 Human Molecular Genetics, 2001, Vol. 10, No. 20 2215–2223

Lessons from the human genome: transitions between euchromatin and heterochromatin

Julie E. Horvath, Jeffrey A. Bailey, Devin P. Locke and Evan E. Eichler*

Department of Genetics and Center for Human Genetics, Case Western Reserve School of Medicine and University Hospitals of Cleveland, BRB 720, 10900 Euclid Avenue, Cleveland, OH 44106, USA

Received July 2, 2001; Accepted July 12, 2001

The publication of the human genome draft sequence provides, for the first time, a global view of the structural properties of the human genome. Initial sequence analysis, in combination with previous published reports, reveals that more than half of the transition regions between euchromatin and centromeric heterochromatin contain duplicated segments. The individual duplications originate from diverse euchromatic regions of the human genome, often containing intron–exon structure of known genes. Multiple duplicons are concatenated together to form larger blocks of wall-to-wall duplications. For a single chromosome, these paralogous segments can span >1 Mb of sequence and define a buffer zone between unique sequence and tandemly repeated satellite sequences. Unusual pericentromeric interspersed repeat elements have been identified at the junctions of many of these duplications. Phylogenetic and comparative studies of pericentromeric sequences suggest that this peculiar genome organization has emerged within the last 30 million years of human evolution and is a source of considerable genomic variation between closely related primate species. Interestingly, not all human pericentromeric regions show this proclivity to duplicate and transpose genomic sequence, suggesting at least two different models for the organization of these regions.

INTRODUCTION

The first studies on the human genome were focused, by necessity, on specific genes and to a lesser extent on non-coding regions such as centromeres and telomeres. The techniques of the time were limiting and typically only small-scale analyses were possible. With the advent of the Human Genome Project, technologies advanced rapidly and the ability to perform large-scale sequencing grew tremendously. For the last 2 years, however, sequence data accumulation has far outpaced data analysis. Recent projects have shown that whole chromosome (1,2) and even entire genome (3–5) analyses, while possible, are still in their infancy. With the completion of the working draft phase of the Human Genome Project, the significance of novel patterns of genomic architecture may now be assessed in a genome-wide fashion.

In the initial overview of our genome organization, we have learned that a mere 1.5% of our genome is coding sequence (3,4) and that gene density varies dramatically among chromosomes and within specific chromosomal regions (6). The majority of genes are located a significant distance from the heterochromatic centromeres and telomeres (3) and there is a transition zone between the genic regions and satellite heterochromatin which we refer to as the pericentromeric region (7). These regions are often overlooked, primarily due to their gene-poor nature and complex organization. Consequently, many of these regions are poorly assembled by current sequence assembly efforts (5). A complete understanding of the relationship between chromosome structure and function requires that these transition regions be fully resolved. Even chromosomes reportedly sequenced to 'completion' such as chromosome 22 have multiple gaps in their pericentromeric regions (1). Recent research investigating the structure of pericentromeric and subtelomeric regions has suggested that these regions are often composed of interchromosomally duplicated (paralogous) genic segments and that some of the difficulties in providing sequencing closure of these regions are due to their paralogous organization (5). In this review we will summarize the emerging properties of highly duplicated pericentromeric regions and discuss implications for their role in evolution.

PERICENTROMERIC DUPLICATION BIAS

Initial characterization of the pericentromeric regions of 2p11, 10p11 and 16p11 indicated that they contained highly duplicated sequences and suggested that a similar phenomenon could be expected on many other human chromosomes (8–12). The sequencing of chromosomes 21 and 22 showed that these small acrocentric chromosomes, like 2p11, 10p11 and 16p11, also contained highly duplicated pericentromeric regions (1,2). Paralogous segments were found along the entire length of chromosomes 21 and 22, but the vast majority of duplicated sequence was restricted to pericentromeric and subtelomeric regions (2,3). Specifically, the most proximal 1 Mb and most distal 30 kb on chromosome 21 and the most proximal 1.5 Mb and distal 50 kb on chromosome 22 harbored the bulk of the duplicated sequences (3). The proximal 1.5 Mb of chromosome

*To whom correspondence should be addressed. Tel: +1 216 368 4883; Fax: +1 216 368 3432; Email: eee@po.cwru.edu

