Human Molecular Genetics, 1999, Vol. 8, No. 2 **151–155**

COMMENTARY

Repetitive conundrums of centromere structure and function

Evan E. Eichler*

Department of Genetics and Center for Human Genetics, Case Western Reserve School of Medicine and University Hospitals of Cleveland, Cleveland, OH 44106, USA

Received November 24, 1998; Accepted November 24, 1998

In the last few years, a paradox has emerged regarding the relationship of centromere structure and its function. Most centromeric DNAs analyzed to date are composed of a remarkably complex array of repeat structures. In contrast, recent analyses of neocentromeric DNA reveal that repetitive DNA is not a prerequisite for centromere activity. The ubiquity of repetitive sequences among diverse species at sites of primary constriction argues that there is a strong evolutionary link between centromere structure and function. Dynamic mutational processes resulting in amplification, deletion and transposition of repetitive sequences appear to occur frequently in such regions, resulting in considerable interspecific diversity in structure and sequence. One possible solution to this conundrum may be that the rapid accumulation of repetitive sequences within centromeric and pericentromeric DNA is a consequence of functionally active centromeres. Emerging repetitive structures at centromeric sites may be an important byproduct of a functional centromere which ensures that site as an evolutionarily favored position in subsequent meiotic and mitotic lineages. The recent identification of large gene duplications in the vicinity of centromeres may be another example of the enhanced mutational lability of such regions of the genome.

INTRODUCTION

The association between repetitive DNA and centromere structure is long-standing. With the exception of point centromeres among species of budding yeast and the enigmatic holocentric chromosomes of *Caenorhabditis elegans* (1), almost every other species of plant, animal and fungus studied to date harbors a spectacular array of both highly repetitive and middle repetitive elements at the site of each chromosome's primary constriction (2). Although the specific details of the DNA structure vary among the diverse species, the molecular architecture is generally the same: blocks of AT-rich tandemly repeated DNA bracketed by clusters of various classes of retroposons (3–6). The recurrent evolutionary theme of repetitive DNA has been taken as evidence (albeit not as a proof) of an implied functional relationship (7). The presence of satellite DNA and other repetitive DNA in these regions of a chromosome are thought to favor the assembly of the kinetochore, thereby ensuring the efficient and timely meiotic and mitotic segregation of chromosomes (8–10). Several recent experiments in both humans and *Drosophila* provide continued support for this model (5,11,12). The repetitive structure of naturally occurring centromeres within the context of mini/microchromosomes, for example, appears to be sufficient to recapitulate, if not all, at least some aspects of centromere function including kinetochore formation and chromosome segregation. Nevertheless, the observation of functional inactivation of one centromere among some stable human dicentric chromosomes (13–15) and the capacity of portions of chromosomes which do not normally carry repetitive centromeric DNA to function as sites for kinetochore assembly (16–18) provide ample evidence that repetitive DNA, by itself, does not explain adequately the molecular basis of a functionally competent centromere (19).

CENTROMERE FUNCTION: WHERE ARE THE REPEATS?

Two articles in this issue further complicate the relationship between centromere structure and function (20,21). Both articles examine in detail the molecular structure of two functionally competent centromeres on human chromosome 10. The structural dichotomy between these two structures could not have been more striking (Fig. 1). Jackson *et al*. (20) focus on the large-scale organization of ~8 Mb of centromere and flanking pericentromeric DNA found on chromosome 10, while Barry *et al*. (21) present the first complete sequence analysis of neocentromere DNA, isolated from cytogenetic band 10q25.2 (17). What is most remarkable about the 80 kb of neocentromere DNA is that the sequence is completely unremarkable. Detailed sequence analysis reveals the complete absence of classic alpha satellite repeat elements commonly associated with human centromeres (22). Nor do the authors find any evidence of other pericentric sequences including beta satellite, gamma satellite, AT-rich

*Tel: +1 216 368 4883; Fax: +1 216 368 3432; Email: eee@po.cwru.edu

sequence (ATRS) or centomere protein B (CENPB) box-binding sequence motifs. The sequence is not particularly enriched for retroposons (Fig. 1a) nor is it, taken as a whole, remarkably AT-rich. Although sporadic occurrences of expressed sequence tags (ESTs) and short sequence motifs (such as Pjalpha, satellites I and II) were identified, their distribution and frequency do not differ significantly from expected values for 'generic' genomic DNA. Furthermore, there are virtually no differences between the sequence compared between the marker, mardel(10), chromosome and the original parental DNA. The absence of a 'magic' repeat sequence undetected by previous restriction or hybridization analysis dispels the notion that repetitive DNA is absolutely required to form a functionally competent centromere. Assuming that the 80 kb region does in fact contain the neocentromere and that it does not reside, for example, in the immediately flanking regions, unusually repetitive DNA does not appear, at least in this case, to be a prerequisite for centromere formation.

