Available online at www.sciencedirect.com

SCIENCE DIRECT®

Genomics 86 (2005) 383 – 395

GENOMICS

www.elsevier.com/locate/ygeno

A rhesus macaque radiation hybrid map and comparative analysis with the human genome

William J. Murphy[a,*], Richa Agarwala[b], Alejandro A. Schäffer[c], Robert Stephens[d], Clarence Smith Jr.[e], Nicole J. Crumpler[a], Victor A. David[e], Stephen J. O'Brien[e]

[a]*Basic Research Laboratory, SAIC-Frederick, Inc., Laboratory of Genomic Diversity, National Cancer Institute, Frederick, MD 21702, USA*
[b]*NLM/NCBI/IEB, National Institutes of Health, Department of Health & Human Services, Bethesda, MD 20894, USA*
[c]*NLM/NCBI/CBB, National Institutes of Health, Department of Health & Human Services, Bethesda, MD 20894, USA*
[d]*Advanced Biomedical Computing Center, National Cancer Institute, Frederick, MD 21702, USA*
[e]*Laboratory of Genomic Diversity, National Cancer Institute, Frederick, MD 21702, USA*

Received 17 March 2005; accepted 25 May 2005
Available online 21 July 2005

Abstract

The genomes of nonhuman primates are powerful references for better understanding the recent evolution of the human genome. Here we compare the order of 802 genomic markers mapped in a rhesus macaque (*Macaca mulatta*) radiation hybrid panel with the human genome, allowing for nearly complete cross-reference to the human genome at an average resolution of 3.5 Mb. At least 23 large-scale chromosomal rearrangements, mostly inversions, are needed to explain the changes in marker order between human and macaque. Analysis of the breakpoints flanking inverted chromosomal segments and estimation of their duplication divergence dates provide additional evidence implicating segmental duplications as a major mechanism of chromosomal rearrangement in recent primate evolution.
Published by Elsevier Inc.

Keywords: Rhesus macaque; Radiation hybrid map; Human genome; Chromosome evolution; Segmental duplication

Introduction

Comparative genome analysis of mankind's primate relatives is important for a comprehensive understanding of those genetic aspects that make human biology unique and for applying primate animal models to inform investigations of human disease [1]. The rhesus macaque (*Macaca mulatta*) is of particular importance as a major nonhuman primate model for studies ranging from cardiovascular disease to AIDS vaccine research. Nonhuman primates like the macaque also provide useful outgroup species for interpreting recent chromosomal changes in the human genome and illustrating the underlying genomic mechanisms operative in karyotypic evolution [1–4].

Primate chromosomes and genomes have been extensively studied using molecular cytogenetic approaches. Chromosome painting of nearly 50 primate species with human flow-sorted chromosome probes has revealed the general patterns of synteny conservation and translocation events distinguishing different primate lineages [5–9]. In addition, there has been limited molecular analysis of intrachromosomal rearrangements in distantly related primate genomes [10], and most of these have largely focused on only a few chromosomes [3,4,11–15]. A few of these studies have suggested a potential role for segmental duplications in primate chromosomal rearrangements [4,12,15,16].

We previously constructed and characterized a radiation hybrid panel as a mapping tool for the rhesus macaque, and as a resource for uncovering chromosomal rearrangements

* Corresponding author. Present address: Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, Mail Stop 4458, College Station, TX 77843-4458, USA. Fax: +1 979 845 9972.
E-mail address: wmurphy@cvm.tamu.edu (W.J. Murphy).

0888-7543/$ - see front matter. Published by Elsevier Inc.
doi:10.1016/j.ygeno.2005.05.013

in recent primate evolution [17]. Here we have exploited this tool by constructing a first-generation comparative map between the rhesus macaque and human genomes. By comparing the order of conserved chromosomal segments with other mammalian genomes we determined those inter- and intrachromosomal rearrangements that occurred in the human lineage since divergence from Old World Monkeys 25 million years ago (Mya) [18]. Examination of the human DNA sequence content at hominoid-specific chromosomal breakpoints extends our insights into the mechanism and timing of primate chromosome rearrangements.

Results

Map construction

A total of 1094 STS markers were designed and genotyped in the 5000-rad macaque RH panel. STS markers were from a variety of sources: *Macaca fascicularis* expressed sequence tags (ESTs) (N = 420), *Macaca* (any macaque species) or *Papio* (baboon) mRNAs from GenBank (N = 285), comparative anchor locus-derived STSs [17] (N = 35), human microsatellites previously optimized and positioned on a baboon linkage map [10] (N = 320), and newly developed *M. mulatta* microsatellites (N = 34). Of these, 292 were dropped for one of the following reasons: weak amplification, high hamster background amplification, excessively high retention frequency (>70% and not predicted to reside on the selectable locus chromosome), and linkage to multiple chromosomes with high LOD scores. The remaining set of 802 markers was used to construct the final RH map, containing 441 MLE-consensus markers (see Methods). Because of the strong conservation of synteny observed between human and macaque genomes using chromosome painting, we assigned markers to chromosome linkage groups assuming complete conservation of synteny between macaque and human genomes. These assignments were further evaluated and verified by 2-point LOD scores (see Methods).

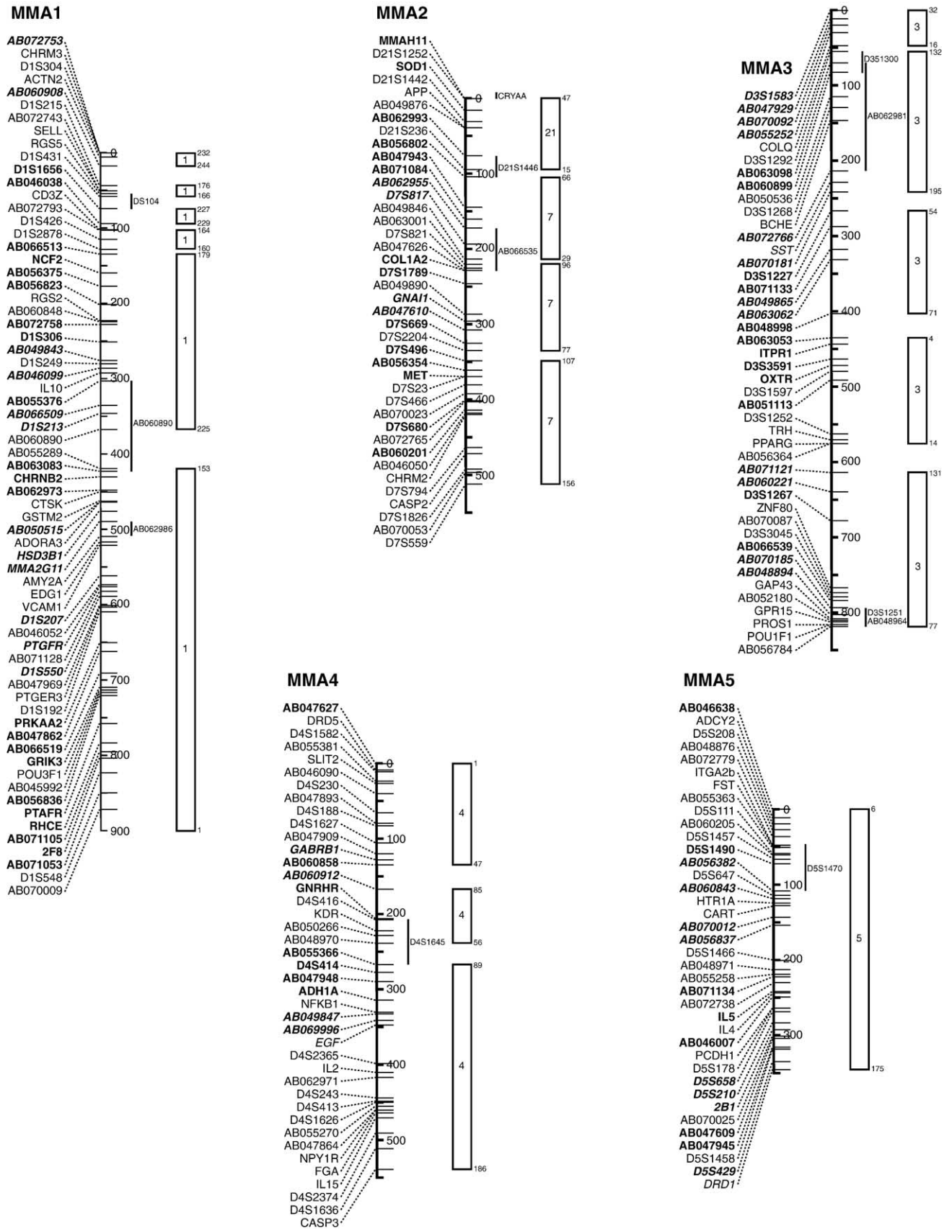
Of the 361 markers not in the MLE-consensus map, 198 markers that were initially removed as being too close were repositioned relative to the MLE-consensus map using RHMAXLIK. The remaining 163 markers that were removed during MLE-consensus map quality checks (see Methods) were placed into their most likely map intervals, comparing best placement to second best placement. These include: (1) markers with excessively high or low retention frequency relative to the chromosomal average, (2) low 2-point LOD scores, (3) uncertain order in the consensus map or flips evaluation of marker order, (4) too many “2” vector entries. Any plausible bins within a LOD of >2.0 of the best placement were documented (Fig. 1).