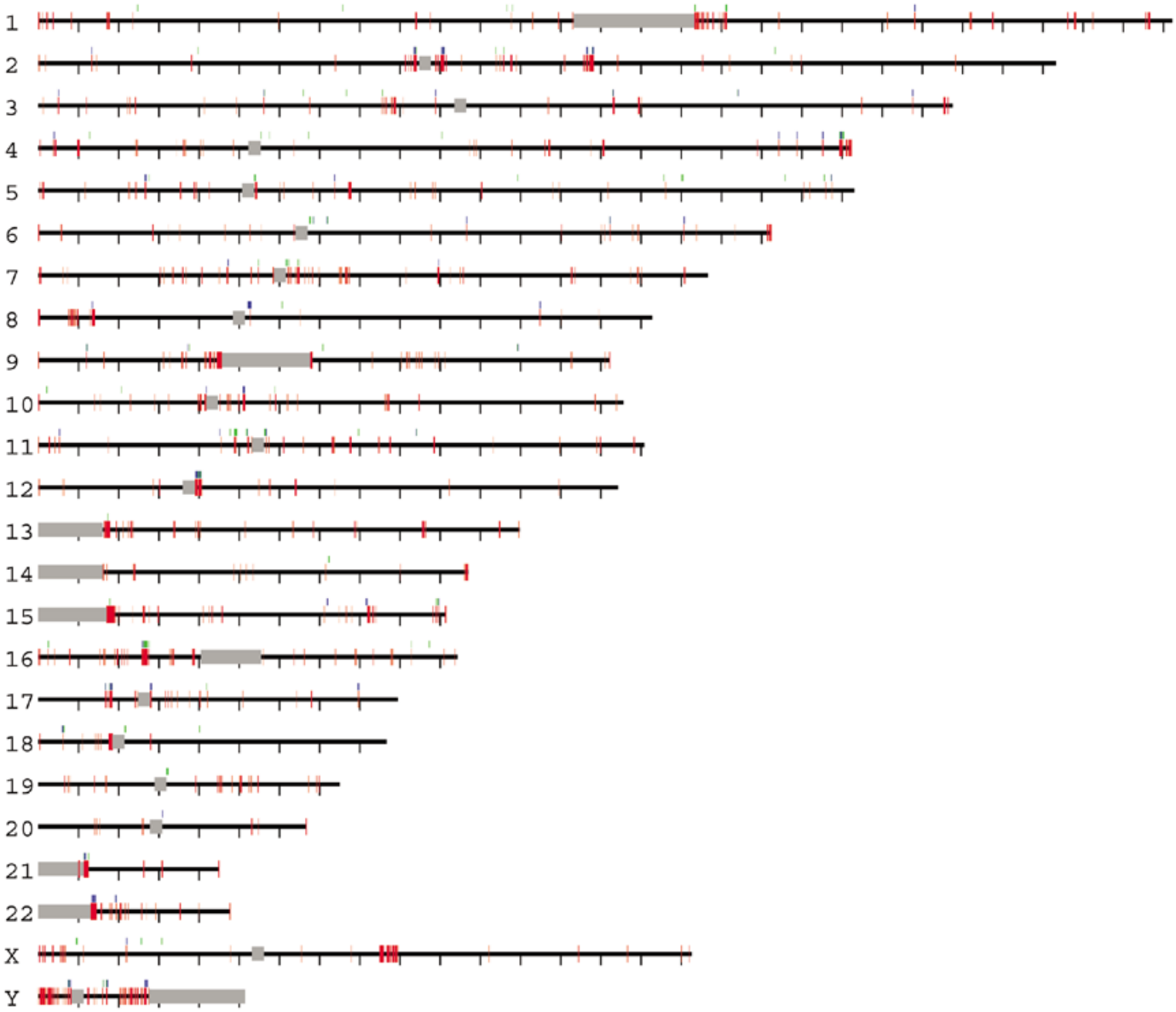

Figure 1. Genome-wide view of interchromosomal duplications. The location of interchromosomal duplicated segments (>90% sequence identity, >10 kb in length) is shown (red bars, not drawn to scale). The analysis is based on the most recent published assembly of the human genome (October 2000, http://genome.ucsc.edu/). Each horizontal black line represents a human chromosome with vertical black bars occurring every 10 Mb. Centromeres, satellite DNAs and acrocentric p arms are indicated by gray boxes. Blue boxes indicate satellite sequences (HSAT I, II, III, γ, β and α satellite) and green lines represent previously characterized pericentromeric specific repeats (CAGGG, GGGCAAAAAGCCG). In some cases (see 16p), pericentromeric regions have not been assembled in the correct location. Approximately 50% of interchromosomal duplications associate with satellite repeat sequences. This figure was adapted from Bailey *et al.* (5).

22 contains 52% of all interchromosomal duplications while containing only 5% of the chromosome sequence (3). The pattern of genome-wide interchromosomal duplications depicting the extent of the pericentromeric bias is shown in Figure 1.

There are at least two possible explanations for the occurrence of duplicated segments within pericentromeric regions. First, pericentromeric regions may be one of the only regions in the genome that are able to accept duplicated material without severe consequence to the organism (9). However, contrary to the predictions of such a model, large genomic regions outside of pericentromeric regions (>10 Mb) have been discovered which are virtually devoid of genes (2,6), and these regions do not show a proclivity to acquire duplicated sequences. This suggests that if the absence of selective constraint is the basis for this pericentromeric bias, lowered gene density is not the cause. A second model proposes that hyper-recombinogenic sequences localized within pericentromeric regions may actively attract segmental duplications. This is based on the observation that unusual polypyrimidine and polypurine mini-satellite-like sequences (similar to χ recombination signals) have been identified precisely at many duplication breakpoints

Table 1. Recent interchromosomal pericentromeric duplications

Gene	Name	Ancestral locus	Duplicated loci	Reference
ABCD1	ATP binding cassette, subfamily D (ALD), member 1	Xq28	2p11, 10p11, 16p11, 22q11	(14)
CTR/CDM	SLC6A8/DXS1357E	Xq28	16p11[a]	(13)
HERC2	hect domain and RLD 2	15q13	15q11[a], 16p11[a]	(20)
FGF7	Fibroblast growth factor (keratinocyte growth factor)	15q15–q21.1	2q21, 9p11, 9q12–q13, 18p11, 18q11, 21q11	(17)
NF1	Neurofibromin 1	17q11.2	14q11[a], 15q11[a], 22q11, 2q21, 18?p/q11, 21q11, 12q12, 1p32, 20p?/q11	(16,31,32)
IGHV	Immunoglobulin heavy variable	14q32.2	15q11.2, 16p11.2	(26,27)
Igk-V	Immunoglobulin κ chain complex variable region	2p11/2p12	1q12, 2p11, 9p11, 9q11, 22q11	(19,28)
IGLV	Immunoglobulin λ variable	22q11	8q11	(21)
N/A	Immunoglobulin D	14q32.3	15q11.2	(26)
VWF	von Willebrand factor	12p13.3	22q11	(22)
N/A	Hs. 135840	4q24	2p11, 10p11, 16p11, 22q11, 14cen, Ycen	(11)
TPTE	Transmembrane phosphatase with tensin homology	21p11	13, 15, 22cen, Y	(30)
PLG	Plasminogen	6q26	2p11/q11	(23)
FRG1	FSHD region gene 1	4q35	9cen, 13p, 14p, 15p, 21p, 20cen, 8, 12, 22	(24)
IGSF3	Immunoglobulin super family 3	1p13	2cen, 13, Y	(29)
GGT	γ-Glutamyltransferase	22q11	13, 20, 22q11[a], Y	(25,29)

[a]Multiple copies are located within this region.
N/A, no formal gene designation.

(8,10,13,14). A phylogenetic analysis of one of these pericentromeric interspersed repeat (PIR) sequences revealed that the repeat existed within the pericentromeric region prior to the integration of the duplicated sequences (15). Not all duplication boundaries, however, are clearly delineated by such repeats. Further specialized comparative sequencing efforts in these regions should provide insight into the mechanistic aspects of this duplication process.

DUPLICON STRUCTURE

Many of the segments duplicated to pericentromeric regions show the presence of ancestral intron–exon gene structure (8,9,11,13,14,16,17). These blocks of duplicated sequence which were transposed to a new location were termed duplicons or orphons to distinguish them from repeat sequences arising by other processes, such as retrotransposition or tandem duplication (14,18). Since these initial studies, a wide variety of partial gene duplications and repeats have been catalogued within pericentromeric regions (Table 1) (8,13,14,16,17,19–32). The duplicons range from <1 kb to as large as 85 kb in length. No pattern in the genomic distribution of the ancestral loci can be discerned. In most cases, the pericentromeric duplicates appear to be non-functional. This is due to the partial nature of the duplication itself, which often occurs without essential upstream regulatory sequences or a complete complement of exons. In some cases, however, mRNA or EST sequences corresponding to the pericentromeric copies have been identified (33,34). Expression profiles for several of these indicate that transcription for these pericentromeric paralogues are restricted to germline, fetal or cancerous tissues. The functional significance of these transcripts is unknown. As more pericentromeric regions are sequenced and analyzed, it is likely that many more partial gene structures embedded within pericentromeric regions will be discovered.

PROXIMITY TO SATELLITES

Detailed analyses of several pericentromeric regions reveal a similar structural organization. Paralogous segments on chromosomes 10, 16, 21 and 22 are all found in close proximity to satellite sequences (1,10,11,29). For example, a variety of satellite sequences in the pericentromeric regions of chromosomes 10 and 22 are embedded within the duplicated segments themselves (1,10,35). Our analysis of 16p11 indicated that a duplicon from Xq28 and another from 4q24 abut monomeric α satellite near the chromosome 16 centromere (11). In addition, chromosome 21 duplicons are structured such that they are located within tens of kilobases of satellite 1 sequences (29,36). The close proximity of these paralogous segments to classically defined pericentromeric repeats suggests that, for at least a subset of pericentromeric regions, duplications demarcate the boundary between satellite and non-satellite sequences.

