© 1997 Oxford University Press
*Human Molecular Genetics*, 1997, Vol. 6, No. 1 **9–16**

ARTICLE

# Emergence and scattering of multiple neurofibromatosis (*NF1*)-related sequences during hominoid evolution suggest a process of pericentromeric interchromosomal transposition

**Vinciane Régnier, Mounira Meddeb, Guillaume Lecointre[1], Florence Richard[2], Annette Duverger, Van Cong Nguyen, Bernard Dutrillaux[2], Alain Bernheim and Gisèle Danglot***

Cytogénétique et Génétique Oncologiques, CNRS URA 1967, Institut Gustave Roussy, 94805 Villejuif, France, [1]Service de Systématique Moléculaire et Laboratoire d'Ichtyologie, CNRS GDR 1005, Muséum National d'Histoire Naturelle, 75231 Paris cedex 05, France and [2]Cytogénétique Moléculaire et Oncologie, CNRS UMR 147, Institut Curie, 75231 Paris cedex 05, France

Received August 22, 1996; Revised and Accepted October 10, 1996

**Type 1 neurofibromatosis (*NF1*) gene encodes for a member of the GTPase activating protein family and is considered to be a tumor suppressor gene. Its very high rate of *de novo* mutation in humans led us to study a specific feature of this gene: the presence of numerous *NF1*-related sequences. According to our results, the human genome contains at least 11 *NF1*-related sequences, nine of which are scattered near centromeric sequences of seven different chromosomes. These *NF1*-related sequences, whose extent is quite varied according to loci, are unprocessed copies of the *NF1* gene, and bear numerous mutations. A phylogenetic analysis of the six largest sequences indicates that they are all derived from a common ancestor, which would have appeared 22–33 million years ago, and was subsequently duplicated several times during hominoid evolution. The most recent duplication and interchromosomal transposition occurred in the last million years suggesting that the process could still be ongoing. Intriguing similarities between the evolution of alpha-satellite DNA and *NF1*-related sequences suggest the involvement of a common genetic mechanism for the generation and pericentric spreading of these *NF1* partial copies.**

## INTRODUCTION

Neurofibromatosis type 1 is a dominantly inherited genetic disorder involving the development of benign and malignant tumors in tissues derived from the neural crest (1). The gene responsible (*NF1*) encodes for a GTPase activating protein which is considered to be a suppressor of Ras (2–4). Apart from its large size (300–350 kb), the *NF1* gene also displays several other unusual characteristics: its estimated *de novo* mutation rate ($1\times10^{-4}$) is one of the highest among human disorders (5), and it is accompanied, in the human genome, by numerous *NF1*-related sequences (6–11).

Related sequences are encountered in genomes as diverse as bacteria, *Drosophila* and mammals. They either belong to families of functional genes with analogous roles (immunoglobulin genes, tyrosine kinase genes, etc.) or to non-functional pseudogenes, believed to play a part in evolution as a reservoir of DNA sequences out of selective constraints, and occasionally used, as a rapidly evolving source of genetic material, for gene conversion or for future creation of novel genes (12). Previous data suggest that *NF1*-homologous sequences could belong to the second class (7,9). While the aim of the previously reported characterizations was to avoid possible misinterpretation in the study of the mutations occurring in NF1 patients, we were interested in more thoroughly studying the *NF1*-related sequences to obtain insight into the molecular events which have led to the generation of this family of sequences. For this purpose, *NF1*-related sequences were characterized by FISH (fluorescence *in situ* hybridization), Southern analysis, cloning and sequencing. A phylogenetic tree showing the relationships between these sequences was constructed. The dating of the emergence of these

*To whom correspondence should be addressed

Sequence data from this article have been deposited with the GenBank/EMBL/DDBJ Data Libraries under Accession Nos. YO7850 (*HcoNF1*), YO7851 (*HcoNF1*-RSa), YO7852 (*HcoNF1*-RSb), YO7853 (*HsaNF1*), YO7854 (*HsaNF1*-RS14A), YO7855 (*HsaNF1*-RS14B), YO7856 (*HsaNF1*-RS15A), YO7857 (*HsaNF1*-RS15B), YO7858 (*HsaNF1*-RS2), YO7859 (*HsaNF1*-RS22), YO7860 (*MfuNF1*)

related sequences was estimated using published rates of neutral substitutions, and finally confirmed by molecular and FISH analysis of *NF1*-related sequences in primate genomes. According to our results, a partial duplication of the *NF1* gene occurred recently in primate evolution (22–33 Myr), and this initial ancestral copy subsequently duplicated, generating new copies which rapidly spread out to the pericentromeric region of other chromosomes. The strategy developed here proved to be efficient and reliable, and we suggest it could be used for the evolutionary study of multigene families and related sequences.

## RESULTS

### Determination of the extent of individual *NF1*-related loci by Southern analysis of monochromosomal somatic cell hybrids

*NF1*-homologous sequences present in the human genome were first analyzed by FISH with an *NF1* cDNA probe covering 99% of the coding region of the gene, in order to confirm the various loci previously observed, and help us to select monochromosomal somatic cell hybrids for molecular analysis. Intense and frequent signals were observed in 15q11 and 14q11, as 84 and 80%, respectively, of the metaphases displayed at least one signal at these loci compared with 83% for the *NF1* gene. Lower signal frequencies were observed in 2q21 and 22q11 (59 and 51% of the metaphase cells had at least one signal at these loci, respectively). Finally, a few faint signals were observed in 12q12, 21q11, 1p32, 18?p/q11 and 20?p/q11 (30–15% of the metaphases showed at least one spot for these loci). No hybridization signal above the background was observed in either 2q33–34 or 15q24-qter (8,10), suggesting an exon content smaller than a few hundred base pairs for these two loci.

The composition of these various *NF1*-related loci was studied by Southern blot analysis of genomic DNA from monochromosomal somatic cell hybrids and the membranes were successively hybridized with the nine different cDNA probes defined in Table 1. Both human and rodent DNA hybridized to the probes. Consequently, a complex pattern of radioactive bands was often observed in the tracks. These bands could be identified by the use of several human DNAs (looking for possible polymorphisms), as well as mouse and hamster DNA which were identically digested and run in parallel. All results are summarized in Figure 1. Probe 1 corresponding to the beginning of the *NF1* coding sequence, and probes 6, 7, 8 and 9 corresponding to its 3′ part (exons 28–49) hybridized to human DNA from chromosome 17 containing cell hybrid, but not to any human DNA from other monochromosomal hybrids. These results confirm that *NF1*-related sequences do not contain these regions. On the contrary, probes 2, 3, 4 and 5, corresponding to the central part of *NF1* cDNA, hybridized to human DNA of several other monochromosomal hybrids. The extent of homology proved to be quite varied. A group of four chromosomes, comprising chromosomes 2, 14, 15 and 22, exhibited extensive regions of homology with the functional *NF1* gene. They all hybridized the four probes covering exons 7–28. In contrast, chromosomes 12, 18 and 21 appeared to contain very restricted homologous regions, as chromosome 12 hybridized only to probes 3 and 4 (exons 12b–23.1), while chromosomes 18 and 21 hybridized solely to probe 2 (exons 7–13). Chromosome 1 and chromosome 20 did not show any positive signal with any of the nine probes. Further analysis of the collected data led to three additional observations. The autoradiographic signals obtained with the *NF1*-related sequences present on chromosomes 14 and 15 were reproducibly more intense than the signals obtained with other related sequences or even with the *NF1* gene. This observation agrees with the presence of several copies of *NF1*-related sequences on chromosome 15 (7,8,10), and suggests an analogous situation on chromosome 14. The migration profile of related sequences frequently differed from those of the functional gene (bands were fewer and shorter), suggesting that deletions had occurred in these sequences. Finally, two of the *NF1*-related sequences, one originating from chromosome 14 and the other from chromosome 22, were very similar to each other, as the six radioactive fragments observed with chromosome 22, all comigrated with those originating from chromosome 14, and only two additional bands were specific for chromosome 14 alone.