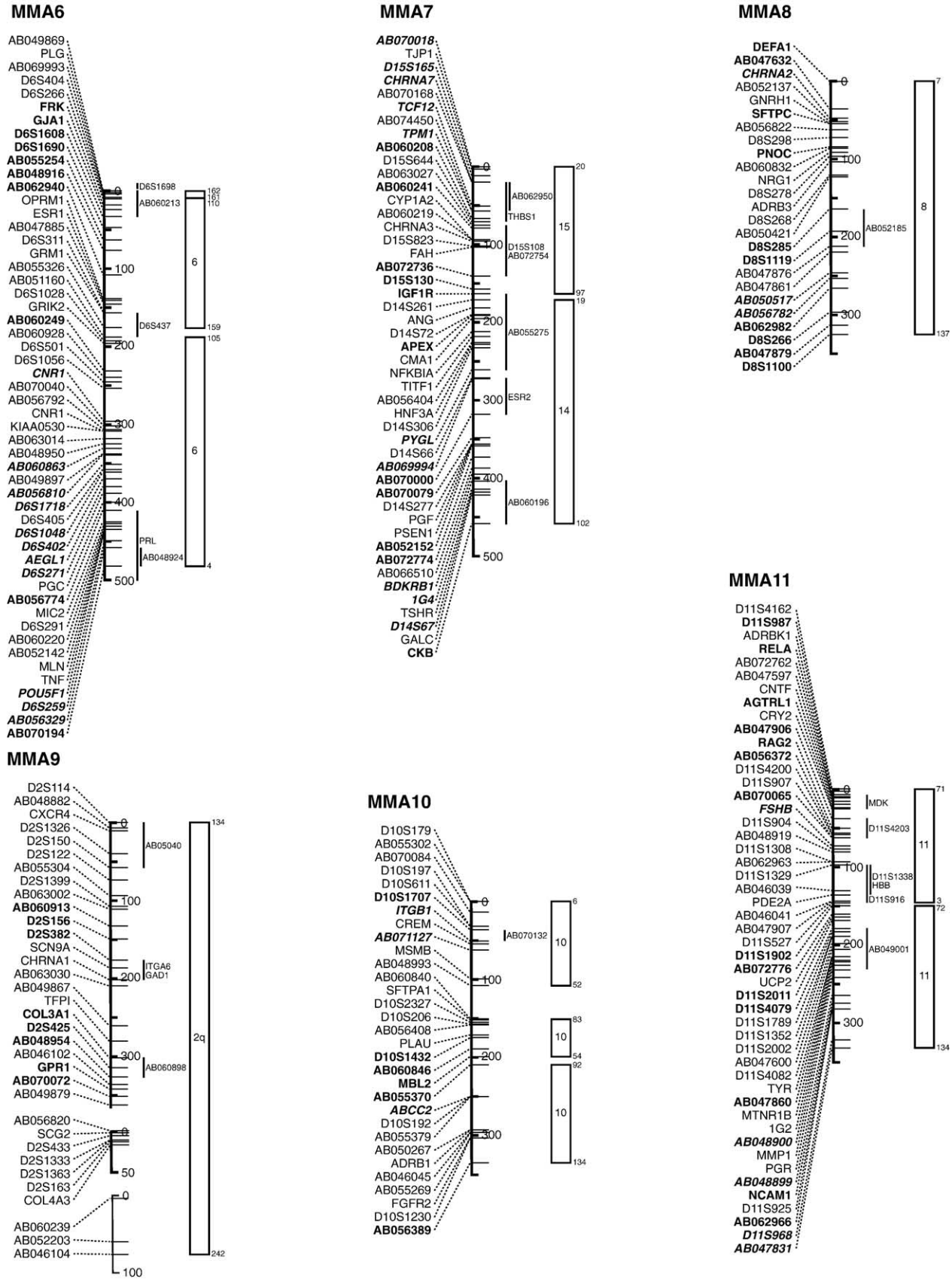
Together these markers cover all 20 macaque autosomes, and X and Y chromosomes (Table 1). Fig. 1 shows RH maps for all macaque chromosomes (hereafter abbreviated Mma-#) and the corresponding comparative map of the human genome. Most chromosomes show a strong degree of marker order conservation between human and macaque, though several are rearranged in two or more blocks relative to human. Two large macaque regions with extreme retention frequency were found: the long arm of Mma9 with very low retention frequency for which we have no explanation, and all of Mma17 with high retention frequency, which contains the locus (*TK1*) selected for in the original RH panel construction experiments [17]. This resulted in multiple linkage groups per chromosome that could not be oriented with the RH data. We have tentatively ordered the linkage groups on Mma9 based on the orientation of included microsatellites from these same regions in the baboon linkage map [10], which is karyotypically colinear with the macaque genome [19].

