. and Li, W.-H. 2001. Genomic differences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am. J. Hum. Genet.* **68:** 444–456.
- Chimpanzee Sequencing and Analysis Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**: 69–87.
- Christophides, G.K., Zdobnov, E., Barillas-Mury, C., Birney, E., Blandin, S., Blass, C., Brey, P.T., Collins, F.H., Danielli, A., Dimopoulos, G., et al. 2002. Immunity-related genes and gene families in *Anopheles* gambiae. Science **298**: 159–165.
- Fortna, A., Kim, Y., MacLaren, E., Marshall, K., Hahn, G., Meltesen, L., Brenton, M., Hink, R., Burgers, S., Hernandez-Boussard, T., et al. 2004. Lineage-specific gene duplication and loss in human and great ape evolution. *PLoS Biol.* 2: e207.
- Fujiyama, A., Watanabe, H., Toyoda, A., Taylor, T.D., Itoh, T., Tsai, S.F., Park, H.S., Yaspo, M.L., Lehrach, H., Chen, Z., et al. 2002. Construction and analysis of a human–chimpanzee comparative clone map. *Science* **295**: 131–134.
- Gibbs, R.A., Weinstock, G.M., Metzker, M.L., Muzny, D.M., Sodergren, E.J., Scherer, S., Scott, G., Steffen, D., Worley, K.C., Burch, P.E., et al. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**: 493–521.
- Grunstein, M., Schedl, P., and Kedes, L. 1976. Isolation and sequence analysis of sea urchin (*Lytechinus pictus*) histone H4 messenger RNA. *J. Mol. Biol.* **104:** 351–369.

- Hill, C.A., Fox, A.N., Pitts, R.J., Kent, L.B., Tan, P.L., Chrystal, M.A., Cravchik, A., Collins, F.H., Robertson, H.M., and Zwiebel, L.J. 2002. G protein-coupled receptors in Anopheles gambiae. Science 298: 176-178.
- Ishkanian, A.S., Malloff, C.A., Watson, S.K., DeLeeuw, R.J., Chi, B., Coe, B.P., Snijders, A., Albertson, D.G., Pinkel, D., Marra, M.A., et al. 2004. A tiling resolution DNA microarray with complete coverage of the human genome. *Nat. Genet.* **36**: 299–303.
- Krzywinski, M., Bosdet, I., Smailus, D., Chiu, R., Mathewson, C., Wye, N., Barber, S., Brown-John, M., Chan, S., Chand, S., et al. 2004. A set of BAC clones spanning the human genome. Nucleic Acids Res. 32: 3651-3660.
- Locke D.P., Segraves, R., Carbone, L., Archidiacono, N., Albertson, D.G., Pinkel, D., and Eichler, E.E. 2003. Large-scale variation among human and great ape genomes determined by array comparative genomic hybridization. Genome Res. 13: 347-357.
- Newman, T.L., Tuzun, E., Morrison, V.A., Hayden, K.E., Ventura, M., McGrath, S.D., Rocchi, M., and Eichler, E.E. 2005. A genome-wide survey of structural variation between human and chimpanzee. Genome Res. 15: 1344-1356.
- Ohno, S. 1970. Evolution by gene duplication. Springer-Verlag, Berlin.
- She, X., Jiang, Z., Clark, R.A., Liu, G., Cheng, Z., Tuzun, E., Church, D.M., Sutton, G., Halpern, A.L., and Eichler, E.E. 2004. Shotgun sequence assembly and recent segmental duplications within the human genome. Nature 431: 927-930.
- Soille, P. 2003. Morphological image analysis: Principles and applications, 2d ed. Springer-Verlag, Berlin, Heidelberg.
- Spiegel, I., Salomon, D., Erne, B., Schaeren-Wiemers, N., and Peles, E. 2002. Caspr3 and caspr4, two novel members of the caspr family are expressed in the nervous system and interact with PDZ domains. Mol. Cell. Neurosci. 20: 283–297.
- Stajich, J.E., Block, D., Boulez, K., Brenner, S.E., Chervitz, S.A., Dagdigian, C., Fuellen, G., Gilbert, J.G., Korf, I., Lapp, H., et al. 2002. The Bioperl toolkit: Perl modules for the life sciences. Genome Res. 12: 1611-1618.

- Thornton, K. and Long, M. 2002. Rapid divergence of gene duplicates on the Drosophila melanogaster X chromosome. Mol. Biol. Evol. 19: 918-925
- Tukey, J.W. 1977. Explanatory data analysis. Addison-Wesley, Reading, MA.
- Watanabe, H., Fujiyama, A., Hattori, M., Taylor, T.D., Toyoda, A., Kuroki, Y., Noguchi, H., BenKahla, A., Lehrach, H., Sudbrak, R., et al. 2004. DNA sequence and comparative analysis of chimpanzee chromosome 22. *Nature* **429**: 382–388. Yang, Z. 1997. PAML: A program package for phylogenetic analysis by
- maximum likelihood. Comput. Appl. Biosci. 13: 555-556.
- Yang, Z. and Nielsen, R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol. Biol. Evol. 17: 32-43.
- Yang, Y.H., Buckley, M.J., and Speed, T.P. 2001. Brief. Bioinformatics **2:** 341–349.
- Yang, Y.H., Dudoit, S., Luu, P., Li, D.M., Peng, V., Ngai, J., and Speed, T.P. 2002. Normalization for cDNA microarray data: A robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res. 30: e15.
- Yunis, J.J., Sawyer, J.R., and Dunham, K. 1980. The origin of man: A chromosomal pictorial legacy. Science 208: 1145-1148.
- Zdobnov, E.M., von Mering, C., Letunic, I., Torrents, D., Suyama, M., Copley, R.R., Christophides, G.K., Thomasova, D., Holt, R.A., Subramanian, G.M., et al. 2002. Comparative genome and proteome analysis of Anopheles gambiae and Drosophila melanogaster. Science **298:** 149–159.
- Zhang, L., Vision, T.J., and Gaut, B.S. 2002. Patterns of nucleotide substitution among simultaneously duplicated gene pairs in Arabidopsis thaliana. Mol. Biol. Evol. 19: 1464-1473.

Received July 19, 2005; accepted in revised form November 9, 2005.