In stark contrast to the mundane neocentromere sequence, the overview of the organization of the normal chromosome 10 centromere and its flanking pericentromeric sequence is astonishingly complex (20). Jackson presents arguably the best developed and well-supported physical maps of any human centromere. Building on previously developed pulsed-field gel electrophoresis (PFGE) and yeast artificial chromosome (YAC) maps in this region (23–25), the authors develop a collection of pericentromeric markers which are used to confirm the orientation of various duplications flanking the alpha satellite centromeric DNA as well as to assess the degree of evolutionary conservation. This analysis indicates an unusually high degree of variability in the organization and content of sequences flanking the alpha satellite and satellite III blocks at or near the chromosome 10 centromere. In addition, the data suggest the presence of uncharacterized recently duplicated segments from a variety of different pericentromeric locations (Fig. 1b). Although final verification of the origin of these segments will require comparative sequence analysis of this region with other mapped genomic sequences, these data support research from other laboratories which have shown that the pericentromeric regions of human chromosomes have been active in the acquisition of duplicated gene segments from elsewhere in the genome (26–37). The apparent evolutionary transience of many of the segments within the pericentromeric regions suggests that many of these putative rearrangements, duplications and inversions are relatively recent events in the human lineage, resulting in striking differences in the chromosome structure among closely related primates such as chimpanzee, gorilla and orangutan. Recent sequence data from the pericentromeric region of chromosome 10 (GenBank accession nos AL022344 and AL022345) confirm that many of the interchromosomal cross-hybridizations reported herein are probably the result of a pericentromeric-directed mechanism of gene duplication.