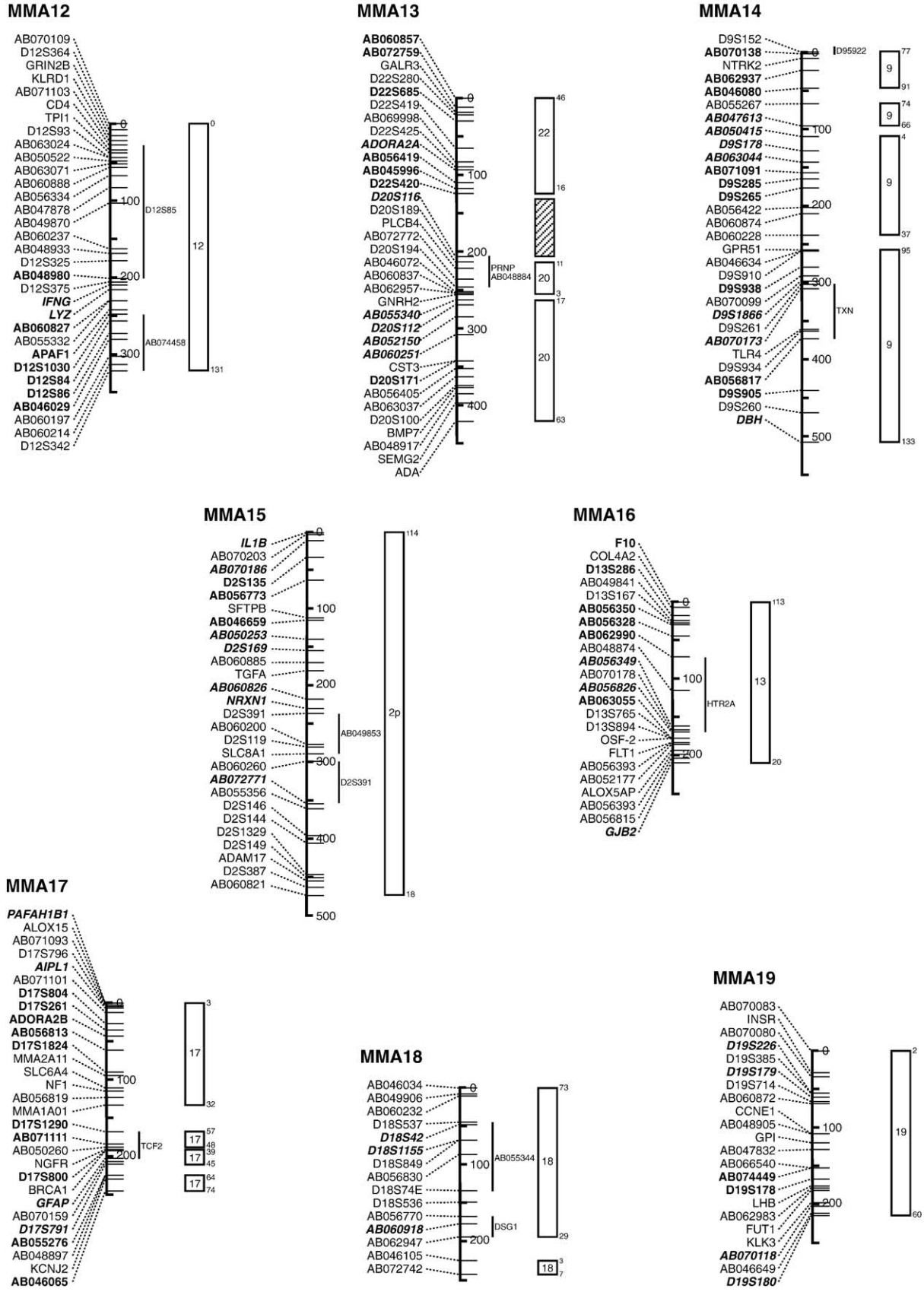
Comparative analysis

Previous reciprocal chromosome painting studies between human and macaque genomes identified 25 syntenic blocks, including the sex chromosomes [20]. Most human chromosomes are syntenic to a single macaque chromosome, with the exceptions of macaque chromosomes 2 (Hsa7 and Hsa21), 7 (Hsa14 and Hsa15), 9 (Hsa2q), 13 (Hsa20 and Hsa22), and 15 (Hsa2p). Examination of conserved marker order (i.e., conserved segments) within chromosomes revealed several chromosomes that require multiple inversions or translocations to transform the macaque-conserved block order in to the human genome order (Fig. 1). Using a rule where we required conserved segments (in this case due to inversions) to be supported by at least 2 MLE-consensus markers positioned at LOD $\geq$ 3.0 and/or concordant marker order in the baboon linkage map, we were able to define 51 blocks of conserved order between macaque and human genomes. We estimate that a minimum of 23 rearrangements can convert one genome into the other (Table 1). This estimate is at best a lower bound of the true number of rearrangements between the two genomes as small micro-inversions, though probably not exceedingly common between these close relatives, likely escaped detection given the resolution limits of RH mapping.

For each human chromosome we compared the order and orientation of conserved segments in macaque with other outgroup genomes to determine the polarity and ancestral form of each chromosome. Using the human genome as a reference, a human-specific break is diagnosed by a conserved segment break in all other species at the same position in the human genome (Fig. 2). A macaque-specific break would be a break only in macaque but no other species. Using this approach we were able to discern which rearrangement events preceded the macaque–human divergence, which were unique to the macaque lineage, and those that were unique to the human lineage (Table 1). For those unique to the human lineage we refined the ancestral


MMA1
AB072753
CHRM3
D1S304
ACTN2
AB060908
D1S215
AB072743
SELL
RGS5
D1S431
D1S1656
AB046038
CD3Z
AB072793
D1S426
D1S2878
AB066513
NCF2
AB056375
AB056823
RGS2
AB060848
AB072758
D1S306
AB049843
D1S249
AB046099
IL10
AB055376
AB066509
D1S213
AB060890
AB055289
AB063083
CHRNB2
AB062973
CTSK
GSTM2
AB050515
ADORA3
HSD3B1
MMA2G11
AMY2A
EDG1
VCAM1
D1S207
AB046052
PTGFR
AB071128
D1S550
AB047969
PTGER3
D1S192
PRKAA2
AB047862
AB066519
GRIK3
POU3F1
AB045992
AB056836
PTAFR
RHCE
AB071105
2F8
AB071053
D1S548
AB070009
DS104
AB060890
AB062986
MMA2
MMAH11
D21S1252
SOD1
D21S1442
APP
AB049876
AB062993
D21S236
AB056802
AB047943
AB071084
AB062955
D7S817
AB049846
AB063001
D7S821
AB047626
COL1A2
D7S1789
AB049890
GNAI1
AB047610
D7S669
D7S2204
D7S496
AB056354
MET
D7S23
D7S466
AB070023
D7S680
AB072765
AB060201
AB046050
CHRM2
D7S794
CASP2
D7S1826
AB070053
D7S559
CRYAA
D21S1446
AB066535
MMA3
D3S1583
AB047929
AB070092
AB055252
COLQ
D3S1292
AB063098
AB060899
AB050536
D3S1268
BCHE
AB072766
SST
AB070181
D3S1227
AB071133
AB049865
AB063062
AB048998
AB063053
ITPR1
D3S3591
OXTR
D3S1597
AB051113
D3S1252
TRH
PPARG
AB056364
AB071121
AB060221
D3S1267
ZNF80
AB070087
D3S3045
AB066539
AB070185
AB048894
GAP43
AB052180
GPR15
PROS1
POU1F1
AB056784
D351300
AB062981
D3S1251
AB048964
MMA4
AB047627
DRD5
D4S1582
AB055381
SLIT2
AB046090
D4S230
AB047893
D4S188
D4S1627
AB047909
GABRB1
AB060858
AB060912
GNRHR
D4S416
KDR
AB050266
AB048970
AB055366
D4S414
AB047948
ADH1A
NFKB1
AB049847
AB069996
EGF
D4S2365
IL2
AB062971
D4S243
D4S413
D4S1626
AB055270
AB047864
NPY1R
FGA
IL15
D4S2374
D4S1636
CASP3
D4S1645
MMA5
AB046638
ADCY2
D5S208
AB048876
AB072779
ITGA2b
FST
AB055363
D5S111
AB060205
D5S1457
D5S1490
AB056382
D5S647
AB060843
HTR1A
CART
AB070012
AB056837
D5S1466
AB048971
AB055258
AB071134
AB072738
IL5
IL4
AB046007
PCDH1
D5S178
D5S658
D5S210
2B1
AB070025
AB047609
AB047945
D5S1458
D5S429
DRD1
D5S1470