A schematic of the chromosome 2 pericentromeric region clearly demonstrates the association of interchromosomal duplications and satellite sequences (Fig. 2). Human chromosome 2 was formed through a fusion event of two ancestral primate chromosomes (syntenic to chimpanzee XII and XIII) early during hominid evolution (<5 million years ago) (37–40). As a consequence of this fusion, α satellite DNA has been mapped to both the active centromere and to the vestigial

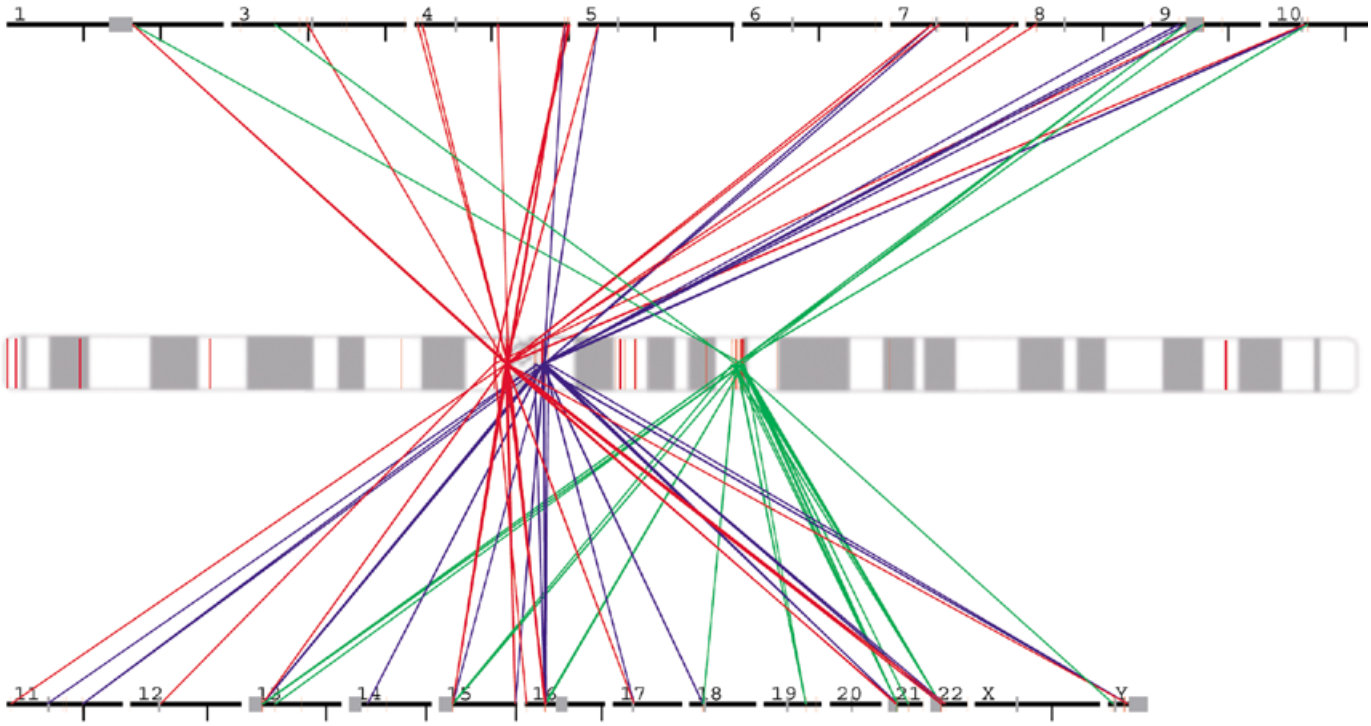

Figure 2. Human chromosome 2 pericentromeric interchromosomal duplications. The pattern of interchromosomal paralogy is shown for 2p11, 2q11 and 2q21 (ancestral centromere). A chromosome 2 ideogram is shown scaled to 10 times the length of other chromosomes. Other chromosomes are represented as horizontal black lines above and below the chromosome 2 ideogram, with vertical bars occurring every 50 Mb. Centromeres are represented as gray boxes. All colored, diagonal lines represent pairwise sequence comparisons that are >90% identical over >10 kb. Red lines represent all segments paralogous to 2p11, blue lines represent segments paralogous to 2q11 and green lines represent segments paralogous to 2q21 (the ancestral primate centromere). The red boxes, representing interchromosomal duplications, are not drawn to scale. Intrachromosomal duplications are not shown (i.e. 2p11 and 2q11).

centromere at 2q21 (41). Interestingly, three clusters of interchromosomal duplications are found on chromosome 2, two regions (2p11 and 2q11) on either side of the active centromere and one cluster of interchromosomal duplications located at the ancestral centromere. Analysis of one of the duplicons from 2q21 (NF1) (16) indicates that the duplications arose in our primate ancestor prior to the chromosomal fusion event.

BREAKPOINT SEQUENCES AND A MODEL OF DUPLICATION

By investigating the sequences flanking the pericentromeric duplications, a general view of the mechanism that gave rise to these complex structures can be proposed. Such analyses have identified short polypyrimidine or polypurine imperfect repeats such as CAGGG, GGGCAAAAAGCCG, GGAA and HSREP522 elements (8,10,15,35). Many of these elements are similar to telomeric repeats, subtelomeric sequences and immunoglobulin switch sequences. CAGGG repeat sequences were identified at the duplication integration boundaries of several duplicons (including the creatine-transporter and a subset of the NF1 duplicons) (13,15). GGGCAAAAAGCCG sequences demarcate the transition points at one end of ALD duplicons (13,14). These findings were corroborated by the analysis of chromosome 10 paralogous segments, which indicated that many duplicons terminate at one of these GC-rich repeat sequences as well (10).

The CAGGG, GGGCAAAAAGCCG and HSREP522 motifs are direct, non-tandem interspersed repeats and contain runs of guanines. Their sequence organization is reminiscent of G-rich (G4) DNA which facilitates DNA–DNA interactions by Hoogsteen pairing (13,15,42). G-rich DNA has been identified in four distinct genomic regions: heavy chain switch recombination regions, rDNA, telomeres and, more recently, pericentromeric regions (15,42). All four of these regions have been shown to be actively involved in recent duplication and/or gene conversion events (43,44). It is thought that these G4 DNA sequences, which have been implicated in meiotic pairing of homologous chromosomes, may also facilitate pairing between non-homologous pericentromeric regions (13,15). Furthermore, the interspersed organization of CAGGG and GGGCAAAA-AGCCG motifs are similar in structure to immunoglobulin switch recombination sequences (13,14) that promote non-homologous sequence exchange events (45). Such sequences, in theory, may have facilitated the integration of duplicated sequences within pericentromeric regions and their dispersal during evolution. Although these GC-rich repeat sequences are found at transition points of many duplicons, it should be noted that some duplicons do not appear to have any repeat structure at their boundaries. In addition, AT-rich sequences incapable of forming G4 DNA have been identified at other boundaries (10).

Large-scale sequence comparisons of 2p11, 10p11, 16p11 and 22q11 indicated that these regions contained many of the same duplicated segments in a similar organization. This