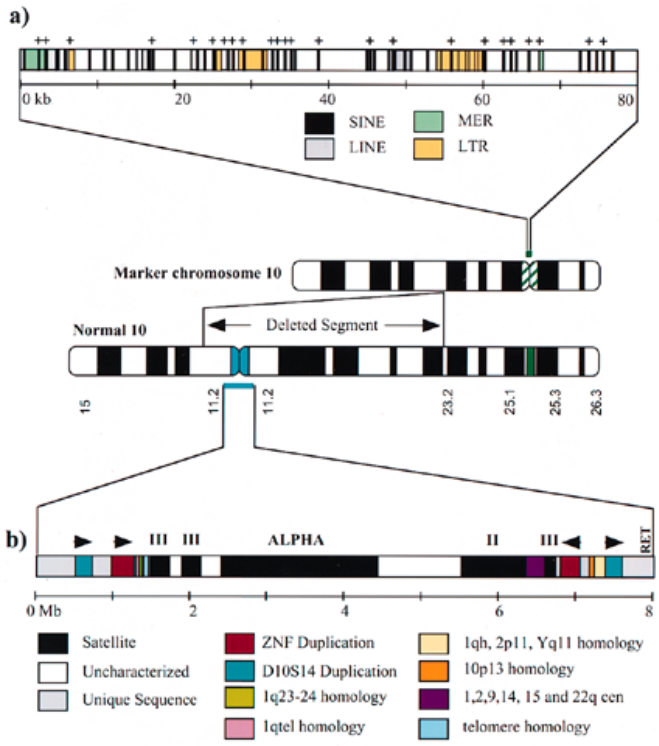

Figure 1. Structure of functionally competent chromosome 10 centromeres. (**a**) Sequence analysis of 80 kb of neocentromere sequence (21). Mardel(10) is formed by breakage and fusion within 10p11.2 and 10q23.2, resulting in the generation of a neocentromere at 10q25.2. A total of 80 kb of sequence (GenBank accession no. AF04284) corresponding to the site of kinetochore formation was analyzed for the presence of repeats using RepeatMasker. The distribution and orientation (+, forward with respect to sequence) of various repeats (SINE, LINE, MER and LTR elements) are depicted. Only repeat segments with length >200 bp were considered. (**b**) Analysis of an ~8 Mb region spanning the normal chromosome 10p11/10q11 centromere (20). Note the presence of various subtelomeric and pericentromeric homologies located distal to the satellite sequence as well as the large duplications of the *ZNF* (zinc finger gene) and *D10S14* region. Sizes of various segments are approximated as the extent of each homology is not precisely known. Approximately 1 Mb of 10p11.2 currently is being sequenced by the Sanger Center and is available within the HTGS division of GenBank (accession nos AL022344 and AL022345). Analysis of this sequence indicates the presence of many interchromosomally duplicated genes and gene segments.

THE FUNCTION OF REPETITIVE STRUCTURE

If these data are generally applicable to other centromeric regions, it would suggest that neocentromeres are less structurally complex than anticipated, while normal centromeres show an unsuspected degree of repetitive complexity which may extend into the megabases of DNA sequence flanking alpha satellite DNA. How can these two very different structures be reconciled with centromere function? One solution to this puzzling conundrum (often used when simple molecular paradigms prove less than failsafe) has been to invoke an epigenetic process to explain the 'marking' of functional centromere DNA (19,38,39). One candidate for this marker, recently proposed by Csink and Henikoff, may be late DNA replication (40). In this resurrected version of Dupraw's model (41), functional competence of a centromere is simply an attribute of the latest replicating segment of DNA. What, then, is the functional relevance of repetitive structure? According to Henikoff, the accumulation of repetitive DNA further retards replication, effectively fixing such sites in the genome to function more competently as a centromere. Therefore, the universality of repetitive DNA at centromeres is an evolutionarily derived character state and not a precursor of centromere function. In other words, centromere function is not a consequence of repetitive structure, but repetitive structure is a consequence of enhancing centromere function. Although this

model could adequately explain phenomena of lateral inhibition among some dicentric chromosomes and the formation of neocentromeres, there are, at this point, very few experimental data in its support. The availability of neocentromere sequence, however, provides a direct opportunity to test the validity of this model. It should be possible, for example, to develop assays to test replication timing of this sequence in both normal as well as neocentromere chromosomes. The finding that this 80 kb of sequence from 10q25.2 is the latest replicating piece of DNA would clearly support the model that a neocentromere is the evolutionary equivalent of a newborn centromere awaiting the maturity conferred by repetitive DNA.

PERICENTROMERIC DUPLICATIONS: COMPLEX DNA BECOMES MORE COMPLICATED

Whatever model is invoked to explain the relationship between centromere structure and function, it is clear that in humans, at least, the model might need to take into account the phenomenon of recent pericentromeric duplication. For many chromosomes, there is now compelling evidence that regions in close proximity to the centromere are hotspots for recent gene duplication events (26–37). If the centromere is a recruitment station for repeats, perhaps it can also serve as a reservoir for the accumulation of transposed genic segments. Interestingly, this peculiar structural property (the wholesale duplication of entire genes or portions of intronic/exonic sequence) has only been reported among primate chromosomes, where it can result in considerable variation in chromosome structure as well as a presumptive proclivity to recurrent cytogenetic rearrangements associated with genetic disease (42). In the absence of detailed maps and sequence analysis from other organisms, it would be premature to conclude that this property is restricted to primate chromosomes. It may be noteworthy that a very similar process of recent gene duplication has now been reported to occur near the subtelomeric boundaries of a variety of human chromosomes (43–46). Is the accumulation of these genic segments a property of 'heterochromatic sequence'? Does their presence in these locations serve a particular function? Disregarding potential evolutionary implications of recent gene duplication and exon shuffling, perhaps the presence of such transposed DNA, harboring multiple protein-binding sites associated with gene function and processing, may function as a gate-keeper, effectively quenching the effects of heterochromatization associated with telomeric and centromeric DNA. Such an effect might serve effectively to 'insulate' bona fide unique genes from heterochromatic position effects due to their close proximity (i.e. *RET* proto-oncogene on 10q11.2) to repetitive sequences. Alternatively, duplicated segments may contribute to the 'epigenetic' basis for centromere formation by playing a role in the formation of the proposed higher-order repeat structures (5,38). In this regard, it is interesting to note that the vast majority of neocentromeres [of which mardel(10) is an exception] are structurally mirror-image chromosomes (39). Perhaps large inverted duplicated sequences somehow aid in specifying the position of a functionally competent centromere. Such inverted structures are not uncommon in pericentromeric DNA (47,48), as evidenced by the large (250 kb) inverted blocks of sequence in 10q11 and 10p11 (20) (Fig. 1b).