MMA6
AB049869
PLG
AB069993
D6S404
D6S266
FRK
GJA1
D6S1608
D6S1690
AB055254
AB048916
AB062940
OPRM1
ESR1
AB047885
D6S311
GRM1
AB055326
AB051160
D6S1028
GRIK2
AB060249
AB060928
D6S501
D6S1056
CNR1
AB070040
AB056792
CNR1
KIAA0530
AB063014
AB048950
AB060863
AB049897
AB056810
D6S1718
D6S405
D6S1048
D6S402
AEGL1
D6S271
PGC
AB056774
MIC2
D6S291
AB060220
AB052142
MLN
TNF
POU5F1
D6S259
AB056329
AB070194
D6S1698
AB060213
D6S437
PRL
AB048924
MMA7
AB070018
TJP1
D15S165
CHRNA7
AB070168
TCF12
AB074450
TPM1
AB060208
D15S644
AB063027
AB060241
CYP1A2
AB060219
CHRNA3
D15S823
FAH
AB072736
D15S130
IGF1R
D14S261
ANG
D14S72
APEX
CMA1
NFKBIA
TITF1
AB056404
HNF3A
D14S306
PYGL
D14S66
AB069994
AB070000
AB070079
D14S277
PGF
PSEN1
AB052152
AB072774
AB066510
BDKRB1
1G4
TSHR
D14S67
GALC
CKB
AB062950
THBS1
D15S108
AB072754
AB055275
ESR2
AB060196
MMA8
DEFA1
AB047632
CHRNA2
AB052137
GNRH1
SFTPC
AB056822
D8S298
PNOC
AB060832
NRG1
D8S278
ADRB3
D8S268
AB050421
D8S285
D8S1119
AB047876
AB047861
AB050517
AB056782
AB062982
D8S266
AB047879
D8S1100
AB052185
MMA9
D2S114
AB048882
CXCR4
D2S1326
D2S150
D2S122
AB055304
D2S1399
AB063002
AB060913
D2S156
D2S382
SCN9A
CHRNA1
AB063030
AB049867
TFPI
COL3A1
D2S425
AB048954
AB046102
GPR1
AB070072
AB049879
AB056820
SCG2
D2S433
D2S1333
D2S1363
D2S163
COL4A3
AB060239
AB052203
AB046104
AB05040
ITGA6
GAD1
AB060898
2q
MMA10
D10S179
AB055302
AB070084
D10S197
D10S611
D10S1707
ITGB1
CREM
AB071127
MSMB
AB048993
AB060840
SFTPA1
D10S2327
D10S206
AB056408
PLAU
D10S1432
AB060846
MBL2
AB055370
ABCC2
D10S192
AB055379
AB050267
ADRB1
AB046045
AB055269
FGFR2
D10S1230
AB056389
AB070132
MMA11
D11S4162
D11S987
ADRBK1
RELA
AB072762
AB047597
CNTF
AGTRL1
CRY2
AB047906
RAG2
AB056372
D11S4200
D11S907
AB070065
FSHB
D11S904
AB048919
D11S1308
AB062963
D11S1329
AB046039
PDE2A
AB046041
AB047907
D11S527
D11S1902
AB072776
UCP2
D11S2011
D11S4079
D11S1789
D11S1352
D11S2002
AB047600
D11S4082
TYR
AB047860
MTNR1B
1G2
AB048900
MMP1
PGR
AB048899
NCAM1
D11S925
AB062966
D11S968
AB047831
MDK
D11S4203
D11S1338
HBB
D11S916
AB049001


MMA12
AB070109
D12S364
GRIN2B
KLRD1
AB071103
CD4
TPI1
D12S93
AB063024
AB050522
AB063071
AB060888
AB056334
AB047878
AB049870
AB060237
AB048933
D12S325
AB048980
D12S375
IFNG
LYZ
AB060827
AB055332
APAF1
D12S1030
D12S84
D12S86
AB046029
AB060197
AB060214
D12S342
D12S85
AB074458
12
MMA13
AB060857
AB072759
GALR3
D22S280
D22S685
D22S419
AB069998
D22S425
ADORA2A
AB056419
AB045996
D22S420
D20S116
D20S189
PLCB4
AB072772
D20S194
AB046072
AB060837
AB062957
GNRH2
AB055340
D20S112
AB052150
AB060251
CST3
D20S171
AB056405
AB063037
D20S100
BMP7
AB048917
SEMG2
ADA
PRNP
AB048884
22
20
20
MMA14
D9S152
AB070138
NTRK2
AB062937
AB046080
AB055267
AB047613
AB050415
D9S178
AB063044
AB071091
D9S285
D9S265
AB056422
AB060874
AB060228
GPR51
AB046634
D9S910
D9S938
AB070099
D9S1866
D9S261
AB070173
TLR4
D9S934
AB056817
D9S905
D9S260
DBH
D95922
TXN
9
9
9
9
MMA15
IL1B
AB070203
AB070186
D2S135
AB056773
SFTPB
AB046659
AB050253
D2S169
AB060885
TGFA
AB060826
NRXN1
D2S391
AB060200
D2S119
SLC8A1
AB060260
AB072771
AB055356
D2S146
D2S144
D2S1329
D2S149
ADAM17
D2S387
AB060821
AB049853
D2S391
2p
MMA16
F10
COL4A2
D13S286
AB049841
D13S167
AB056350
AB056328
AB062990
AB048874
AB056349
AB070178
AB056826
AB063055
D13S765
D13S894
OSF-2
FLT1
AB056393
AB052177
ALOX5AP
AB056393
AB056815
GJB2
HTR2A
13
MMA17
PAFAH1B1
ALOX15
AB071093
D17S796
AIPL1
AB071101
D17S804
D17S261
ADORA2B
AB056813
D17S1824
MMA2A11
SLC6A4
NF1
AB056819
MMA1A01
D17S1290
AB071111
AB050260
NGFR
D17S800
BRCA1
GFAP
AB070159
D17S791
AB055276
AB048897
KCNJ2
AB046065
TCF2
17
17
17
17
MMA18
AB046034
AB049906
AB060232
D18S537
D18S42
D18S1155
D18S849
AB056830
D18S74E
D18S536
AB056770
AB060918
AB062947
AB046105
AB072742
AB055344
DSG1
18
18
MMA19
AB070083
INSR
AB070080
D19S226
D19S385
D19S179
D19S714
AB060872
CCNE1
AB048905
GPI
AB047832
AB066540
AB074449
D19S178
LHB
AB062983
FUT1
KLK3
AB070118
AB046649
D19S180
19

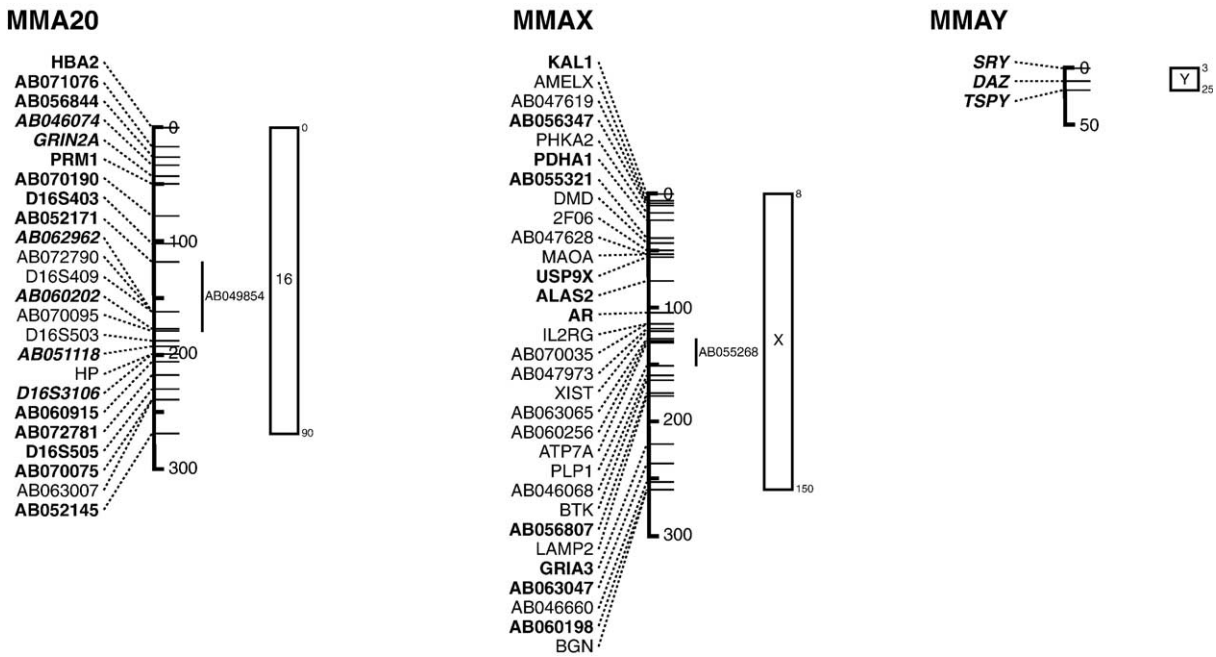


Fig. 1. Comparative maps between each macaque chromosome and conserved segments in the human genome, from NCBI Build 33 (right). Each chromosome subfigure displays the macaque STS markers (left), their positions in the macaque RH map (center), and blocks of conserved order with the human genome (right). The human chromosomes are identified by numbers inside of the blocks, and the orientation and order of the blocks are denoted by the end coordinates (given in megabase pairs). The scale on the macaque RH map is in centirays$_{5000}$. Markers placed with respect to the MLE-consensus map are positioned to the right of the RH map with vertical bars showing their most likely intervals. RH markers in boldface are those that are positioned in the MLE-consensus map with odds $\geq$ 1000:1; those in italics are positioned at odds $\geq$ 100:1.

breakpoints through comparison to other species and determined the smallest breakpoint interval using the boundaries of the mouse- (and rat-) conserved segment blocks (Fig. 2).