**Table 1.** Description of the nine *NF1* cDNA probes

| Probe | Size (bp) | Exons[a] | Primer used |
|---|---|---|---|
| 1 | 885 | 2–7 | 1F: 5′-CAGGACAGCAGAACACACATACCAAAGTCA-3′ |
| | | | 1R: 5′-CACTTTCTGTCAGCTGCCTACTTCCTCCAT-3′ |
| 2 | 1243 | 7–13 | 2F: 5′-TGGACAGTCTACGAAAAGCTCTTGCTGG-3′ |
| | | | 2R: 5′-CAGAGGTGGCGGAAACAGGACATG-3′ |
| 3 | 1215 | (12B–13)–19A | 3F: 5′-GGAAGGGAAAAGGGAACTCCTCTATGGA-3′ |
| | | | 3R: 5′-GACATTTTACATCATCATCTGCTGCTTGGT-3′ |
| 4 | 1014 | 17–23–1 | 4F: 5′-AAGGCAGCTCTGAACATCTAGGGCAAGCTA-3′ |
| | | | 4R: 5′-GCCAATCAGAGGATGTGATCACAATTCGTA-3′ |
| 5 | 1061 | 23–1–28 | 5F: 5′-TACGAATTGTGATCACATCCTCTGATTGGC-3′ |
| | | | 5R: 5′-GATATAGACTGCGGAGACGTTGTCGTAAGC-3′ |
| 6 | 1090 | (27B–28)–31 | 6F: 5′-TTATGTTGCACGGAGGTTCAAAACTGGT-3′ |
| | | | 6R: 5′-TCTTTGTCGTTTGGCATCATCATTATGCT-3′ |
| 7 | 1043 | 30–36 | 7F: 5′-AGCCACACCTCACGTTAGAATTTTTGGA-3′ |
| | | | 7R: 5′-CTGCTTTATCTGCCCATGAGACACTCGT-3′ |
| 8 | 1049 | 35–43 | 8F: 5′-CATGCATGAGAGATATTCCAACGTGCA-3′ |
| | | | 8R: 5′-TGTGATCCCTGATTCCATTTCTTGCCT-3′ |
| 9 | 1037 | (42–43)–49 | 9F: 5′-CCAACACTAAGAAGTTGCTTGGAACAAGGA-3′ |
| | | | 9R: 5′-TAAAACAGGAAGTGCAGCATTACAACATGG-3′ |

[a] Primers are located within the exons indicated or at the boundary between exons in brackets.

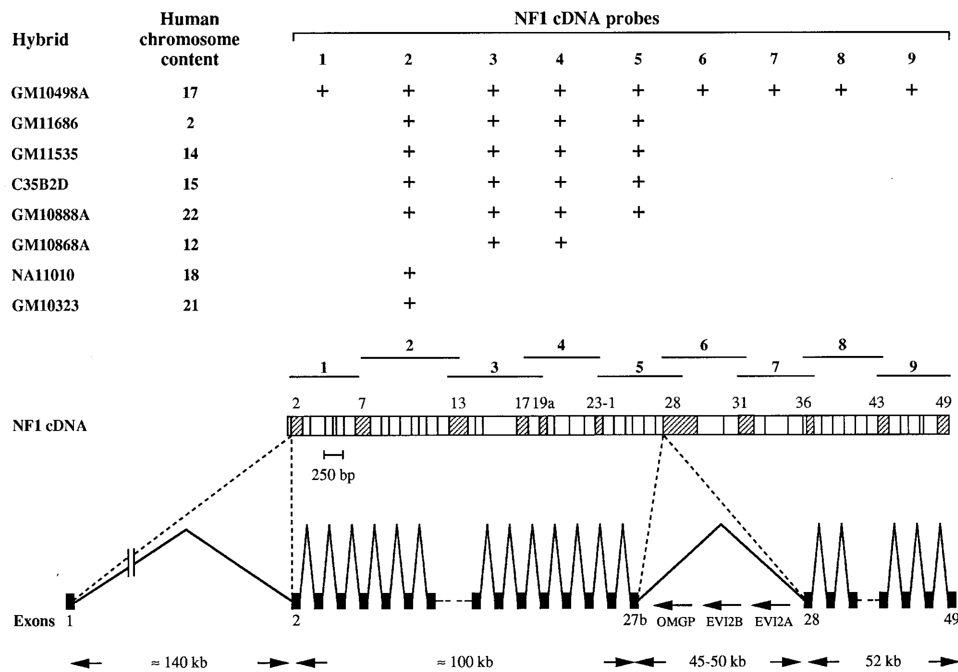


**Figure 1.** Compilation of the hybridization results of the nine *NF1* cDNA probes to Southern blots of monochromosomal human–rodent hybrid DNA. Presence of *NF1* cross-hybridizing fragment(s) on the chromosome are indicated by +. Results are presented in relation to the position of the nine overlapping probes on the *NF1* cDNA as well as to a scheme of the genomic organization of the *NF1* gene (2).

## Sequencing and multiple alignment of *NF1*-related sequences

A region of approximately 1000 bp was PCR-amplified from genomic DNA of monochromosomal somatic cell hybrids containing either the common largest *NF1*-related sequence, or the functional gene. The reactions, performed with specific primers located in exons 13 and 15 of the gene, generated a 1050 bp PCR product with the human/mouse cell hybrids containing chromosome 15 or 17, and a 1020 bp product with the hybrid containing chromosome 2. No amplification was observed with total mouse DNA. PCR amplification from human/hamster cell hybrids containing chromosome 14 or 22 yielded a common 1130 bp product. A faint additional 960 bp band of hamster origin was also observed. The human amplified DNA fragments were cloned into pBluescript and the subsequent complete sequencing of several independent clones was carried out. The *NF1* sequence obtained from chromosome 17 complies with the previously published sequence (GenBank nos. U17679, U17680), and all undetermined positions in introns could be identified (Fig. 2). Sequencing of three clones from chromosome 2 and four clones from chromosome 22 allowed the establishment of the *NF1*-related sequence present in each chromosome. In the case of clones from chromosomes 14 and 15, seven and six independent clones, respectively, were sequenced because the primary results suggested the presence of two distinct sequences on each chromosome. Indeed, two families of clones could be obtained for each chromosome. The clones within a family never differed by more than 2 bp, and each difference was confined to a single clone. On the contrary, the two chromosome 14 sequences differed in 18 bp of 1130 bp, and the two chromosome 15 sequences showed 11 differences of 1050 bp analyzed.