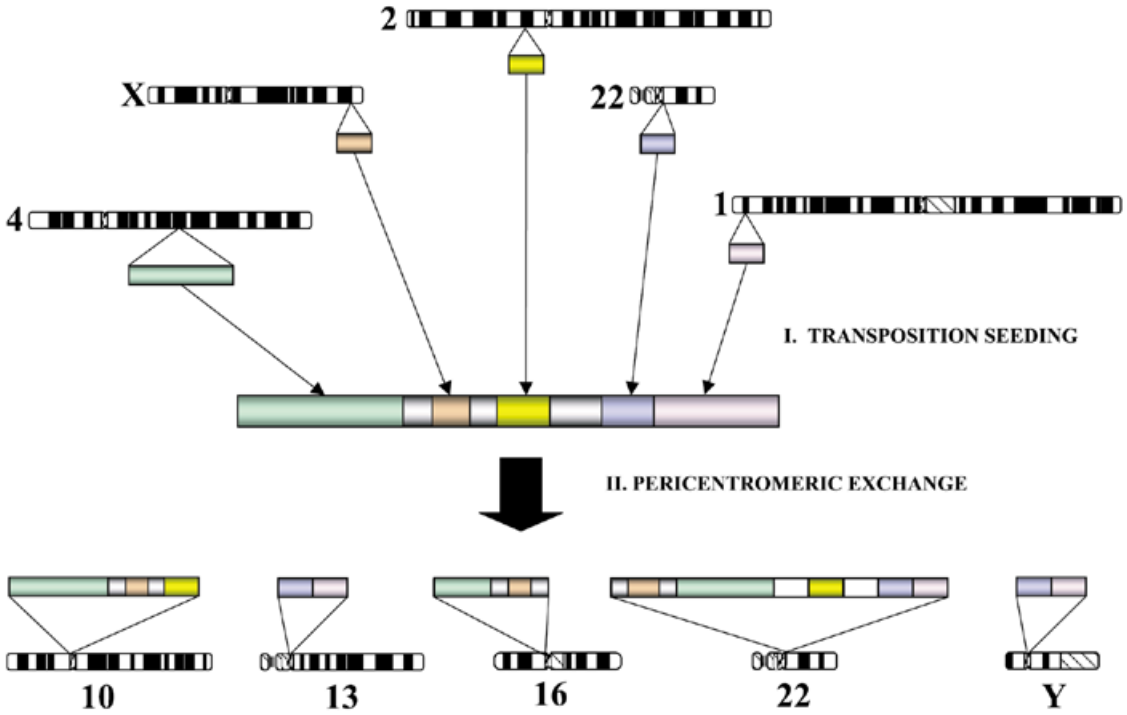

Figure 3. Two-step model of pericentromeric duplication. A two-step model is proposed to explain the accumulation of pericentromerically duplicated segments originating from non-homologous chromosomes. First, genomic segments duplicatively transpose to an ancestral pericentromeric region (transposition seeding). Multiple duplicated segments accumulate in the pericentromeric region, generating a mosaic of duplicated sequences. Later, larger blocks containing multiple duplicated segments are spread to other pericentromeric regions (pericentromeric exchange). These pericentromeric regions are subject to further rearrangements, scrambling the order of several segments. A 200 kb segment from the pericentromeric region of 2p11 is shown as an example. Colored bars represent the ancestral duplicated segments (duplicons) and open boxes represent uncharacterized intervening sequence.

organization indicated a pattern of duplications within duplications. We have proposed a two-step duplication model to explain these properties (11,14,32) (Fig. 3). Initially, material from donor loci invades a pericentromeric region. This material often contains genic sequence with intact exon–intron structure and therefore must be duplicated through a DNA transposition not retroposition process. The GC-rich repeat sequences may have a structural susceptibility to such DNA transposition events or other characteristics of pericentromeric DNA may preferentially recruit these duplicated segments. As more duplicated material is added to a pericentromeric or subtelomeric region it becomes a mosaic of duplicated segments. At a later time in evolution, a block of this mosaic sequence is transposed, spreading duplicated segments to non-homologous chromosomes. Subsequent deletion and inversion events may alter the organization of each specific pericentromeric region (8). In this fashion, the pericentromeric regions of multiple chromosomes became populated with partial gene duplications.

ACTIVE REGIONS OF PERICENTROMERIC DUPLICATION

By analyzing the entire working draft genome sequence, it appears that there are at least two types of pericentromeric regions: those that were 'plastic' and accepted recent duplicated sequences from non-homologous chromosomes and those that were 'non-plastic' (Fig. 1). Recently, we performed a genome-wide scan for paralogous segments based on the most recent published version of the human genome (5). We examined all duplications >1 kb in length and in excess of 90% sequence identity. By these criteria, we found 3.62% of the genome to be involved in duplication events, and this could be broken into overlapping duplication classes: 1.77% were duplicated interchromosomally, whereas 2.29% were duplicated intrachromosomally (5). Overall, this analysis showed a 6-fold enrichment of duplication events in pericentromeric and subtelomeric regions across the genome. The pericentromeric regions participating in interchromosomal duplications include: 1q, 2p, 2q, 5q, 7p, 7q, 9p, 9q, 10p, 10q, 11p, 11q, 12q, 13q, 14q, 15q, 16p, 16q, 17p, 17q, 18p, 21q, 22q, Yp and Yq (Fig. 1) (5). Of the 48 pericentromeric regions, 54% (26/48) have been shown to participate in interchromosomal duplications both by *in silico* and FISH analysis (5,46) (Fig. 1). As the Human Genome Project enters its final 2 year completion phase and specialized efforts target the pericentromeric gap regions of the human genome (47), it is likely that the number of chromosomes showing patterns of pericentromeric duplication will increase.

QUIESCENT REGIONS OF PERICENTROMERIC DUPLICATION

Detailed mapping and sequencing of certain pericentromeric regions clearly indicates that a mosaic pattern of interchromosomal duplication is not a universal property of all human chromosomes. The well-characterized pericentromeric

region of the X chromosome shows no evidence of pericentromeric duplications to or from other chromosomes (48). Furthermore, the pericentromeric region of chromosome 19 has been extensively mapped (49) and contains numerous β satellite sequences and multiple ZNF genes, but none of these have recently been duplicated to non-homologous chromosomes (50). Although intrachromosomal duplications are present within the pericentromeric region of 19p12, these duplications are relatively ancient, dating to a time before the emergence of the anthropoids (>40 million years ago) (50,51). Similarly, efforts to map the euchromatic/heterochromatic transition region for chromosome 5 have provided no evidence for recent pericentromeric interchromosomal duplications (52). Combining both *in silico* and comparative FISH data from the Human Genome Project (3,5,46) identifies 10 pericentromeric regions that are negative by both measures for interchromosomal duplications (3p, 4q, 5p, 6q, 8p, 8q, 19p, 19q, Xp and Xq). Final verification of their quiescent duplicative nature will require detailed sequence analyses of the transition regions.

COHORTS OF PERICENTROMERIC DUPLICATION

FISH analysis of randomly selected BACs from the RPCI-11 library showed that 5.4% of 1243 clones hybridized to more than one chromosome (46). More than half of these multi-site clones showed signals to multiple pericentromeric locations within the genome. The distribution of shared sites among pericentromeric regions was not random. These data, combined with sequence analysis, suggested that certain chromosomes have transposed sequence preferentially, defining subsets or cohorts of chromosomes that participate in pericentromeric duplication. For example, 2p11, 10q11, 16p11 and 22q11 define a subset of chromosomes that frequently cross-hybridize by FISH, a result that is verified by >97% sequence identity over hundreds of kilobases of sequence (8,10). Similarly, analysis of the pattern of shared pericentromeric signals showed that chromosomes 7, 10 and Y often had co-hybridizing signals. As would be expected, acrocentric chromosomes 13, 14, 15, 21 and 22 (46) define a cohort based on the high degree of sequence identity of rRNA DNA and shared α and β satellite DNA. Unexpectedly, however, the acrocentric chromosomes also share homology with the pericentromeric regions of 3p and 4q, a property which, in part, is due to the presence of duplicated genomic segments including transmembrane phosphatase with tensin homology (TPTE) and chAB4 related duplicons (30,53–60). The molecular or genetic mechanisms responsible for forming these particular non-homologous associations are not known. It is possible that the proximity of specific chromosomal territories in germline precursor cells, as is seen for chromosomes 9 and 22 in mitotic cells, may increase the probability of particular exchange events prior to meiosis (61).