Table 1
Mapping summary of 1st generation macaque-human comparative map

Chromosome	Human homologues	Number of loci	Number of conserved segments	Min. # rearrangements
1	1	71	6	4
2	7&21	43	4	3
3	3	54	5	3
4	4	43	3	1
5	5	39	1	0
6	6	57	3	1
7	14&15	56	2	1
8	8	30	1	0
9	2q	38	1	0
10	10	32	3	1
11	11	55	2	1
12	12	34	1	0
13	20&22	36	3	2
14	9	33	4	3
15	2p	28	1	0
16	13	24	1	0
17	17	30	4	2
18	18	17	2	1
19	19	22	1	0
20	16	25	1	0
X	X	32	1	0
Y	Y	3	1	0
Total		802	51	23

Nearly all of the conserved blocks on human chromosomes 3, 7, 11, 17, and 18 are the result of recent rearrangements in the human lineage since divergence from a common ancestor with the rhesus macaque approximately 25 Mya ([4,12,18]; Supplemental Data). By contrast, internal rearrangements involving macaque chromosomes 4, 6, 10, and 14 most likely occurred during the evolutionary history of the Old World Monkey lineage. For some chromosomes the data and breakpoint refinement were not conclusive (MMA1, 13, 14) so we made no attempt to interpret these. Because the five chromosomes which were rearranged in the human lineage are able to be refined using the comparative data (notably the mouse and rat genome sequences) (Supplemental Data), we focused on these for our breakpoint region sequence analysis.

Combining these results with recent findings on single human chromosomes allows us to make general phylogenomic scenarios for human chromosomes [21]. Fig. 2 shows the 5 ancestral blocks of conserved order found in human, Bornean orangutan [15], macaque and cat (as an outgroup). One possible rearrangement scenario, consisting of three inversions, can convert macaque chromosome 3 into human chromosome 3. Using multiple species for comparison (Fig. 2), we refined the breakpoints boundaries on Hsa3 (human chromosome 3 [“Hsa3”] sequence coordinates) and examined their genomic content with the UCSC Genome Browser (April 2003 Build). Fig. 3 also shows a magnified view of one of the chromosome breakpoints at roughly 75–76 Mb. The breakpoint was characterized by a large number

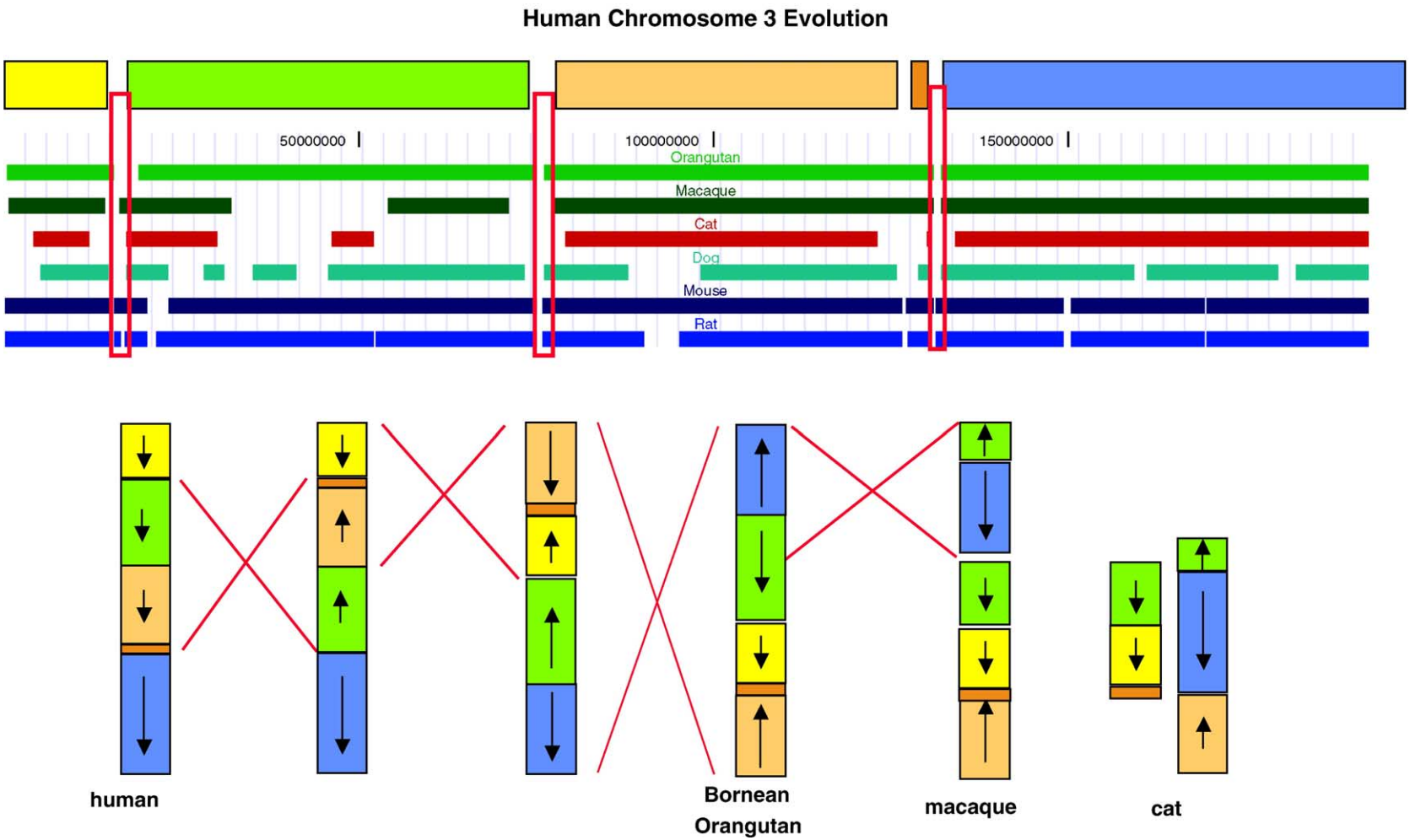


Fig. 2. A potential evolutionary rearrangement scenario for human chromosome 3. The top portion of the figure shows human chromosome 3 genome sequence coordinates (displayed horizontally), with colored blocks of conserved order for homologous chromosome segments in orangutan, rhesus macaque, cat, dog, mouse, and rat shown below the scale as custom track output from the UCSC genome browser. Shown above the human scale are five boxes of conserved order that represent the ancestral segments of Hsa3. Red vertical boxes define human-specific breakpoints. (Bottom) One possible rearrangement scenario converting the macaque chromosome into the human chromosome, estimated from the Genome Rearrangements Web Server [21]. The cat is shown as an outgroup taxon for reference. Note the second step in the scenario (from right to left) is not an actual rearrangment, but reverses the entire chromosome to be in the same orientation as the human genome.