What mutational mechanisms are responsible for the accumulation of duplicated segments within the pericentromeric region? One hypothesis put forward is that these apparent transpositions are a direct consequence of one of the many mutational forces operating on alpha satellite DNA (32). Although the data thus far are limited, there is also the suggestion that there may be specific GC-rich repetitive signals which may be involved in 'directing' duplications to the pericentromeric regions of human chromosomes (27,30,31). Recent molecular and cytogenetic analysis of marsupial interspecific hybrid chromosomes indicates that massive pericentromeric-directed invasions of transposons can occur within a single generation (49). The significance of these and other observations with respect to centromeric function awaits further clarification.

RESOLUTION OF THE CENTROMERE STRUCTURE–FUNCTION PARADOX

How will the repetitive conundrum(s) of centromere structure and function ever be resolved? Perhaps the first step is, for the time being, to dissociate these two different aspects of the centromere and to focus on the development of more comprehensive functional and structural analysis of such regions of the genome. Detailed structural analysis will entail continued sequencing of DNA associated with both neocentromere and normal centromeric regions of the genome. Of particular interest will be the elucidation of the structure of the sequence which exists at the boundaries of alpha and non-alpha satellite DNA. Such information could possibly shed light on how centromeres have evolved, as well as the types of mutational processes which have been important in their formation (22).

Unfortunately, the outlook for understanding such regions as a byproduct of the Human Genome Project (HGP) may be bleak. In the recently announced new 5 year plan for the US HGP, an ambitious schedule was outlined to develop a 'working draft' of the complete sequence by the end of 2001, with its final completion by the end of 2003 (50). Acknowledged was the fact that certain regions of the genome would probably be underrepresented, in particular areas within 'the centromeres and other constitutive heterochromatic regions'. Other areas due to their repetitive nature will pose serious challenges to the sequence-and-assemble machine of the HGP although such regions are expected to be rare. Our notion of their frequency, however, may be more than slightly biased by our inability to detect and analyze such regions. Our current understanding of the structure of the human genome is based on what we have already mapped and sequenced. Traditional sequence-tagged site (STS)-based or YAC physical mapping strategies often disintegrate and crumble as mappers approach the centromere. Efforts to develop detailed physical maps of various chromosomes indicate that the gaps often lie within regions, substantial in size, composed of repetitive DNA (51,52). The fact that some of the best-mapped, most proximal markers are often >1 Mb away from classic satellite DNA gives further testament to this fact (53). Recent estimates suggest that pericentromeric duplications may account for as much as 30–50 Mb of human DNA, above and beyond the 300 Mb already occupied by various classes of satellite DNA. Similar amounts of recently duplicated material may exist near the sub-telomeric regions as well as other regions of the human genome (54). In total, as much as 15–20% of the human genome could conceivably consist of highly repetitive DNA or large fragments of recently duplicated DNA. There is a distinct possibility that such difficult regions may be 'swept under the carpet' in the race to sequence the human genome.

Understanding the function of centromeres ultimately will require detailed analysis of both the nature of the repetitive structure and the apparent lack of repetition associated with traditional centromeres and neocentromeres, respectively. From the structural perspective of normal centromeres, there is probably little scientific merit in targeting for sequencing the many megabases of tandemly repeated alpha satellite present in the human genome. There is, however, a need for a few such regions to be studied and understood in detail. We should at least be confident of the homogeneity of such sequence and the potential patterns of repetition which may be encountered both within and flanking the primary points of constriction (22). Such patterns and their variation are inevitably important in our understanding of the mutational processes that create such regions and the association of these structures with the *trans*-acting components important for kinetochore formation. Perhaps the secrets underlying the proposed 'epigenetic' function–structure of centromeres are to be found within the pericentromeric graveyard of genes and transposons which flank the centromere. Alternatively, pericentromeric duplications may simply be byproducts of an active centromere with no immediate consequence to its competence. Nevertheless, these two papers (20,21) clearly underscore our current deficit in the understanding of centromere structure and function. Further analyses are warranted to resolve the molecular role of repetitive DNA in one of the most important biological functions of every living eukaryotic cell.