Multiple alignment of the *NF1* gene sequence and the six *NF1*-related sequences (*NF1*-RS2, *NF1*-RS14A, *NF1*-RS14B, *NF1*-RS15A, *NF1*-RS15B, *NF1*-RS22) showed that the exon–intron organization of *NF1* gene was well conserved in the related sequences (Fig. 2). A 78 bp insertion containing a partially repeated monomer was observed in the intron 13-like sequence of *NF1*-RS14A, *NF1*-RS14B and *NF1*-RS22. Numerous nucleotide substitutions, small insertions and deletions were encountered in all the *NF1*-related sequences analyzed. Substitutions include aberrant splice signals, as an A to G transition, at position +3 in the splice donor site of intron 14-like sequence, was present in all sequences, as well as a G to A transition in the first position of exon 15-like sequence in *NF1*-RS14 and *NF1*-RS22. Percent divergence revealed that the *NF1*-homologous sequences shared a high degree of identity (Table 2). The most divergent *NF1*-related sequences (located on chromosomes 15 and 14/22) differed by 9.4%, whereas the copies present on chromosomes 14 and 22 shared the fewest differences (0.21%). The average rate of neutral nucleotide substitution in primates is estimated to be 1.5–2×$10^{-9}$ substitutions per nucleotide site per year (13,14). Accordingly, *NF1*-related sequences would have begun to diverge from each other in the last 23–31 million years.