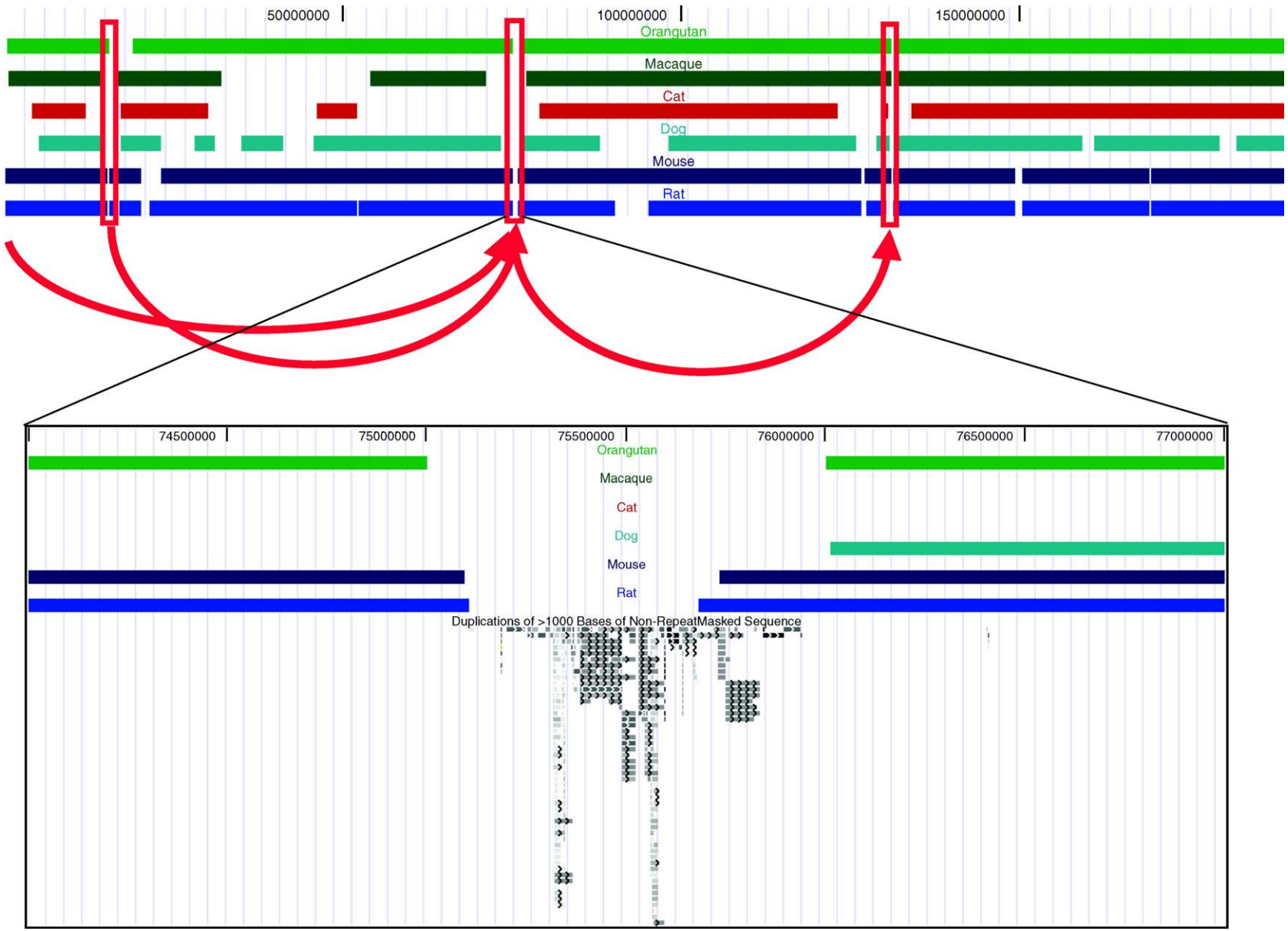


Fig. 3. A close-up view of one human-specific breakpoint region (boxed in red at ~75–78 Mb) on Hsa3 is shown to highlight the segmental duplications that populate the evolutionary breakpoint region (taken from the UCSC browser). The small gray boxes in the breakpoint interval (~71.4–71.6 Mb) each represent a block of sequence that is duplicated elsewhere in the human genome at least once. Red lines connect similarities between breakpoints indicating duplications flanking conserved segments.

of segmental duplications (>100 kb of total duplicated sequence). Some of these duplications are found at the other rearrangement breakpoints on Hsa3 with similar high densities of segmental duplications (Table 2).

Similar analysis of the other breakpoint regions on Hsa3 and other human chromosomes revealed a similar pattern (Table 2). For example, the analysis of human chromosome 11 identifies a single inversion that distinguishes it from its macaque homologue (Fig. 1). Focused analysis of the human-specific breakpoint regions with macaque, cat, dog, mouse, and rat genomes also shows a large percentage of segmental duplications at these breakpoints. Many of the duplicated copies map to the breakpoint that defines the opposite end of the conserved segment (Table 2). A comparable analysis of Hsa7 is consistent with the paracentric inversion breakpoints identified by Müller et al. [4] and their correlation with segmental duplications, while Hsa17 shows a similar pattern (Table 2).

If chromosome rearrangements are really facilitated by duplicate copies flanking an inverted segment, then the estimated divergence time of the duplicated sequences should correspond to or predate the origin of the lineage containing the rearranged chromosome. We tested this by calculating divergence times using the pairwise sequence divergence of duplicated segments (>10 kb) flanking the rearranged segments (Table 2). Applying this approach to breakpoint duplications on Hsa3 and Hsa11, the estimated duplication ages are roughly 12 Mya or greater for both sets of rearrangements on both chromosomes. This is consistent with the rearrangements occurring coincident with or after divergence of the human branch from orangutan roughly 12 Mya [22]. We note that one of the duplications on Hsa11 produces a divergence estimate of approximately 10 Mya, which is younger than the human–orangutan divergence. This inconsistency may be explained by either imprecision associated with the divergence estimation, or that this rearrangement event actually occurred after the human–orangutan divergence in the lineage leading to humans, chimpanzees, and gorillas. The breakpoint data from Hsa17 also supports a duplication driven rearrangement hypothesis for this chromosome, as two breakpoint pairs (36 + 45.9 and 36 + 60.7Mb) contain recent duplications in the 5–6 Mya range. While we cannot determine the exact timing of these rearrangements in the human lineage without additional primate genome maps, we can predict that these Hsa17 inversions occurred sometime after the orangutan–human divergence (<12 Mya), and possibly (given the divergence estimate) after the human/chimp divergence from gorilla 7–8 Mya.

We further applied our approach to previously reported duplications flanking rearrangement breakpoints on other human chromosomes. For the duplication pair on human chromosome 18, previously hypothesized to have facilitated the inversion that distinguishes humans from all other primates and mammals examined so far [12], we estimate that the duplication occurred roughly 3 Mya (Table 2), consistent with the inversion occurring after the human–chimpanzee divergence 5–6 Mya [18]. We also examined seven duplications in the breakpoints identified by Müller et al. [4] for the paracentric inversion in human 7 that distinguishes orangutan and macaque from humans and great apes (Table 2). We find that six of the duplications predate the human–gorilla divergence (7–8 Mya), suggesting that they accumulated prior to the inversion that occurred in the human/chimpanzee lineage and may have facilitated the rearrangement [4].

Discussion

We have presented the first ordered whole genome map of the rhesus macaque based on 802 markers distributed across all chromosomes. This map serves as a first step toward future mapping and sequencing efforts, to aid in long-range sequence scaffold assembly and interpretation of the detailed syntenic conservation between human and macaque genomes. Our results support the majority of rearrangements observed in an ordered autosome comparison of baboon and human [10], confirming a strong degree of genomic conservation between baboon and macaque at the G-banding level [19], while refining many of the breakpoint boundaries.

By comparison to other mammalian genome maps we were able to identify those rearrangements that have occurred during the past 25 million years of human evolution. These comparisons revealed a striking correspondence between evolutionary breakpoint regions and accumulation of segmental duplications. In fact, all of the breakpoint regions that occurred in the human lineage since it diverged from Old World monkeys contained $\geq$10 kb of duplicated sequence. The idea that duplications might have provided the substrate for nonallelic homologous recombination has been recently proposed for human chromosomes 7 and 18 [4,12] and other primate chromosome rearrangements [16]. Our findings further extend this observation to nearly all recent, large-scale chromosome inversions in the human genome that occurred in the last 25 million years.

Two previous studies explicitly examined the association between evolutionary breakpoints and the presence of segmental duplications in human and mouse genome sequences [23,24]. Both studies concluded that a far greater number of human–mouse evolutionary chromosomal breakpoints contain duplications than would be expected by chance. This observation led Armengol et al. [23] to suggest a role for segmental duplications in chromosomal rearrangement.