ACKNOWLEDGEMENT

I would like to thank Dr Huntington Willard for useful suggestions in the preparation of this manuscript.

REFERENCES

1. Pluta, A.F., Mackay, A.M., Ainsztein, A.M., Goldberg, I.G. and Earnshaw, W.C. (1995) The centromere: hub of chromosomal activities. *Science*, **270**, 1591–1594.
2. Choo, K. (1997) *The Centromere*. Oxford University Press, Oxford.
3. Dong, F., Miller, J.T., Jackson, S.A., Wang, G.L., Ronald, P.C. and Jiang, J. (1998) Rice (*Oryza sativa*) centromeric regions consist of complex DNA. *Proc. Natl Acad. Sci. USA*, **95**, 8135–8140.
4. Cambareri, E.B., Aisner, R. and Carbon, J. (1998) Structure of the chromosome VII centromere region in *Neurospora crassa*: degenerate transposons and simple repeats. *Mol. Cell. Biol.*, **18**, 5465–5477.
5. Sun, X., Wahlstrom, J. and Karpen, G. (1997) Molecular structure of a functional *Drosophila* centromere. *Cell*, **97**, 1007–1019.
6. Round, E.K., Flowers, S.K. and Richards, E.J. (1997) *Arabidopsis thaliana* centromere regions: genetic map positions and repetitive DNA structure. *Genome Res.*, **7**, 1045–1053.
7. Willard, H.F. (1990) Centromeres of mammalian chromosomes. *Trends Genet.*, **6**, 410–416.
8. Haaf, T., Warburton, P.E. and Willard, H.F. (1992) Integration of human alpha-satellite DNA into simian chromosomes: centromere protein binding and disruption of normal chromosome segregation. *Cell*, **70**, 681–696.
9. Larin, Z., Fricker, M.D. and Tyler-Smith, C. (1994) *De novo* formation of several features of a centromere following introduction of a Y alphoid YAC into mammalian cells. *Hum. Mol. Genet.*, **3**, 689–695.
10. Tyler-Smith, C., Oakey, R.J., Larin, Z., Fisher, R.B., Crocker, M., Affara, N.A., Ferguson-Smith, M.A., Muenke, M., Zuffardi, O. and Jobling, M.A. (1993) Localization of DNA sequences required for human centromere function through an analysis of rearranged Y chromosomes. *Nature Genet.*, **5**, 368–375.
11. Harrington, J.J., Van Bokkelen, G., Mays, R.W., Gustashaw, K. and Willard, H.F. (1997) Formation of *de novo* centromeres and construction of first-generation human artificial microchromosomes. *Nature Genet.*, **15**, 345–355.
12. Ikeno, M., Grimes, B., Okazaki, T., Nakano, M., Saitoh, K., Hoshino, H., McGill, N.I., Cooke, H. and Masumoto, H. (1998) Construction of YAC-based mammalian artificial chromosomes. *Nature Biotechnol.*, **16**, 431–439.
13. Earnshaw, W.C., Ratrie, H.D. and Stetten, G. (1989) Visualization of centromere proteins CENP-B and CENP-C on a stable dicentric chromosome in cytological spreads. *Chromosoma*, **98**, 1–12.
14. Page, S.L., Earnshaw, W.C., Choo, K.H. and Shaffer, L.G. (1995) Further evidence that CENP-C is a necessary component of active centromeres: studies of a dic(X;15) with simultaneous immunofluorescence and FISH. *Hum. Mol. Genet.*, **4**, 289–294.
15. Sullivan, B.A. and Schwartz, S. (1995) Identification of centromeric antigens in dicentric Robertsonian translocations: CENP-C and CENP-E are necessary components of functional centromeres. *Hum. Mol. Genet.*, **4**, 2189–2197.