## Confirmation of dating and history of *NF1*-related sequences

To confirm this chronological evaluation, and unambiguously locate the appearance of *NF1*-related sequences in relation to paleontological data, we further analyzed the genomes of two primate species: the gibbon and the macaque, known to have diverged from the human lineage 22 million years and 33 million years ago, respectively (15,16). DNA purified from macaque

```
                      10                  30                  50                  70                  90                 110                 130
                 . exon 13 .                                                                                                                   .--->
MfuNF1        ATGTCCTGTTTCCGCCACCTCTGTGAGGAAGCAGATATCCGGTGTGGGGTGGATGAAGTGTCAGTGCATAACCTCTTGCCCAACTAT-AACACATTCATGGAGTTTGCCTCTGTCAGCAATATGATGTCAACAGGTAAAT
HcoNF1        ...............................................................................................................C............................
HsaNF1        ............................................................................................................................................
HsaNF1-RS15A  .....................................C.T.............................C...............G...G...............................G..................
HsaNF1-RS15B  .....................................C.T.............................C...............G...G...............................G..................
HsaNF1-RS2    ............T........................C..A............................................C........................--.........G..................
HcoNF1-RSa    ............T........................C........................G....................G.C........................--...........T................
HcoNF1-RSb    ............T..........--............C....C.....C....................................C........................--......G.....................
HsaNF1-RS14B  .....................................C...............C........G....G.................C........................--.........G..................
HsaNF1-RS22   .....................................C...............C........G....G.................C.T......................--.........G..................
HsaNF1-RS14A  .....................................C...............C........G....G.................C.T......................--.........G..................

                     150                 170                 190                 210                 230                 250                 270
MfuNF1        GTGAATAGTGGTTTTTTTTACTCAGTCTGCCTCAAAGCACATGGCATCTGATTGTGAGAATGTATTTAAGGTCACTACTTTGTAAGTGTACAGGGGAAATTCAAATAGCTTACTTGAAATCCTTCTCTGAACAAAGTAA-
HcoNF1        .................................................................C.....................T................G...................T...............
HsaNF1        .....................................................T.................................T................G...................T...............
HsaNF1-RS15A  ..................-.....AC......G..-....G..A.....A......................T..............T................G...................T...........A..T
HsaNF1-RS15B  ..................-.....AC......G..-....G........A......................T..............T.........G......G...................T...........A..T
HsaNF1-RS2    .G................-.....AC......G..-....................................T....T.C.......T................G...................T.............G.
HcoNF1-RSa    .G.........G............AC......A..-....................................T....T.C.......T................G...................T...............
HcoNF1-RSb    .G................-.....AC......A..-....................................T....T.C.......T...............GG...................T...............
HsaNF1-RS14B  A.................-.....AC..C...A..-..................................A.T...CT.C.......T......A.........G...................T...............
HsaNF1-RS22   A.................-.....AC..C...A..-..................................A.T..ACT.C.......T......A.........G...................T...............
HsaNF1-RS14A  A.................-.....AC..C...A..-..................................A.T..ACT.C.......T......A.........G...................T...............

                     290                 310                 330                 350                 370                 390                 410
MfuNF1        --------GATGAAAATAAAAAGCATTTGAAATAAATTGTTAGTCTCCTACTGAGTTGTTAATATGCTATACATTTTAAAAATAACTGGCAGTGTATGTTACTATCAGTTATGATTATGTAATTTAC-------------
HcoNF1        ........------G...........................A...T.C............................................A..............................................
HsaNF1        ..............................................T.C............................................AG.............................................
HsaNF1-RS15A  GAAACAAA.........G.................--------.....C........T..............G....................A.......T........G....C..A.....................
HsaNF1-RS15B  GAAACAAA.........G.................--------.....C.G......T.............TG....................A................G....C..A.....................
HsaNF1-RS2    ........................................G.......C........T.......................C...........A....................C............ATATGGA......
HcoNF1-RSa    ................................................C........T..............G........C...........A....................C............ATATGGA......
HcoNF1-RSb    .............................----T..........T...C........T.......................C...........A....................C............ATATGGA......
HsaNF1-RS14B  ........................G.......................C........T.............G.........T...........ACG..................C............GTTTGTATGACTA
HsaNF1-RS22   ........................G.......................C........T.....G.......G.........T...........ACG..................C............GTTTGTATGACTA
HsaNF1-RS14A  ........................G.......................C........T.....G.......G.........T...........ACG..................C............GTTTGTATGACTA

                     430                 450                 470                 490                 510                 530                 550
MfuNF1        -----------------------------------------------------------------TCTGCCTAAGTTGTAGATAGTGAAGTTTTATGAACATTGGCATATGTTTCGAGTAAATTATAGGTGGGAATAGTT
HcoNF1        ...............................................................................................CA.................T.......................C.
HsaNF1        ...............................................................................................CA.........G.......T.........G.............C.
HsaNF1-RS15A  ...................................................................CA............................C................T.......................C.
HsaNF1-RS15B  .................................................................C.CA..........................C.C................T.......................A.
HsaNF1-RS2    .......................................................GTTATGGATAA..C.-------------------------------------------AT.....................G.C.
HcoNF1-RSa    .......................................................GTTATGGATAA..C.-------------------------------------------AT..........C............C.
HcoNF1-RSb    .......................................................GTTATGGATAA..C.-------------------------------------------AT..........G......C.....C.
HsaNF1-RS14B  TGTAATTTATGTTATCCATATACATAGTCAAATTACATATGACTATGTAATTTATGTTATCCATAA..C.G..T...A.G..........-....CAC................T.......................C.
HsaNF1-RS22   TGTAATTTATGTTATCCATATACATAGTCAAATTACATATGACTATGTAATTTATGTTATCCATAA..C.G..T...A.G..........-....CAC..........T.....T.......................C.
HsaNF1-RS14A  TGTAATTTATGTTATCCATATACATAGTCAAATTACATATGACTATGTAATTTATGTTATCCATAA..C.G..T...A.G..........-....CAC..........T.....T.......................C.

                     570                 590                 610                 630                 650                 670                 690
MfuNF1        ATTTTATTCTGTGGACATTGTAGAGTCCAAGATAAGTACCTTTCTTCTGAGGTTAGTGAAAGGAAGTTTTTGGCTTTATCATTTGAAGCATTTGCTCTACTCTTCCTACTCTGCCTTTTGGGTGGGGCTTATCAGGTTCT
HcoNF1        ............................G...............C.G...................................................G............---.............T............
HsaNF1        ............................................C.G...................................................G............---...........A..............
HsaNF1-RS15A  .....G.G...................T................C.G.......................................G...........G...C....................A........G.......
HsaNF1-RS15B  .....G.G..................TT................C.G.......................................G...........G...C....................A........G.......
HsaNF1-RS2    .....G.G...................T................C.G.............C.-..................C....G...........G...C.....................................
HcoNF1-RSa    .....G.G.C.................T................C.G.............C.........................G...C.......G...C..G...............-....-.............
HcoNF1-RSb    G....G.G...................T................C.G.............C.-.......................G...........G...C..G....................-.............
HsaNF1-RS14B  .......G...................T................C.G.............C.........................G...........G...C.....................................
HsaNF1-RS22   .......G...................T................C.G.............C.........................G...........G...C.......................-...........G.
HsaNF1-RS14A  .......G...................T................C.G.............C.........................G...........G...C.......................-...........G.

                     710                 730                 750                 770                 790                 810                 830
                                                            <-.--      . exon 14 .                                                     ---.>
MfuNF1        CCATTGGCAGGCAGGGCTCTAAGTGCAGTAACTTGATTTGCTGTTGTATTTGCTTAGGAAGAGCAGCACTTCAGAAAAGAGTGATGGCACTGCTGAGGCGCATTGAGCATCCCACTGCAGGAAACACTGAGGTATGCCCT
HcoNF1        ............................................................................................................................................
HsaNF1        ............................................................................................................................................
HsaNF1-RS15A  .............................G........G.....................C................-...............G........C.T............................G...T..
HsaNF1-RS15B  ....G........................G........G.....................C................-...............G........C.T............................G...T..
HsaNF1-RS2    .............................G........G.........G...........C................-........................C..............................G..A...
HcoNF1-RSa    ................................................G...........C................-........................C..............................G......
HcoNF1-RSb    ................................................G...........C................-........................C..............................G......
HsaNF1-RS14B  ................................................G.........C.C................-........................C..............................G...T..
HsaNF1-RS22   ................................................G.........C.C...........C....-........................C..............................G...T..
HsaNF1-RS14A  ................................................G.........C.C...........C....-........................C..............................G......

                     850                 870                 890                 910                 930                 950                 970
MfuNF1        TAGCAACAGAGACACCCCTTCCAGGCGCCCACCCTCAATTTTGGAAGCCTCTTGTTACGTATGTGTGATCAGGAATAGCTTTTGAAGTAAATCTAAGATATGTGCATATTACAAGTATAATATCTGAGTATTTAATATAC
HcoNF1        ..........A........C.....................-................A..................................C..............................................
HsaNF1        ..........A........C.....................-................A..................................C......C.......................................
HsaNF1-RS15A  ......G...A........C.T........................A...A....C..A.......A..........................C............G.G..G...G...............C..T.....
HsaNF1-RS15B  ......G...A........C.T........................A...A.......A.......A..........................C............G.G..G...G...............C..T.....
HsaNF1-RS2    ..........A........CT.....................................A..........G.......................C...............CG......C.............C.......T
HcoNF1-RSa    .........GA........CT.....................................AC.........G.......................C.....................................C.......T
HcoNF1-RSb    ..........A........C......................................AC.........G.......................C.....................................C.......T
HsaNF1-RS14B  ......G...A........C.................G....................A..........G.......................C.....................................C.......T
HsaNF1-RS22   A.......A.A.....T..C......A.....G..G......................A.......................-..........C......C..............................C.......T
HsaNF1-RS14A  A.......A.A.....T..C......A.....G..G......................A.......................-..........C.....................................C.......T

                     990                1010                1030                1050                1070
                                                                                              <---. exon 15 .
MfuNF1        CTCAAGTTTGAAACTTGACTGTTGCTCATGAATGTTTAGCTCTAGACTTAAAGTTGTTTTCAAGTGATAACTGCCTTCATTTTAGGCTTGGGAAGATACACATGCAAA
HcoNF1        A................G....A...G..TG.........................C.............T.....................................
HsaNF1        A................G....A...G..TG.................--......C.............T.....................................
HsaNF1-RS15A  A................G.CA.....G..TG.C.......................C.............T.....................................
HsaNF1-RS15B  A................G.CA.....G..TG.C.......................C.............T.....................................
HsaNF1-RS2    A........C..G....G........G..TG.........................C................T..................................
HcoNF1-RSa    A...........G...AG........G..TG.........................C................T.....................C............
HcoNF1-RSb    A...........G....G........G..TG.........................C................T..................................
HsaNF1-RS14B  A...........G....G........G..TG.........................C................T...........A......................
HsaNF1-RS22   A...........G....G........G..TG.........................C................T...........A......................
HsaNF1-RS14A  A...........G....G........G..TG.........................C................T...........A......................
```

(*Macaca fuscata*) and gibbon (*Hylobates concolor*) was PCR amplified with the primers used for amplification of the human *NF1* homologous sequences. A single DNA fragment was obtained from the macaque, while three DNA fragments of similar molecular weights were observed with the gibbon. The sequencing of these DNAs indicated that the *NF1* material amplified from the macaque contained a unique sequence, highly homologous to the human functional gene, while that amplified from the gibbon contained one sequence homologous to the human functional gene, and two different sequences homologous to the human related sequence *NF1*-RS2 (*HcoNF1*-RSa and *HcoNF1*-RSb, Fig. 2). The absence of *NF1*-homologous sequences in the macaque was further confirmed by two different approaches. To overcome the possible presence of mutations in the macaque sequence complementary to the primers used in amplification reactions, four other primer couples, homologous to various positions of this *NF1* region, were used to try to amplify potential macaque *NF1*-related sequences. However, all primer couples only allowed the synthesis of the functional gene. Furthermore, FISH experiments were performed on macaque metaphase spreads from another macaque species (*Macaca sylvana*). The whole human *NF1* cDNA fluorescent probe revealed a unique locus in the pericentric region of monkey chromosomes homologous to human chromosome 17. Overall, these observations suggest that no *NF1*-related sequences are present in the macaque genome.