EVOLUTIONARY TIMING OF DUPLICATIONS

Duplicated segments such as ALD, the 4q24 segment, CTR and regions of 21q are found only in human, chimpanzee and gorilla but not orang-utan, suggesting the pericentromeric copies emerged <12 million years ago (11,13,14,62,63). In contrast, FISH analysis of KGF duplicons and PCR amplification of NF1 segments indicate that these were duplicated before the divergence of human and orang-utan (16,17). Based on sequence similarity, all of these paralogous segments are >90% identical, further supporting their recent emergence. Not surprisingly, both quantitative and qualitative differences in the organization of these regions among man and the great apes have been observed (11,13,19). This suggests that pericentromeric regions are evolutionarily malleable and are able to diverge rapidly (9).

IMPLICATIONS

Genomic instability

The presence of duplicated sequences can be costly for individuals of a species because large blocks of highly similar sequence can cause genomic instability, via mis-alignment of paralogous segments and unequal crossing-over during meiosis (64–66). In the last few years, intrachromosomal duplications (also known as REPs or LCR, low copy repeat sequences) have been implicated in many recurrent chromosomal rearrangements associated with microdeletion and microduplication syndromes (65,67–73). It is noteworthy that many of the same regions associated with intrachromosomal rearrangement and disease have been active for pericentromeric interchromosomal duplication events (74). Sequence analysis of both 10q11 and 22q11 reveal that regions of intrachromosomal duplication are generally located distal to the zones of pericentromeric duplication. As more of the genome has been sequenced, considerable overlap between these two categories of recent duplication has begun to emerge (3,5).

Is it possible that pericentromeric duplications may underlie other forms of large-scale structural variation in the genome? Sites of spontaneous interstitial duplication associated with congenital malformations were recently reviewed by Brewer *et al.* (75). They found that half of all chromosomal regions involved in these duplications occurred within pericentromeric regions. Similarly, a study of translocation breakpoints examined in solid tumor cell lines found that rearrangements most often (60% of all tumors) occurred within pericentromeric DNA (76). In addition, pericentric inversions of chromosome 9 are one of the most common karyotype variations observed in the human population (77). FISH analysis of proximal chromosome 9 clones indicated signals at both 9p12 and 9q13, suggesting that highly identical sequences reside at both locations (78). The second most common inversion event diagnosed cytogenetically in humans results from pericentromeric inversions of chromosome 2, specifically at 2p11 and 2q13 (77). Analysis of chromosome 2 sequences indicated a paralogous segment (LIS2) located at both 2p11 and 2q13. This paralogous segment has recently been shown to be part of much larger duplication with ~98.6% sequence identity (79). Such large blocks of paralogous sequence bracketing the centromere may provide the necessary substrate for recurrent pericentric inversion events. In short, human pericentromeric regions exhibit genomic instability at several different levels. These same regions are enriched for duplicated segments. Although cause and consequence can not be distinguished, the available data suggest that these two processes (duplication and genomic instability) are closely intertwined and contribute significantly to disease etiology and genomic structural variation.

Evolutionary innovation

Ohno *et al.* (80) first postulated the importance of gene duplication in 1968 as the main driving force of evolution. Once a gene was duplicated, one copy was no longer constrained by selection and any mutations that occurred in the duplicate copy could potentially lead to new expression patterns or altered function, leaving the original copy to provide its required function (81). Over the past decade, numerous genes have been identified that exist in multiple copies in the human genome: among them are the Hox (82), T-cell receptor (83), globin (84) and Sm/Lg rRNA genes (85). In contrast to these ancient duplications, the pericentromerically duplicated segments reviewed here arose relatively recently in evolutionary time (11,13,14,16,17). In a few cases, polymorphic variation has been observed within the human population (43,86), suggesting that the duplication process may be ongoing. Assuming this process occurred over longer evolutionary periods of time, what could be its impact?

A surprising finding of the recent genome sequence papers was that humans had fewer genes (3) than had been predicted previously (87,88). The estimated number of genes (~35 000) represented only a 2-fold increase over that of the fruit fly and worm (3). While this result was disconcerting to some, when analyzed in more detail it was found that human genes/proteins typically contain multiple domains whereas fruit fly and worm genes/proteins are generally half as complex, containing fewer mixed domains (3). Humans also have significantly more segmental duplications than flies and worms, a nearly 10-fold increase in the fraction of segmental duplications (3). While pericentromeric and subtelomeric duplications can be costly to an organism by increasing the opportunity for genomic instability, they provide an evolutionary means to mix domains of previously unrelated genes together in novel combinations, a vehicle for exon shuffling. If these novel combinations maintain or acquire transcriptional potency, it could result in novel gene innovations. Several examples of such chimeric transcripts/genes arising from pericentromeric duplications have recently been described (35,89). In this light, pericentromeric regions could be viewed as workshops of evolutionary invention, an assemblage of chromosomal pieces that is continually expanded, churned and purged (74). Alternatively, chromosomal fusion events (as is the case for 2q21) might allow pericentromeric DNA to be immediately 'euchromatized,' providing an opportunity for the emergence of new fusion transcripts. Although successful gene innovation by this process is probably a rare occurrence, the sheer number of such juxtapositions and duplications in the last 15 million years suggests that pericentromeric duplications have had both structural and functional import in the evolution of our genome.

ACKNOWLEDGEMENTS

The authors are supported by grants from NIGMS, NHGRI, DOE and the March of Dimes. J.E.H. was supported in part by NIH GM08613, Genetics Training grant. D.P.L. was supported in part by NIH HD07518, Medical Genetics Training grant.