16. Depinet, T., Zackowski, J., Earnshaw, W., Kaffe, S., Sekhon, G., Stallard, R., Sullivan, B., Vance, G., Van Dyke, D., Willard, H., Zinn, A. and Schwartz, S. (1997) Characterization of neo-centromeres in marker chromosomes lacking detectable alpha-satellite DNA. *Hum. Mol. Genet.*, **6**, 1195–1204.
17. du Sart, D., Cancilla, M.R., Earle, E., Mao, J.I., Saffery, R., Tainton, K.M., Kalitsis, P., Martyn, J., Barry, A.E. and Choo, K.H. (1997) A functional neo-centromere formed through activation of a latent human centromere and consisting of non-alpha-satellite DNA. *Nature Genet.*, **16**, 144–153.
18. Williams, B.C., Murphy, T.D., Goldberg, M.L. and Karpen, G.H. (1998) Neocentromere activity of structurally acentric mini-chromosomes in *Drosophila*. *Nature Genet.*, **18**, 30–37.
19. Willard, H.F. (1998) Centromeres: the missing link in the development of human artificial chromosomes. *Curr. Opin. Genet. Dev.*, **8**, 219–225.
20. Jackson, M.S., Rocchi, M., Thompson, G., Hearn, T., Crosier, M., Guy, J., Kirk, D., Mulligan, L., Ricco, A., Piccininni, S., Marzella, R., Viggiano, L. and Archidiacono, N. (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.
21. Barry, A.E., Howman, E.V., Cancilla, M.R., Saffery, R. and Choo, K.H. (1999) Sequence analysis of an 80-kb human analphoid neocentromere DNA. *Hum. Mol. Genet.*, **8**, 217–227.
22. Warburton, P. and Willard, H. (1996) Evolution of centromeric satellite DNA: molecular organization within and between human and primate chromosomes. In Jackson, M., Strachan, T. and Dover, G. (eds), *Human Genome Evolution*. BIOS Scientific Publishers, Oxford.
23. Jackson, M.S., See, C.G., Mulligan, L.M. and Lauffart, B.F. (1996) A 9.75-Mb map across the centromere of human chromosome 10. *Genomics*, **33**, 258–270.
24. Jackson, M.S., Slijepcevic, P. and Ponder, B.A. (1993) The organisation of repetitive sequences in the pericentromeric region of human chromosome 10. *Nucleic Acids Res.*, **21**, 5865–5874.
25. Tunnacliffe, A., Liu, L., Moore, J.K., Leversha, M.A., Jackson, M.S., Papi, L., Ferguson-Smith, M.A., Thiesen, H.J. and Ponder, B.A.J. (1993) Duplicated KOX zinc finger gene clusters flank the centromere of human chromosome 10: evidence for a pericentric inversion during primate evolution. *Nucleic Acids Res.*, **21**, 1409–1417.
26. Wong, Z., Royle, N. and Jeffreys, A. (1990) A novel human DNA polymorphism resulting from transfer of DNA from chromosome 6 to chromosome 16. *Genomics*, **7**, 222–234.
27. Borden, P., Jaenichen, R. and Zachau, H. (1990) Structural features of transposed human Vk genes and implications for the mechanism of their transpositions. *Nucleic Acids Res.*, **18**, 2101–2107.
28. 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.
29. 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.
30. 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.