A phylogenetic analysis of the *NF1* functional genes (from human, macaque and gibbon) and the *NF1*-related sequences (six from human and two from gibbon) was conducted by maximum parsimony, using macaque *NF1* functional gene as outgroup (Fig. 3). The two analyses performed (see Materials and Methods) yielded the same most parsimonious tree, differing only by branch lengths. In both cases, a very high consistency index (CI) and retention index (RI) were observed, indicating a low level of homoplasy. Nodes were highly supported under the bootstrap test, and only a slight variation in internal branch length according to optimization of homoplasy was observed, allowing good confidence of the tree. As shown in Figure 3, a common branch of all related sequences separates from the functional genes branch. This common stem is strongly supported by 18–24 independent genetic events. This ancestral *NF1*-related sequence shows two successive branch divisions before the splitting between Hylobatidae and Hominidae. The first division gives rise to a branch leading to the *NF1*-RS15 family, supported by 33–41 mutational events, and a second branch, common for all other related sequences, which is supported by 18–22 mutations. This last common branch in turn divides, leading to the *NF1*-RS2 family, and the *NF1*-RS14/22 stem. Finally, it is noteworthy that the related sequences have accumulated a mean of 56–77 genetic changes since the first duplication node whereas only 10–14 events have occurred on the functional genes branch.

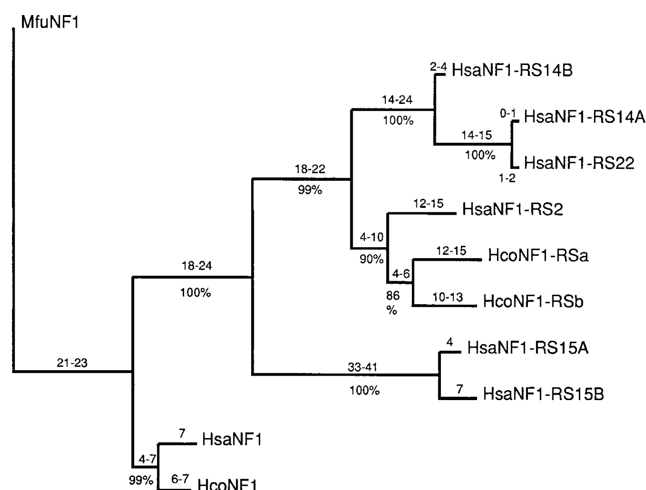


**Figure 3.** Phylogenetic tree showing relationships between the *NF1* gene and the *NF1*-related sequences in human and gibbon. The tree shown here is the most parsimonious tree obtained from the recoded data set (aligned sequences in Fig. 2) with exclusion of the region of nucleotides 408–485, by an exhaustive search of PAUP 3.1.1. The tree length is 215 steps. The data set has very few homoplasies: CI = 0.912, RI = 0.912. Branch length is proportional to the number of nucleotide changes. Numbers above each branch are the minimum and maximum length according to the optimization chosen to place homoplasies. Numbers below each branch are the bootstrap proportions.

## DISCUSSION

The aim of this study was to characterize human *NF1*-related sequences, to obtain preliminary insights into the molecular mechanisms which have led to their production. The FISH that we performed with the *NF1* cDNA probe allowed the simultaneous detection of numerous *NF1*-related sequences and showed a great variation in frequency and intensity of signals. Overall, *NF1*-related sequences appear to be scattered in nine different loci of the human genome (these results and ref. 10). It is noteworthy that seven of these loci are associated with a centromere, as six are pericentric and one is in the 2q21 region, which is known to be the relic of an ancient suppressed centromere. Human chromosome 2 has been quite recently created during the course of evolution, by the fusion of two acrocentric chromosomes still present in all great apes (simian chromosomes 12 and 13) (17). Telomeric sequences present in the 2q13 region, and alphoid DNA in 2q21 attest of the telomere–telomere fusion and the initial dicentric nature of human chromosome 2 (18,19). The study of *NF1*-related sequences present in each locus indicated that four chromosomes (chromosomes 2, 14, 15 and 22), exhibited extensive *NF1* homologous regions (encompassing exons 7–27b), while others (chromosomes 12, 18 and 21) contained much smaller *NF1*-related sequences. Although the technique used here does not allow the determination of the precise boundary of the *NF1*-related sequences, it is worth noting that the same region appears to be present in the four largest *NF1*-related sequences and that the small *NF1* homologous sequences present in the other chromosomes are also parts of this common region.

**Figure 2.** Aligned DNA sequences of a 1000 bp region of the *NF1* gene from human (*Hsa*), gibbon (*Hylobates concolor*, *Hco*), macaque (*Macaca fuscata*, *Mfu*) and *NF1*-related sequences from human and gibbon. The sequence extends from exon 13 to exon 15 of the *NF1* genomic sequence. Exon boundaries are indicated by arrows. Dots indicate identical nucleotides, and dashes indicate deletions, relative to the macaque *NF1* sequence. Monomer repeats in *NF1*-RS14 and *NF1*-RS22 are underlined. In comparison with the intron 13 sequence previously described (GenBank database, accession nos U17679, U17680), eight ambiguous positions were resolved and three nucleotide changes were noticed. A cytosine residue described at position 7 and a thymidine residue described at position 37 in sequence U17679 were missing in our data. Finally, a single guanosine residue was observed in position 5 of sequence U1780, instead of the reported N and C residues.

**Table 2.** Percentage of divergence between the *NF1* gene and its related sequences on chromosomes 2, 14, 15 and 22[a]

| | *NF1* | *NF1*-RS2 | *NF1*-RS14A | *NF1*-RS14B | *NF1*-RS15A | *NF1*-RS15B | *NF1*-RS22 |
|---|---|---|---|---|---|---|---|
| *NF1* | — | 6.6 | 8 | 6.9 | 7.6 | 7.9 | 8 |
| *NF1*-RS2 | — | — | 5.2 | 4 | 6.9 | 7.2 | 5.4 |
| *NF1*-RS14A | — | — | — | 1.9 | 9.3 | 9.4 | .21 |
| *NF1*-RS14B | — | — | — | — | 7.8 | 7.9 | 1.9 |
| *NF1*-RS15A | — | — | — | — | — | 1.2 | 9.3 |
| *NF1*-RS15B | — | — | — | — | — | — | 9.4 |

[a]The pairwise comparisons were performed on the sequences recoded as described in Materials and Methods, and deleted for region 408–485.