Most segmental duplications in the human genome likely arose within the past 35–40 million years during the primate radiation [22,25]. Because the vast majority of the human–mouse synteny blocks are due to chromosomal rearrangements occurring in the mouse/rat lineage, primate-

Table 2
Paired Segmental Duplication (SD) content at human-specific inversion breakpoints

Human chromosome	Breakpoint A (Mbp)[a]	Breakpoint B (Mbp)[a]	Number of paired SDs (>10kb)	Size (bp)	Coordinates of duplicate pair(Mb)	% identity approx.	Duplication date (Mya)[b]
3	0–64,290	75,057,848–75,705,918	1	27,778	63,001–90,885/75,100,445–75,128,222	96.2	12
3	15,154,959–15,236,968	75,057,848–75,705,918	1	13,373	15,166,391–15,179,763/75,392,638–75,406,067	95.9	12
3	75,057,848–75,705,918	130,896,841–131,213,246	5	31,535	75,431,032–75,463,517/130,999,522–131,031,056	94.3	23
	"	"		44,870	75,205,102–75,249,876/131,014,803–131,059,672	94.4	23
	"	"		21,076	75,252,950–75,268,978/131,060,015–131,081,090	95.8	13
	"	"		26,832	75,269,082–75,295,440/131,090,674–131,117,505	95.5	15
	"	"		79,396	75,305,422–75,374,646/131,117,505–131,196,900	93.9	≥25
7	76,173,866/76,305,767	101,830,259–101,900,313	7	106,202	76,294,363–76,400,000/101,885,615–101,991,816	98.6	4
	"	"		17,077	76,240,904–76,257,980/101,523,375–101,541,135	93.6	≥25
	"	"		11,957	76,246,024–76,257,980/101,841,240–101,853,838	93.9	≥25
	"	"		11,965	76,246,024–76,257,988/101,742,166–101,754,791	93.9	≥25
	"	"		21,109	76,268,968–76,290,076/101,523,375–101,544,790	96.7	9
	"	"		18,669	76,273,944–76,292,612/101,736,006–101,754,791	97.8	6
	"	"		18,669	76,273,944–76,292,612/101,835,161–101,853,838	97.8	6
11	3,148,064–3,600,022	71,452,377–71,849,730	3	97,812	3,439,425–3,534,281/71,600,005–71,697,816	96.6	10
	"	"		39,959	3,385,220–3,424,992/71,698,563–71,738,521	95.2	16
	"	"		10,209	3,572,940–3,583,148/71,825,175–71,835,336	94.5	22
17	35,929,570–36,252,395	45,859,974–45,902,536	1	105,843	36,274,617–36,381,204/45,894,158–46,000,000	98.3	5
17	35,929,570–36,252,395	60,668,108–60,759,220	5	28,218	36,013,200–36,041,643/60,668,658–60,696,875	97.9	6
	"	"		15,535	36,041,346–36,057,952/60,639,154–60,654,688	97.8	6
	"	"		14,861	36,066,753–36,081,569/60,681,791–60,696,651	98.0	6
	"	"		24,764	36,186,043–36,211,013/60,672,112–60,696,875	97.8	6
	"	"		18,726	36,242,375–36,261,026/60,677,964–60,696,689	98.2	5
18	p-terminus	18.4–18.5	1	19,469	102,547–121,692/18,408,332–18,427,800	99.0	3

[a] Estimated from rightmost and leftmost boundaries of adjacent mouse synteny blocks flanking the breakpoint region (April 2003 Build).
[b] Estimated using data from Ref. [22]: see methods.

specific segmental duplications would not be predicted to provide the substrate for rearrangement in these murid rodent breakpoints [24]. Therefore there would be no expectation to observe duplications at the same sites of mouse-specific breakpoints. However, these two concepts are not mutually exclusive. While many fragile regions may exist [26,27], and these may be prone to accumulating segmental duplications, only those breakpoint regions that promote a chromosomal rearrangement should fit the additional criterion of having a duplicate sequence found at the ends of each rearranged segment [12]. Here we show that this is supported in the majority of the cases where rearrangements could be reliably assigned to the human lineage, and are further supported by molecular dating of duplication events coincident with or preceding the phylogenetic age of the rearrangement. The recent completion of the chimp genome sequence and future acquisition of other primate genome sequences will allow this observation to be more broadly tested in the context of duplications unique to additional primate lineages.

Methods

Primer design

Human microsatellite primers (MapPairs) were obtained from Research Genetics based on those markers optimized by Rogers et al. [10]. Coding sequence primers were designed in either 3′-untranslated regions of macaque ESTs and mRNAs derived from GenBank. All primers were designed with Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi).

Radiation hybrid genotyping

Methods for genotyping have been described in detail elsewhere [17,28]. Briefly, PCR-based typing was performed on the 5000-rad rhesus macaque whole genome radiation hybrid panel [17]. All STSs were optimized using either a standard PCR protocol (10 min hot-start at 95°C followed by 35 cycles of 95°C for 15 s, 58–62°C for 15 s, 72°C for 30 s) or a touchdown PCR protocol that lowered the annealing temperature from 60°C to 50°C by 2°C over 10 cycles. Optimized markers were then genotyped in the 93-clone RH panel, prealiquoted into 384 well plates, using Taq-Gold DNA polymerase (Applied Biosystems Inc.) and scored in duplicate on 2% agarose gels stained with VistaGreen (Amersham).

Microsatellite library construction and screening

A $(dG{\cdot}dT)_n$ $(dC{\cdot}dA)_n$-enriched microsatellite library was constructed from DNA of a male rhesus macaque and recombinants were sequenced as previously described [29]. A software tool was developed for processing of the sequences of recombinants which removed vector and linker sequences, discarded clones with fewer than 10 uninterrupted dinucleotide repeats, and designed PCR primers using Primer 3.0.

Map construction

Macaque STS-derived RH genotyping vectors were assigned to chromosomes based on 1:1 syntenic conservation of the human and macaque genomes. We then tested these groups using the program RH2PT in the software package RHMAP [30] and previous syntenic predictions. An initial LOD threshold of 12.0 was used and then subsequent intrachromosome groups were assembled using decreasing LOD thresholds, with 8.0 being the minimum LOD threshold used for any group. For ordering, chromosomes were broken down into two groups corresponding to the arms of all metacentric and submetacentric chromosomes, or the entire chromosome of acrocentric elements. In some cases shifts in retention frequency required multiple groups/arm. Markers within each chromosome arm or linkage group were ordered using a reduction from the problem of RH mapping to the traveling salesman problem (TSP) [31], as implemented in the software rh_tsp_map [32].

The computations to construct the radiation hybrid map were done using programs from the rh_tsp_map/CONCORDE [32,33] and RHMAP [30] software packages. Given an input set of markers, rh_tsp_map transforms the problem of finding the MLE ordering to an instance of the well-studied traveling salesman problem (TSP). Using the CONCORDE package for TSP along with QSopt optimization package (http://www.isye.gatech.edu/~wcook/qsopt), all the TSP instances we generated were solved to guaranteed optimality. This approach assumes that input linkage groups are defined correctly, and we took many precautions, as described below, to use only markers and linkage groups that could be assigned with high confidence. As noted previously, in a few regions (such as the Ig heavy-chain locus region) where retention frequency is highly variable, we used the RHMAXLIK option of RHMAP to check that it gives the same order as rh_tsp_map/CONCORDE. This checking was done because RHMAP models variable retention frequency, although it does not guarantee an optimal map with respect to its model. Following Agarwala et al. [32], we used three definitions of the maximum likelihood, base_mle, extended_mle, and normalized_mle, that differ in their treatment of “2” entries in the vector data for defining the likelihood. During preliminary attempts at map construction, we used the merge, map_eval and flips programs of rh_tsp_map to identify linkage groups for which the optimal orders differed among the three MLE criteria, the placement of one marker was not optimal with respect to the map defined by the remaining markers, or the two best orders differed by LOD < 0.1. In such cases, we dropped problematic markers, and recomputed optimal marker orders until the two consistency criteria were met.