31. 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.
32. 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.
33. Zimonjic, D., Kelley, M., Rubin, J., Aaronson, S. and Popescu, N. (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**.
34. Frippiat, J.P., Dard, P., Marsh, S., Winter, G. and Lefranc, M.P. (1997) Immunoglobulin lambda light chain orphans on human chromosome 8q11.2. *Eur. J. Immunol*., **27**, 1260–1265.
35. Kehrer-Sawatzki, H., Schwickardt, T., Assum, G., Rocchi, G. and Krone, W. (1997) A third neurofibromatosis type 1 (NF1) pseudogene at chromosome 15q11.2. *Hum. Genet*., **100**, 595–600.
36. 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.
37. Potier, M., Dutriaux, A., Orti, R., Groet, J., Gibelin, N., Karadima, G., Lutfalla, G., Lynn, A., Van Broeckhoven, C., Chakravarti, A., Petersen, M., Nizetic, D., Delabar, J. and Rossier, J. (1998) Two sequence-ready contigs spanning the two copies of a 200-kb duplication on human 21q: partial sequence and polymorphisms. *Genomics*, **51**, 417–426.
38. Karpen, G.H. and Allshire, R.C. (1997) The case for epigenetic effects on centromere identity and function. *Trends Genet*., **13**, 489–496.
39. Choo, K.H. (1997) Centromere DNA dynamics: latent centromeres and neocentromere formation. *Am. J. Hum. Genet*., **61**, 1225–1233.
40. Csink, A.K. and Henikoff, S. (1998) Something from nothing: the evolution and utility of satellite repeats. *Trends Genet*., **14**, 200–204.
41. DuPraw, E.J. (1968) *Cell and Molecular Biology.* Academic Press, New York.
42. Eichler, E.E. (1998) Masquerading repeats: paralogous pitfalls of the human genome. *Genome Res*., **8**, 758–762.
43. Trask, B., Friedman, C., Martin-Gallardo, A., Rowen, L., Akinbami, C., Blankenship, J., Collins, C., Giorgi, D., Iadonato, S., Johnson, F., Kuo, W., Massa, H., Morrish, T., Naylor, S., Nguyen, O., Rouquier, S., Smith, T., Wong, D., Younglbom, J. and van den Engh, G. (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. Rouquier, S., Taviaux, S., Trask, B.J., Brand-Arpon, V., van den Engh, G., Demaille, J. and Giorgi, D. (1998) Distribution of olfactory receptor genes in the human genome. *Nature Genet*., **18**, 243–250.
45. Kermouni, A., Van Roost, E., Arden, K.C., Vermeesch, J.R., Weiss, S., Godelaine, D., Flint, J., Lurquin, C., Szikora, J.P., Higgs, D.R. *et al*. (1995) The IL-9 receptor gene (IL9R): genomic structure, chromosomal localization in the pseudoautosomal region of the long arm of the sex chromosomes, and identification of IL9R pseudogenes at 9qter, 10pter, 16pter, and 18pter. *Genomics*, **29**, 371–382.
46. Amann, J., Valentine, M., Kidd, V.J. and Lahti, J.M. (1996) Localization of *chi1*-related helicase genes to human chromosome regions 12p11 and 12p13: similarity between parts of these genes and conserved human telomeric-associated DNA. *Genomics*, **32**, 260–265.
47. Wohr, G., Fink, T. and Assum, G. (1996) A palindromic structure in the pericentromeric region of various human chromosomes. *Genome Res*., **6**, 267–279.
48. Eichler, E.E., Hoffman, S.M., Adamson, A.A., Gordon, L.A., McCready, P., Lamerdin, J.E. and Mohrenweiser, H.W. (1998) Complex beta-satellite repeat structures and the expansion of the zinc finger gene cluster in 19p12. *Genome Res*., **8**, 791–808.
49. O'Neill, R.J., O'Neill, M.J. and Graves, J.A. (1998) Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature*, **393**, 68–72.
50. Collins, F.S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R. and Walters, L. (1998) New goals for the U.S. Human Genome Project: 1998. *Science*, **282**, 682–689.
51. McDermid, H.E., McTaggart, K.E., Riazi, M.A., Hudson, T.J., Budarf, M.L., Emanuel, B.S. and Bell, C.J. (1996) Long-range mapping and construction of a YAC contig within the cat eye syndrome critical region. *Genome Res*., **6**, 1149–1159.
52. Bouffard, G.G., Idol, J.R., Braden, V.V., Iyer, L.M., Cunningham, A.F., Weintraub, L.A., Touchman, J.W., Mohr-Tidwell, R.M., Peluso, D.C., Fulton, R.S., Ueltzen, M.S., Weissenbach, J., Magness, C.L. and Green, E.D. (1997) A physical map of human chromosome 7: an integrated YAC contig map with average STS spacing of 79 kb. *Genome Res*., **7**, 673–692.
53. Mahtani, M.M. and Willard, H.F. (1998) Physical and genetic mapping of the human X chromosome centromere: repression of recombination. *Genome Res*., **8**, 100–110.
54. Mazzarella, R. and Schlessinger, D. (1998) Pathological consequences of sequence duplications in the human genome. *Genome Res*., **8**, 1007–1021.