Sequencing of 1000 bp in the common part of the four largest *NF1*-related regions revealed the presence of two slightly different copies on chromosomes 14 and 15, then leading to a total of 11 *NF1*-related sequences in the human genome. As fluorescent hybridization signals were very intense in 14q11 and 15q11 regions, and as the 15q24–qter locus should be small (not visible in FISH with our cDNA probe) and centered around exon 24 (8), which is far from the exon 13–15 region sequenced here, we suggest that both copies are present in the pericentric region of chromosomes 14 and 15. The sequencing data showed conservation of intron–exon organization in all of the six sequences studied. An analogous situation has been described by Suzuki *et al.* (9) in the *NF1*-related sequence present in chromosome 21. Multiple points mutations, most often transitions, as well as insertions and deletions were observed, strengthening the hypothesis of unprocessed pseudogenes (6–10). Pairwise comparisons indicated that *NF1*-related sequences must have begun to diverge from each other, 23–31 million years ago. This dating was confirmed by the study of *NF1* sequences in two monkey species. PCR approaches as well as FISH experiments failed to find any evidence of *NF1*-related sequences in the macaque, whose divergence from hominoids is assumed to have occurred 33 million years ago, while two *NF1*-homologous sequences, phylogenetically related to human *NF1*-RS2, were observed in the gibbon, supposed to have split from the great apes lineage 22 million years ago. Although the absence of *NF1*-related sequences in the macaque is not fully conclusive, as these sequences could have been lost during evolution of the Cercopithecoid lineage, pairwise comparisons in human sequences as well as genome analysis of primates both converge toward the appearance of *NF1*-related sequences in the same chronological period. The phylogenetic tree deduced from all of the *NF1* sequences determined in humans and in the two primates indicates a common ancestral sequence for all related sequences. The branch length leading to the related sequences is greater than that leading to the functional genes, supporting the absence of a functional role for these sequences in the cell. The grouping of related sequences cannot result from an artifact such as long branch attraction (20), as this phenomenon needs high levels of homoplasy. On the contrary, our data base has an exceptionally low level of homoplasy (RI of 0.912). As our Southern blot results show a common extent for all the largest human homologous sequences, the best model fitting with all these data supposes the occurrence of a duplication of the central part of the *NF1* gene, ~22–33 million years ago. This partial unprocessed *NF1* copy would have led, after duplication, to the sequences actually present on human chromosomes 2 and 15. The ancestor of the actual related sequence on chromosome 2 subsequently duplicated, gave rise to the family of sequences now located on human chromosomes 14 and 22. According to the most parsimonious tree, both branchings would have occurred before the splitting between gibbon and human lineage (–22 million years). Related sequences, homologous to human *NF1*-RS15 and *NF1*-RS14 should then have been observed in the gibbon genome. Further studies will be necessary to determine whether these sequences were not detected because of technical problems (point mutations or deletions in primer sites) or whether they were lost during the course of Hylobatidae evolution.

It is noteworthy that *NF1*-RS14A and *NF1*-RS22, which are present on two different chromosomes, show only 0.21% divergence. According to a divergence rate of $1.2{\times}10^{-9}$ in most recent human evolution (13,21), this observation indicates that one of the related sequences present in the pericentromeric region of chromosome 14 was duplicated and transposed to the pericentromeric region of chromosome 22 less than one million years ago. The initial *NF1*-related sequence which emerged 22–33 million years ago, was therefore regularly duplicated during hominoid evolution and this process of duplication/transposition could still be ongoing in human lineage.

The spreading of the *NF1*-related sequences on multiple chromosomes is particularly puzzling. Interchromosomal exchanges do occur during the course of evolution, but they are far too rare to explain the rapid scattering of *NF1*-related sequences in the recent evolution of primates. For example, human chromosome 17, which bears the *NF1* gene is completely syntenic to its mouse homologue (mChr11) (22), whereas rodents are supposed to have diverged from the human lineage 75–110 million years ago. The great majority of *NF1*-related sequences are pericentric, like the *NF1* functional gene, suggesting a region-specific directed mechanism for the generation of these interchromosomal duplications. Such a mechanism, specifically attached to pericentromeric regions, is also supported by quite recent observations by Eichler *et al.* (23) and Wöhr *et al.* (24). The first described a 27 kb synteny between human 16p11 and Xq28 regions. The Xq28 locus contains two functional genes, while the 16p11 locus contains homologous pseudogenes. The second reports the presence of ~40 copies of a 240 kb anonymous sequence inserted between α satellite and satellite III of 10 human chromosomes. In both cases, multiple copies of these sequences are present in the pericentromeric regions of non-orthologous chromosomes of primates, especially great apes.

Centromeric sequences consist of millions of base pairs of highly organized repeated DNA families. The predominant class is the alpha satellite DNA, which is exclusively present in human and primates. Analogies between the evolution of alpha satellite DNA and *NF1*-related sequences are intriguing. While human

alpha satellite sequences are all derived from three ancient suprafamilies, it is worth noting that six of the seven *NF1*-related loci associated with a centromere are situated on chromosomes belonging to the same family (suprafamily 2 containing chromosomes 2, 4, 8, 9, 13, 14, 15, 18, 20, 21 and 22). Furthermore the phylogenetic tree of human alpha satellite DNA shows striking similarities with the phylogenetic tree of our *NF1*-related sequences, especially exhibiting high homology between alpha satellite sequences of chromosomes 14, 15 and 22 (25). Alpha satellite sequences of a given subfamily were shown to extend over 250 to >4000 kb but show only 1–5% divergence. Such an organization requires that the similarity of sequences must be maintained by mechanisms such as transposition, unequal crossing-over and sequence conversion, leading to a permanent process of homogenization (26). These mechanisms, which lead to a concerted evolution of alpha satellite subfamilies, are supposed to be very efficient within a given centromere, less frequent between homologous chromosomes and even more rare between non-homologous chromosomes (27). Our results show the preferential spreading of *NF1*-related sequences in the pericentric region of chromosomes containing highly homologous alpha satellite sequences. This finding suggests that the efficient mechanisms which generated and scattered the multiple *NF1*-related sequences on numerous chromosomes in a short period of evolution might be related to those involved in homogenization of alpha satellite sequences.

## MATERIALS AND METHODS

### *NF1* cDNA probes

Nine overlapping *NF1* cDNA fragments of 800–1200 bp (characterized in Table 1) were obtained by RT–PCR from normal human brain total RNA as previously described (28,29) and cloned in pBluescript vector. The plasmid probes were labeled by nick-translation in the presence of $^{32}$P dCTP for Southern analysis, and Biotin16-dUTP (Boehringer Mannheim) or Biotin14-dATP (Gibco-BRL) for FISH.