REFERENCES

1. Dunham, I., Shimizu, N., Roe, B.A., Chissoe, S., Hunt, A.R., Collins, J.E., Bruskiewich, R., Beare, D.M., Clamp, M., Smink, L.J. *et al.* (1999) The DNA sequence of human chromosome 22. *Nature*, **402**, 489–495.
2. Hattori, M., Fujiyama, A., Taylor, T.D., Watanabe, H., Yada, T., Park, H.S., Toyoda, A., Ishii, K., Totoki, Y., Choi, D.K. *et al.* (2000) The DNA sequence of human chromosome 21. *Nature*, **405**, 311–339.
3. The International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
4. Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A. *et al.* (2001) The sequence of the human genome. *Science*, **291**, 1304–1351.
5. Bailey, J.A., Yavor, A.M., Massa, H.F., Trask, B.J. and Eichler, E.E. (2001) Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res.*, **11**, 1005–1017.
6. Caron, H., van Schaik, B., van der Mee, M., Baas, F., Riggins, G., van Sluis, P., Hermus, M.C., van Asperen, R., Boon, K., Voute, P.A. *et al.* (2001) The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science*, **291**, 1289–1292.
7. Eichler, E.E. (1998) Masquerading repeats: paralogous pitfalls of the human genome. [Published erratum appears in *Genome Res.* (1998) **8**, 1095]. *Genome Res.*, **8**, 758–762.
8. Horvath, J.E., Schwartz, S. and Eichler, E.E. (2000) The mosaic structure of human pericentromeric DNA: a strategy for characterizing complex regions of the human genome. *Genome Res.*, **10**, 839–852.
9. Jackson, M.S., Rocchi, M., Thompson, G., Hearn, T., Crosier, M., Guy, J., Kirk, D., Mulligan, L., Ricco, A., Piccininni, S. *et al.* (1999) Sequences flanking the centromere of human chromosome 10 are a complex patchwork of arm-specific sequences, stable duplications and unstable sequences with homologies to telomeric and other centromeric locations. *Hum. Mol. Genet.*, **8**, 205–215.
10. Guy, J., Spalluto, C., McMurray, A., Hearn, T., Crosier, M., Viggiano, L., Miolla, V., Archidiacono, N., Rocchi, M., Scott, C. *et al.* (2000) Genomic sequence and transcriptional profile of the boundary between pericentromeric satellites and genes on human chromosome arm 10q. *Hum. Mol. Genet.*, **9**, 2029–2042.
11. Horvath, J.E., Viggiano, L., Loftus, B.J., Adams, M.D., Archidiacono, N., Rocchi, M. and Eichler, E.E. (2000) Molecular structure and evolution of an α satellite/non-α satellite junction at 16p11. *Hum. Mol. Genet.*, **9**, 113–123.
12. Loftus, B.J., Kim, U.J., Sneddon, V.P., Kalush, F., Brandon, R., Fuhrmann, J., Mason, T., Crosby, M.L., Barnstead, M., Cronin, L. *et al.* (1999) Genome duplications and other features in 12 Mb of DNA sequence from human chromosome 16p and 16q. *Genomics*, **60**, 295–308.
13. Eichler, E.E., Lu, F., Shen, Y., Antonacci, R., Jurecic, V., Doggett, N.A., Moyzis, R.K., Baldini, A., Gibbs, R.A. and Nelson, D.L. (1996) Duplication of a gene-rich cluster between 16p11.1 and Xq28: a novel pericentromeric-directed mechanism for paralogous genome evolution. *Hum. Mol. Genet.*, **5**, 899–912.
14. Eichler, E.E., Budarf, M.L., Rocchi, M., Deaven, L.L., Doggett, N.A., Baldini, A., Nelson, D.L. and Mohrenweiser, H.W. (1997) Interchromosomal duplications of the adrenoleukodystrophy locus: a phenomenon of pericentromeric plasticity. *Hum. Mol. Genet.*, **6**, 991–1002.
15. Eichler, E.E., Archidiacono, N. and Rocchi, M. (1999) CAGGG repeats and the pericentromeric duplication of the hominoid genome. *Genome Res.*, **9**, 1048–1058.
16. Regnier, V., Meddeb, M., Lecointre, G., Richard, F., Duverger, A., Nguyen, V.C., Dutrillaux, B., Bernheim, A. and Danglot, G. (1997) Emergence and scattering of multiple neurofibromatosis (NF1)-related sequences during hominoid evolution suggest a process of pericentromeric interchromosomal transposition. *Hum. Mol. Genet.*, **6**, 9–16.
17. Zimonjic, D.B., Kelley, M.J., Rubin, J.S., Aaronson, S.A. and Popescu, N.C. (1997) Fluorescence *in situ* hybridization analysis of keratinocyte growth factor gene amplification and dispersion in evolution of great apes and humans. *Proc. Natl Acad. Sci. USA*, **94**, 11461–11465.
18. Childs, G., Maxson, R., Cohn, R.H. and Kedes, L. (1981) Orphons: dispersed genetic elements derived from tandem repetitive genes of eucaryotes. *Cell*, **23**, 651–663.
19. Arnold, N., Wienberg, J., Ermert, K. and Zachau, H.G. (1995) Comparative mapping of DNA probes derived from the V κ immunoglobulin gene regions on human and great ape chromosomes by fluorescence *in situ* hybridization. *Genomics*, **26**, 147–150.
20. Buiting, K., Greger, V., Brownstein, B.H., Mohr, R.M., Voiculescu, I., Winterpacht, A., Zabel, B. and Horsthemke, B. (1992) A putative gene

family in 15q11–13 and 16p11.2: possible implications for Prader–Willi and Angelman syndromes. *Proc. Natl Acad. Sci. USA*, **89**, 5457–5461.
21. Frippiat, J.P., Dard, P., Marsh, S., Winter, G. and Lefranc, M.P. (1997) Immunoglobulin λ light chain orphons on human chromosome 8q11.2. *Eur. J. Immunol.*, **27**, 1260–1265.
22. Mancuso, D.J., Tuley, E.A., Westfield, L.A., Lester-Mancuso, T.L., Le Beau, M.M., Sorace, J.M. and Sadler, J.E. (1991) Human von Willebrand factor gene and pseudogene: structural analysis and differentiation by polymerase chain reaction. *Biochemistry*, **30**, 253–269.
23. Frank, S.L., Klisak, I., Sparkes, R.S. and Lusis, A.J. (1989) A gene homologous to plasminogen located on human chromosome 2q11-p11. *Genomics*, **4**, 449–451.
24. van Deutekom, J.C., Lemmers, R.J., Grewal, P.K., van Geel, M., Romberg, S., Dauwerse, H.G., Wright, T.J., Padberg, G.W., Hofker, M.H., Hewitt, J.E. *et al*. (1996) Identification of the first gene (FRG1) from the FSHD region on human chromosome 4q35. *Hum. Mol. Genet.*, **5**, 581–590.
25. Collins, J.E., Mungall, A.J., Badcock, K.L., Fay, J.M. and Dunham, I. (1997) The organization of the γ-glutamyl transferase genes and other low copy repeats in human chromosome 22q11. *Genome Res.*, **7**, 522–531.
26. Tomlinson, I.M., Cook, G.P., Carter, N.P., Elaswarapu, R., Smith, S., Walter, G., Buluwela, L., Rabbitts, T.H. and Winter, G. (1994) Human immunoglobulin VH and D segments on chromosomes 15q11.2 and 16p11.2. *Hum. Mol. Genet.*, **3**, 853–860.
27. Nagaoka, H., Ozawa, K., Matsuda, F., Hayashida, H., Matsumura, R., Haino, M., Shin, E.K., Fukita, Y., Imai, T., Anand, R. *et al*. (1994) Recent translocation of variable and diversity segments of the human immunoglobulin heavy chain from chromosome 14 to chromosomes 15 and 16. *Genomics*, **22**, 189–197.
28. Borden, P., Jaenichen, R. and Zachau, H.G. (1990) Structural features of transposed human VK genes and implications for the mechanism of their transpositions. *Nucleic Acids Res.*, **18**, 2101–2107.
29. Ruault, M., Trichet, V., Gimenez, S., Boyle, S., Gardiner, K., Rolland, M., Roizes, G. and De Sario, A. (1999) Juxta-centromeric region of human chromosome 21 is enriched for pseudogenes and gene fragments. *Gene*, **239**, 55–64.
30. Chen, H., Rossier, C., Morris, M.A., Scott, H.S., Gos, A., Bairoch, A. and Antonarakis, S.E. (1999) A testis-specific gene, TPTE, encodes a putative transmembrane tyrosine phosphatase and maps to the pericentromeric region of human chromosomes 21 and 13, and to chromosomes 15, 22, and Y. *Hum. Genet.*, **105**, 399–409.
31. Kehrer-Sawatzki, H., Schwickardt, T., Assum, G., Rocchi, M. and Krone, W. (1997) A third neurofibromatosis type 1 (NF1) pseudogene at chromosome 15q11.2. *Hum. Genet.*, **100**, 595–600.
32. Luijten, M., Wang, Y., Smith, B.T., Westerveld, A., Smink, L.J., Dunham, I., Roe, B.A. and Hulsebos, T.J. (2000) Mechanism of spreading of the highly related neurofibromatosis type 1 (NF1) pseudogenes on chromosomes 2, 14 and 22. *Eur. J. Hum. Genet.*, **8**, 209–214.
33. Iyer, G.S., Krahe, R., Goodwin, L.A., Doggett, N.A., Siciliano, M.J., Funanage, V.L. and Proujansky, R. (1996) Identification of a testis-expressed creatine transporter gene at 16p11.2 and confirmation of the X-linked locus to Xq28. *Genomics*, **34**, 143–146.
34. Ji, Y., Rebert, N.A., Joslin, J.M., Higgins, M.J., Schultz, R.A. and Nicholls, R.D. (2000) Structure of the highly conserved HERC2 gene and of multiple partially duplicated paralogs in human. *Genome Res.*, **10**, 319–329.
35. Footz, T.K., Brinkman-Mills, P., Banting, G.S., Maier, S.A., Riazi, M.A., Bridgland, L., Hu, S., Birren, B., Minoshima, S., Shimizu, N. *et al*. (2001) Analysis of the Cat Eye syndrome critical region in humans and the region of conserved synteny in mice: a search for candidate genes at or near the human chromosome 22 pericentromere. *Genome Res.*, **11**, 1053–1070.
36. Vogt, P. (1990) Potential genetic functions of tandem repeated DNA sequence blocks in the human genome are based on a highly conserved 'chromatin folding code.' *Hum. Genet.*, **84**, 301–336.
37. Haaf, T. and Bray-Ward, P. (1996) Region-specific YAC banding and painting probes for comparative genome mapping: implications for the evolution of human chromosome 2. *Chromosoma*, **104**, 537–544.
38. Avarello, R., Pedicini, A., Caiulo, A., Zuffardi, O. and Fraccaro, M. (1992) Evidence for an ancestral alphoid domain on the long arm of human chromosome 2. *Hum. Genet.*, **89**, 247–249.
39. Ijdo, J.W., Baldini, A., Ward, D.C., Reeders, S.T. and Wells, R.A. (1991) Origin of human chromosome 2: an ancestral telomere-telomere fusion. *Proc. Natl Acad. Sci. USA*, **88**, 9051–9055.
40. Luke, S. and Verma, R.S. (1992) Origin of human chromosome 2. *Nat. Genet.*, **2**, 11–12.
41. Baldini, A., Ried, T., Shridhar, V., Ogura, K., D'Aiuto, L., Rocchi, M. and Ward, D.C. (1993) An alphoid DNA sequence conserved in all human and great ape chromosomes: evidence for ancient centromeric sequences at human chromosomal regions 2q21 and 9q13. *Hum. Genet.*, **90**, 577–583.
42. Dempsey, L.A., Sun, H., Hanakahi, L.A. and Maizels, N. (1999) G4 DNA binding by LR1 and its subunits, nucleolin and hnRNP D, A role for G-G pairing in immunoglobulin switch recombination. *J. Biol. Chem.*, **274**, 1066–1071.
43. Trask, B.J., Friedman, C., Martin-Gallardo, A., Rowen, L., Akinbami, C., Blankenship, J., Collins, C., Giorgi, D., Iadonato, S., Johnson, F. *et al*. (1998) Members of the olfactory receptor gene family are contained in large blocks of DNA duplicated polymorphically near the ends of human chromosomes. *Hum. Mol. Genet.*, **7**, 13–26.
44. Riethman, H.C., Xiang, Z., Paul, S., Morse, E., Hu, X.L., Flint, J., Chi, H.C., Grady, D.L. and Moyzis, R.K. (2001) Integration of telomere sequences with the draft human genome sequence. *Nature*, **409**, 948–951.
45. Sen, D. and Gilbert, W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, **334**, 364–366.
46. Cheung, V.G., Nowak, N., Jang, W., Kirsch, I.R., Zhao, S., Chen, X.N., Furey, T.S., Kim, U.J., Kuo, W.L., Olivier, M. *et al*. (2001) Integration of cytogenetic landmarks into the draft sequence of the human genome. *Nature*, **409**, 953–958.
47. Eichler, E.E. (2001) Segmental duplications: what's missing, misassigned, and misassembled—and should we care? *Genome Res.*, **11**, 653–656.
48. Schueler, M. and Willard, H.F. (2001) *Content and Evolution of the Centromere/Non-centromere Border of the Human X Chromosome*. Genome Sequencing and Biology meeting, Cold Spring Harbor, NY.
49. Ashworth, L.K., Batzer, M.A., Brandriff, B., Branscomb, E., de Jong, P., Garcia, E., Garnes, J.A., Gordon, L.A., Lamerdin, J.E., Lennon, G. *et al*. (1995) An integrated metric physical map of human chromosome 19. *Nat. Genet.*, **11**, 422–427.
50. Eichler, E.E., Hoffman, S.M., Adamson, A.A., Gordon, L.A., McCready, P., Lamerdin, J.E. and Mohrenweiser, H.W. (1998) Complex β-satellite repeat structures and the expansion of the zinc finger gene cluster in 19p12. *Genome Res.*, **8**, 791–808.
51. Bellefroid, E.J., Marine, J.C., Matera, A.G., Bourguignon, C., Desai, T., Healy, K.C., Bray-Ward, P., Martial, J.A., Ihle, J.N. and Ward, D.C. (1995) Emergence of the ZNF91 Kruppel-associated box-containing zinc finger gene family in the last common ancestor of anthropoidea. *Proc. Natl Acad. Sci. USA*, **92**, 10757–10761.
52. Puechberty, J., Laurent, A.M., Gimenez, S., Billault, A., Brun-Laurent, M.E., Calenda, A., Marcais, B., Prades, C., Ioannou, P., Yurov, Y. *et al*. (1999) Genetic and physical analyses of the centromeric and pericentromeric regions of human chromosome 5: recombination across 5cen. *Genomics*, **56**, 274–287.
53. Greig, G.M., Warburton, P.E. and Willard, H.F. (1993) Organization and evolution of an α satellite DNA subset shared by human chromosomes 13 and 21. *J. Mol. Evol.*, **37**, 464–475.
54. Jorgensen, A.L., Kolvraa, S., Jones, C. and Bak, A.L. (1988) A subfamily of alphoid repetitive DNA shared by the NOR-bearing human chromosomes 14 and 22. *Genomics*, **3**, 100–109.
55. Choo, K.H., Vissel, B., Brown, R., Filby, R.G. and Earle, E. (1988) Homologous α satellite sequences on human acrocentric chromosomes with selectivity for chromosomes 13, 14 and 21: implications for recombination between nonhomologues and Robertsonian translocations. *Nucleic Acids Res.*, **16**, 1273–1284.
56. Vissel, B. and Choo, K.H. (1991) Four distinct α satellite subfamilies shared by human chromosomes 13, 14 and 21. *Nucleic Acids Res.*, **19**, 271–277.
57. Vissel, B. and Choo, K.H. (1992) Evolutionary relationships of multiple α satellite subfamilies in the centromeres of human chromosomes 13, 14, and 21. *J. Mol. Evol.*, **35**, 137–146.
58. Choo, K.H., Vissel, B. and Earle, E. (1989) Evolution of α-satellite DNA on human acrocentric chromosomes. *Genomics*, **5**, 332–344.
59. Wohr, G., Fink, T. and Assum, G. (1996) A palindromic structure in the pericentromeric region of various human chromosomes. *Genome Res.*, **6**, 267–279.
60. Assum, G., Fink, T., Klett, C., Lengl, B., Schanbacher, M., Uhl, S. and Wohr, G. (1991) A new multisequence family in human. *Genomics*, **11**, 397–409.
61. Kozubek, S., Lukasova, E., Mareckova, A., Skalnikova, M., Kozubek, M., Bartova, E., Kroha, V., Krahulcova, E. and Slotova, J. (1999) The