### FISH analysis of *NF1* sequences in human and macaque genomes

Normal stimulated human lymphocytes were harvested and metaphase spreads were prepared using standard methods. *In situ* hybridization was performed as described previously (30,31). The human *NF1* probe labeled with biotin (80 ng) was mixed with human Cot1 DNA and sonicated salmon sperm, and applied to slides. After hybridization, the signals were revealed by the addition of FITC–avidin, biotinylated anti-avidin antibody and a second layer of FITC–avidin. The FISH signals were either photographed or acquired with a computer-driven cooled CCD color camera. Localization of the various fluorescent signals and analysis of their respective frequencies were performed by examination of 70 metaphase cells.

Macaque metaphase spreads were prepared after fibroblast culture of a skin biopsy of a *Macaca sylvana* specimen (Muséum National d'Histoire Naturelle, Paris). FISH hybridization of human *NF1* probe to macaque metaphases was performed with the following modifications. After RNAse treatment, slides were incubated in the presence of pepsin (0.5 mg/ml in $10^{-2}$ N HCl for 10 min at 37°C), postfixed in ethanol/acetic acid and then dehydrated. After 3 days hybridization at 37°C, washings were performed at 38°C.

### Southern analysis of monochromosomal somatic cell hybrids

Genomic DNA from 10 monochromosomal human rodent somatic cell hybrids containing chromosomes 1, 2, 12, 14, 17, 18, 20, 21 and 22 (NA13139, GM11686, GM10868A, GM11535, GM10498A, NA11010, NA13140, GM10323, GM10888A, respectively, obtained from the NIGMS Human Genetic Mutant Cell Repository, Camden, NJ, USA) and C35B2D containing chromosome 15 (32) were analyzed. DNA from hybrid cell lines and from human, mouse and hamster cell lines was digested with the restriction enzymes *Eco*RI, *Hin*dIII and/or *Taq*I, electrophoresed and transferred on to Hybond-N+ membranes (Amersham). The membranes were successively hybridized with the nine *NF1* cDNA probes defined in Table 1. Prehybridization, hybridization and washings were performed according to standard procedures.

DNA samples (300 ng) from monochromosomal somatic cell hybrids containing human chromosomes 2, 14, 15, 17 and 22 (GM11686, GM11535, NA11715, GM10498A and GM10888A) were used as templates for PCR amplification with two *NF1* specific primers located in exons 13 and 15 (5′-CCTGACACT-GAAGCTGTTCTGGTTGCC-3′ and 5′-AAGGATTAGCTTT-GTTGCTTGTTCCCAT-3′, respectively). In addition, DNA from two species of primates (*Macaca fuscata* and *Hylobates concolor*, Muséum National d'Histoire Naturelle, Paris) was also used. Amplification conditions were: denaturation for 45 s at 95°C, annealing for 3 min at 68–69°C, and extension for 3 min at 72°C for 30–40 cycles. PCR products were purified on small columns, and directly cloned in a pBluescript vector. DNA sequencing reactions were performed using a PRISM Ready Reaction DyeDeoxy Terminator Cycle Sequencing Kit (Applied Biosystems). Extension products were analyzed on the Applied Biosystems Model 373A DNA Sequencing System. Three to seven independent clones derived from each ligation reaction were sequenced on both strands using vector-based primers as well as the following primers designed to match at best *NF1* and *NF1*-related sequences: 5′-GCTGTTGTAGTTGCTTAGG-3′; 5′-AAAAACTCAGTGGGAGACTA-3′; 5′-TTGTTAGTCTC-CCACTGAGTT-3′; 5′-CAGCAGTGCCATCACTCTTT-3′.

### Computer-assisted sequence analysis

Multiple nucleotide alignment was performed with the CLUSTAL W V1.5 program (33). Alignment was controlled using the facilities of the ED program of the MUST package (34). Transitional saturation was tested, using the NET and COMPMAT programs of MUST, by plotting pairwise transitional differences against pairwise transversional differences. Phylogenetic analyses were performed with the maximum parsimony method using PAUP V3.1.1 (35). In the absence of transitional saturation, no weighting scheme was applied. For analysis, several contiguous gaps were recoded as a single character. When some sequences differed from others within insertions, Barriel's coding method (36) was used. A first analysis was performed using the whole recoded data. In a second analysis, a problematic region including multiple superimposed insertion–deletion–substitution events (region 408–485) was deleted because of uncertainties in recoding interpretations. Exhaustive searches

were performed in each case under ACCTRAN optimization. Robustness of nodes was analyzed in terms of branch length and bootstrap analyses (37), which were performed with PAUP and 1000 iterations.

## ACKNOWLEDGMENTS

We thank Dr D. Malleyran and Dr F. Claro, from the Parc Zoologique de Paris (Muséum National d'Histoire Naturelle) at Vincennes for having provided tissue samples of *Hylobates concolor* and *Macaca fuscata*. We are grateful to T. Chedeville (C.T.S. of Créteil) for normal blood samples and C. Scamps for technical advices. This work was supported by grants from the CNRS, Institut Gustave Roussy, the association pour la Recherche contre le Cancer (ARC), and the Ligue Nationale contre le Cancer. V.R. is supported by a doctoral fellowship from the Ministère de l'Education Nationale, de l'Enseignement Supérieur et de la Recherche.

## ABBREVIATIONS

RT, reverse transcription; PCR, polymerase chain reaction; bp, base pair; Myr, million years; FISH, fluorescence *in situ* hybridization; CI, consistency index; RI, retention index.