topological organization of chromosomes 9 and 22 in cell nuclei has a determinative role in the induction of t(9,22) translocations and in the pathogenesis of t(9,22) leukemias. *Chromosoma*, **108**, 426–435.
62. Orti, R., Potier, M.C., Maunoury, C., Prieur, M., Creau, N. and Delabar, J.M. (1998) Conservation of pericentromeric duplications of a 200-kb part of the human 21q22.1 region in primates. *Cytogenet. Cell Genet.*, **83**, 262–265.
63. Goodman, M. (1999) The genomic record of Humankind's evolutionary roots. *Am. J. Hum. Genet.*, **64**, 31–39.
64. Lupski, J.R. (1998) Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet.*, **14**, 417–422.
65. Ji, Y., Eichler, E.E., Schwartz, S. and Nicholls, R.D. (2000) Structure of chromosomal duplicons and their role in mediating human genomic disorders. *Genome Res.*, **10**, 597–610.
66. Mazzarella, R. and Schlessinger, D. (1997) Duplication and distribution of repetitive elements and non-unique regions in the human genome. *Gene*, **205**, 29–38.
67. Amos-Landgraf, J.M., Ji, Y., Gottlieb, W., Depinet, T., Wandstrat, A.E., Cassidy, S.B., Driscoll, D.J., Rogan, P.K., Schwartz, S. and Nicholls, R.D. (1999) Chromosome breakage in the Prader–Willi and Angelman syndromes involves recombination between large, transcribed repeats at proximal and distal breakpoints. *Am. J. Hum. Genet.*, **65**, 370–386.
68. Chen, K.S., Manian, P., Koeuth, T., Potocki, L., Zhao, Q., Chinault, A.C., Lee, C.C. and Lupski, J.R. (1997) Homologous recombination of a flanking repeat gene cluster is a mechanism for a common contiguous gene deletion syndrome. *Nat. Genet.*, **17**, 154–163.
69. Reiter, L.T., Murakami, T., Koeuth, T., Gibbs, R.A. and Lupski, J.R. (1997) The human COX10 gene is disrupted during homologous recombination between the 24 kb proximal and distal CMT1A-REPs. *Hum. Mol. Genet.*, **6**, 1595–1603.
70. Christian, S.L., Fantes, J.A., Mewborn, S.K., Huang, B. and Ledbetter, D.H. (1999) Large genomic duplicons map to sites of instability in the Prader–Willi/Angelman syndrome chromosome region (15q11–q13). *Hum. Mol. Genet.*, **8**, 1025–1037.
71. Edelmann, L., Pandita, R.K., Spiteri, E., Funke, B., Goldberg, R., Palanisamy, N., Chaganti, R.S., Magenis, E., Shprintzen, R.J. and Morrow, B.E. (1999) A common molecular basis for rearrangement disorders on chromosome 22q11. *Hum. Mol. Genet.*, **8**, 1157–1167.
72. Shaikh, T.H., Kurahashi, H., Saitta, S.C., O'Hare, A.M., Hu, P., Roe, B.A., Driscoll, D.A., McDonald-McGinn, D.M., Zackai, E.H., Budarf, M.L. *et al.* (2000) Chromosome 22-specific low copy repeats and the 22q11.2 deletion syndrome: genomic organization and deletion endpoint analysis. *Hum. Mol. Genet.*, **9**, 489–501.
73. Potocki, L., Chen, K.S., Park, S.S., Osterholm, D.E., Withers, M.A., Kimonis, V., Summers, A.M., Meschino, W.S., Anyane-Yeboa, K., Kashork, C.D. *et al.* (2000) Molecular mechanism for duplication 17p11.2-the homologous recombination reciprocal of the Smith–Magenis microdeletion. *Nat. Genet.*, **24**, 84–87.
74. Eichler, E.E. (1999) Repetitive conundrums of centromere structure and function. *Hum. Mol. Genet.*, **8**, 151–155.
75. Brewer, C., Holloway, S., Zawalnyski, P., Schinzel, A. and FitzPatrick, D. (1999) A chromosomal duplication map of malformations: regions of suspected haplo- and triplolethality—and tolerance of segmental aneuploidy—in humans. *Am. J. Hum. Genet.*, **64**, 1702–1708.
76. Padilla-Nash, H.M., Heselmeyer-Haddad, K., Wangsa, D., Zhang, H., Ghadimi, B.M., Macville, M., Augustus, M., Schrock, E., Hilgenfeld, E. and Ried, T. (2001) Jumping translocations are common in solid tumor cell lines and result in recurrent fusions of whole chromosome arms. *Genes Chromosomes Cancer*, **30**, 349–363.
77. Kaiser, P. (1984) Pericentric inversions. Problems and significance for clinical genetics. *Hum. Genet.*, **68**, 1–47.
78. Park, J.P., Wojiski, S.A., Spellman, R.A., Rhodes, C.H. and Mohandas, T.K. (1998) Human chromosome 9 pericentric homologies: implications for chromosome 9 heteromorphisms. *Cytogenet. Cell Genet.*, **82**, 192–194.
79. Fogli, A., Giglio, S., Arrigo, G., Lo Nigro, C., Zollo, M., Viggiano, L., Rocchi, M., Archidiacono, N., Zuffardi, O. and Carrozzo, R. (1999) Identification of two paralogous regions mapping to the short and long arms of human chromosome 2 comprising LIS1 pseudogenes. *Cytogenet. Cell Genet.*, **86**, 225–232.
80. Ohno, S., Wolf, U. and Atkin, N.B. (1968) Evolution from fish to mammals by gene duplication. *Hereditas*, **59**, 169–187.
81. Ohno, S. (1970) *Evolution by Gene Duplication*. Springer-Verlag, New York, NY.
82. Krumlauf, R. (1994) Hox genes in vertebrate development. *Cell*, **78**, 191–201.
83. Hood, L., Kronenberg, M. and Hunkapiller, T. (1985) T cell antigen receptors and the immunoglobulin supergene family. *Cell*, **40**, 225–229.
84. Orkin, S.H. and Kazazian, H.H.,Jr (1984) The mutation and polymorphism of the human β-globin gene and its surrounding DNA. *Annu. Rev. Genet.*, **18**, 131–171.
85. Srivastava, A.K. and Schlessinger, D. (1991) Structure and organization of ribosomal DNA. *Biochimie*, **73**, 631–638.
86. Ritchie, R.J., Mattei, M.G. and Lalande, M. (1998) A large polymorphic repeat in the pericentromeric region of human chromosome 15q contains three partial gene duplications. *Hum. Mol. Genet.*, **7**, 1253–1260.
87. Liang, F., Holt, I., Pertea, G., Karamycheva, S., Salzberg, S.L. and Quackenbush, J. (2000) Gene index analysis of the human genome estimates approximately 120 000 genes. *Nat. Genet.*, **25**, 239–240.
88. Fields, C., Adams, M.D., White, O. and Venter, J.C. (1994) How many genes in the human genome? *Nat. Genet.*, **7**, 345–346.
89. Inoue, K., Dewar, K., Katsanis, N., Reiter, L.T., Lander, E.S., Devon, K.L., Wyman, D.W., Lupski, J.R. and Birren, B. (2001) The 1.4-Mb CMT1A duplication/HNPP deletion genomic region reveals unique genome architectural features and provides insights into the recent evolution of new genes. *Genome Res.*, **11**, 1018–1033.