## REFERENCES

1. Riccardi, V. M. (1992) *Neurofibromatosis: phenotype, natural history, and pathogenesis.* Johns Hopkins University Press, Baltimore.
2. Li, Y., O'Connell, P., Breidenbach, H. H., Cawthon, R., Stevens, J., Xu, G., Neil, S., Robertson, M., White, R. and Viskochil, D. (1995) Genomic organization of the neurofibromatosis 1 gene (*NF1*). *Genomics*, **25**, 9–18.
3. DeClue, J. E., Papageorge, A. G., Fletcher, J. A., Diehl, S. R., Ratner, N., Vass, W. C. and Lowy, D. R. (1992) Abnormal regulation of mammalian p21*ras* contributes to malignant tumor growth in von Recklinghausen (type 1) neurofibromatosis. *Cell*, **69**, 265–273.
4. Basu, T. N., Gutmann, D. H., Fletcher, J. A., Glover, T. W., Collins, F. S. and Downward, J. (1992) Aberrant regulation of ras proteins in malignant tumour cells from type 1 neurofibromatosis patients. *Nature*, **356**, 713–715.
5. Huson, S. M., Compston, D. A., Clark, P. and Harper, P. S. (1989) A genetic study of von Recklinghausen neurofibromatosis in south east Wales. I. Prevalence, fitness, mutation rate, and effect of parental transmission on severity. *J. Med. Genet.*, **26**, 704–711.
6. Marchuk, D. A., Tavakkol, R., Wallace, M. R., Brownstein, B. H., Taillon-Miller, P., Fong, C. T., Legius, E., Andersen, L. B., Glover, T. W. and Collins, F. S. (1992) A yeast artificial chromosome contig encompassing the type 1 neurofibromatosis gene. *Genomics*, **13**, 672–680.
7. Legius, E., Marchuk, D. A., Hall, B. K., Andersen, L. B., Wallace, M. R., Collins, F. S. and Glover, T. W. (1992) *NF1*-related locus on chromosome 15. *Genomics*, **13**, 1816–1818.
8. Gasparini, P., Grifa, A., Origone, P., Coviello, D., Antonacci, R. and Rocchi, M. (1993) Detection of a neurofibromatosis type 1 (*NF1*) homologous sequence by PCR: implication for the diagnosis and screening of genetic diseases. *Mol. Cell. Probes*, **7**, 415–418.
9. Suzuki, S., Ozawa, N., Taga, C., Kano, T., Hattori, M. and Sakaki, Y. (1994) Genomic analysis of a *NF1*-related pseudogene on human chromosome 21. *Gene*, **147**, 277–280.
10. Purandare, S. M., Breidenbach, H. H., Li, Y., Zhu, X. L., Sawada, S., Neil, S. M., Brothman, A., White, R., Cawthon, R. and Viskochil, D. (1995) Identification of neurofibromatosis 1 (*NF1*) homologous loci by direct sequencing, fluorescence in situ hybridization, and PCR amplification of somatic cell hybrids. *Genomics*, **30**, 476–485.
11. Hulsebos, T. J. M., Bijleveld, E. H., Riegman, P. H. J., Smink, L. J. and Dunham, I. (1996) Identification and characterization of NF1-related loci on human chromosomes 22, 14 and 2. *Hum. Genet.*, **98**, 7–11.
12. Fitch, D. H., Bailey, W. J., Tagle, D. A., Goodman, M., Sieu, L. and Slightom, J. L. (1991) Duplication of the γ-globin gene mediated by L1 long interspersed repetitive elements in an early ancestor of simian primates. *Proc. Natl Acad. Sci. USA*, **88**, 7396–7400.
13. Miyamoto, M. M., Slightom, J. L. and Goodman, M. (1987) Phylogenetic relations of humans and African apes from DNA sequences in the ψη-globin region. *Science*, **238**, 369–373.
14. Sakoyama, Y., Hong, K. J., Byun, S. M., Hisajima, H., Ueda, S., Yaoita, Y., Hayashida, H., Miyata, T. and Honjo, T. (1987) Nucleotide sequences of immunoglobulin ε genes of chimpanzee and orangutan: DNA molecular clock and hominoid evolution. *Proc. Natl Acad. Sci. USA*, 84, 1080–1084.
15. Pilbeam, D. (1984) The descent of hominoids and hominids. *Sci. Am.*, **250**, 84–96.
16. Conroy, G. C. (1990) *Primate evolution*. Norton, New York.
17. Dutrillaux, B., Rethore, M. O. and Lejeune, J. (1975) Comparison of the karyotype of the orangutan (*Pongo pygmaeus*) to those of man, chimpanzee, and gorilla. *Ann. Genet.*, **18**, 153–161.
18. 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.
19. 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.
20. Felsenstein, J. (1978) Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.*, **27**, 401–410.
21. Gibbons, A. (1995) When it comes to evolution, humans are in the slow class. *Science*, **267**, 1907–1908.
22. DeBry, R. W. and Seldin, M. F. (1996) Human/mouse homology relationships. *Genomics*, **33**, 337–351.
23. Eichler, E. E., Lu, F., 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.
24. Wöhr, G., Fink, T. and Assum, G. (1996) A palindromic structure in the pericentromeric region of various human chromosomes. *Genome Res.*, **6**, 267–279.
25. Alexandrov, I. A., Mitkevich, S. P. and Yurov, Y. B. (1988) The phylogeny of human chromosome specific alpha satellites. *Chromosoma*, **96**, 443–453.
26. Jorgensen, A. L., Bostock, C. J. and Bak, A. L. (1987) Homologous subfamilies of human alphoid repetitive DNA on different nucleolus organizing chromosomes. *Proc. Natl Acad. Sci. USA*, **84**, 1075–1079.
27. Warburton, P. E. and Willard, H. F. (1990) Genomic analysis of sequence variation in tandemly repeated DNA. Evidence for localized homogeneous sequence domains within arrays of α-satellite DNA. *J. Mol. Biol.*, **216**, 3–16.
28. Régnier, V., Danglot, G., Nguyen, V. C. and Bernheim, A. (1995) A Tsp509I variant in exon 13 of the neurofibromatosis type 1 (*NF1*) gene allows the identification of both alleles at the mRNA level. *Hum. Genet.*, **96**, 131–132.
29. Danglot, G., Régnier, V., Fauvet, D., Vassal, G., Kujas, M. and Bernheim, A. (1995) Neurofibromatosis 1 (*NF1*) mRNAs expressed in the central nervous system are differentially spliced in the 5′ part of the gene. *Hum. Mol. Genet.*, **4**, 915–920.
30. Slim, R., Weissenbach, J., Nguyen, V. C., Danglot, G. and Bernheim, A. (1991) Relative order determination of four Yp cosmids on metaphase and interphase chromosomes by two-color competitive *in situ* hybridization. *Hum. Genet.*, **88**, 21–26.
31. Rosier, M. F., Goguel, A. F., Martin, A., Le Paslier, D., Couillin, P., Houlgatte, R., Bernheim, A., Auffray, C. and Devignes, M. D. (1994) A 1.7-Mb YAC contig around the human BDNF gene (11p13): integration of the physical, genetic, and cytogenetic maps in relation to WAGR syndrome. *Genomics*, **24**, 69–77.
32. Diriong, S., Danglot, G., Bernheim, A. and Nguyen, V. C. (1995) A new hybrid cell line containing only human chromosome 15 selected through fluorescence *in situ* hybridization and characterized by Alu-PCR amplification of the human DNA. *Ann. Genet.*, **38**, 27–31.
33. Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
34. Philippe, H. (1993) MUST: a computer package of Management Utilities for Sequences and Trees. *Nucleic Acids Res.*, **21**, 5264–5272.
35. Swofford, D. L. (1993) *Phylogenetic Analysis Using Parsimony (PAUP) 3.1.1.* Illinois Natural History Survey, Champaign, IL.
36. Barriel, V. (1994) Phylogénies moléculaires et insertions-délétions de nucléotides. *C. R. Acad. Sci., Paris, Sc. Vie*, **317**, 693–701.
37. Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**, 783–791.