We used several tests to drop or set aside markers for placement with respect to the MLE-consensus map because they could not be reliably ordered as part of consensus map construction itself. These methods were applied multiple times iteratively, until all markers passed each test of quality. The first test of quality assumed that a good marker (vector) assigned to one chromosome should not have high 2-point LOD scores with markers assigned to a different chromosome [34]. Specifically, we dropped entirely any marker that had 2-point LOD scores > 15.0 with at least two markers assigned to (a) different chromosome(s). The second test of quality relates to the assumption of equal retention frequency in rh_tsp_map. We identified any markers whose retention frequency was >0.2 lower than the average for its chromosome and/or could not be reliably ordered when analyzed with 2–4 nearby markers. The markers that failed the second quality test were reserved for subsequent placement in the best intermarker interval(s) of the final MLE-consensus map. In the course of applying the first test for high interchromosome scores, we observed that many markers with isolated high scores to another marker on another chromosome had high retention frequency. As a third test of reliability we dropped any marker with rf > 0.48, except those on chromosomes Mma17 and MmaY because these two chromosomes had high average rf. The 0.48 threshold was sufficient to eliminate most of the remaining high interchromosome scores. The fourth test of quality may be considered a dual of the first test: a marker should not have low 2-point LOD scores when compared to markers predicted by synteny to be on the same chromosome and nearby. Specifically, we dropped markers that had no 2-point scores > 3.0. The fifth test of marker quality had three parts all related to the reliability of marker ordering. The first part was that the three definitions of the MLE criterion should yield the same marker order. The second part is that if any marker is temporarily deleted and then placed back into the map relative to the remaining marker, the new placement should agree with the initial position of the marker. The third part is that no window of 7 consecutive markers can be permuted so as to yield a better log-likelihood or even a log-likelihood within 0.1 of the best. For preliminary maps where any of these reliability tests failed, we tried dropping different markers until we identified the problematic one(s). Markers identified as problematic by either the third, fourth, or fifth test were deferred for placement.

When constructing consensus maps, we excluded any marker with too many "2" (ambiguous) vector entries, to reduce the chance that the different MLE definitions will lead to different optimal orders. To reduce the chance of having different orders with similar likelihoods and regions of local uncertainty, rh_tsp_map does not use one of a pair of markers if they are too close to each other. In this study, we excluded vectors with more than two ambiguities and considered two markers too close if either their Hamming distance or their likelihood-derived distance was <3.0. Markers omitted either due to too many 2's or to being too close to another marker were set aside for later placement with respect to the MLE-consensus map. Those markers that were too close were mapped to their most likely positions in RHMAXLIK after enforcing a marker order constraint for the MLE consensus loci.

Comparative analyses

For each macaque locus, human physical positions in the draft human sequence are based upon BLAST (http://www.ncbi.nlm.nih.gov/blast/) analysis to the April 2003 human genome assembly (used because segmental duplication data were only available for this build at the time of analysis). Comparative mapping data from the cat genome is from Menotti-Raymond et al. [35]. The dog comparisons were made to RH maps from Guyon et al. [36] and Andelfinger et al. [37]. Orangutan data is from Tsend-Ayush [15]. Mouse and rat conserved segments are from Bourque et al. [38]. The Genome Rearrangements server (http://nbcr.sdsc.edu/GRIMM/mgr.cgi [21]) was used to obtain the minimum number of rearrangements needed to transform pieces of the macaque genome into the corresponding pieces of the human genome, and to identify at least some scenarios that achieve those lower bounds. In all cases signed data on marker order were used.

Divergence estimates for paired segmental duplications were obtained using their percentage nucleotide sequence identity (based on Jukes-Cantor distances in the UCSC browser). These percentages were converted to time (in million years) using the percent identity/divergence time estimates given in Ref. [22], which assume a neutral rate of evolution.

Acknowledgments

We thank Mary Carrington, Michael Dean, Jillana Minchoff, Jason Morris, Kevin Scott, Jennifer Tabler, Emma Teeling, Anne Truelove, and Michael Smith for technical assistance and/or insightful discussion on these topics. We thank William Cook for assistance in integrating the rh_tsp_map distance functions into CONCORDE and for making available early versions of QSopt.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2005.05.013.

References

[1] E.E. Eichler, P.J. de Jong, Biomedical applications and studies of molecular evolution: A proposal for a primate genomic library resource, Genome Res. 12 (2002) 673–678.

[2] J.J. Yunis, O. Prakash, The origins of man: A chromosomal pictoral legacy, Science 215 (1982) 1525–1530.

[3] S. Müller, R. Stanyon, P. Finelli, N. Archidiacono, J. Wienberg, Molecular cytogenetic dissection of human chromosomes 3 and 21 evolution, Proc. Natl. Acad. Sci.USA 97 (2000) 206–211.
[4] S. Müller, P. Finelli, M. Neusser, J. Wienberg, The evolutionary history of human chromosome 7, Genomics 84 (2004) 458–467.
[5] D. Haig, A brief history of human autosomes, Philos. Trans. R. Soc. Lond. B: Biol. Sci. 354 (1999) 1447–1470.
[6] S.J. O'Brien, R. Stanyon, Phylogenomics. Ancestral primate viewed, Nature 402 (1999) 365–366.
[7] J. Wienberg, L. Froenicke, R. Stanyon, Insights into mammalian genome organization and evolution by molecular cytogenetics, in: M.S. Clark (Ed.), Comparative Genomics, Kluwer Academic, 2000, pp. 207–244.
[8] R. Stanyon, U. Koehler, S. Consigliere, Chromosome painting reveals that galagos have highly derived karyotypes, Am. J. Phys. Anthropol. 117 (2002) 319–326.
[9] F. Bigoni, R. Stanyon, R. Wimmer, W. Schempp, Chromosome painting shows that the proboscid monkey (*Nasalis larvatus*) has a derived karyotype and is phylogenetically nested within the Asian Colobines, Am. J. Primatol. 60 (2003) 85–93.
[10] J. Rogers, M.C. Mahaney, S.M. Witte, S. Nair, D. Newman, S. Wedel, L.A. Rodriguez, K.S. Rice, S.H. Slifer, A. Perelygin, et al., A genetic linkage map of the baboon (*Papio hamadryas*) genome based on human microsatellite polymorphisms, Genomics 67 (2000) 237–247.
[11] L. Carbone, M. Ventura, S. Tempesta, M. Rocchi, N. Archidiacono, Evolutionary history of chromosome 10 in primates, Chromosoma 111 (2002) 267–272.
[12] B.K. Dennehey, D.G. Gutches, E.H. McConkey, K.S. Krauter, Inversion, duplication, and changes in gene context are associated with human chromosome 18 evolution, Genomics 83 (2003) 493–501.
[13] V. Eder, M. Ventura, M. Ianigro, M. Teti, M. Rocchi, N. Archidiacono, Chromosome 6 phylogeny in primates and centromere repositioning, Mol. Biol. Evol. 20 (2003) 1506–1512.
[14] A. Ruiz-Herrera, F. García, C. Azzalin, E. Giulotto, J. Egozcue, M. Ponsà, M. Garcia, Distribution of intrachromosomal telomeric sequences (ITS) on *Macaca fascicularis* (Primates) chromosomes and their implication for chromosome evolution, Hum. Genet. 110 (2002) 578–586.
[15] E. Tsend-Ayush, F. Grützner, Y. Yue, B. Grossmann, U. Hänsel, R. Sudbrak, T. Haaf, Plasticity of human chromosome 3 evolution during primate evolution, Genomics 83 (2004) 193–202.
[16] D.P. Locke, N. Archidiacono, D. Misceo, M. Francesca Cardone, S. Deschamps, B. Roe, M. Rocchi, E.E. Eichler, Refinement of a chimpanzee pericentric inversion breakpoint to a segmental duplication cluster, Genome Biol. 4 (2003) R50.
[17] W.J. Murphy, J. Page, C. Smith Jr., R.C. Desrosiers, S.J. O'Brien, A radiation hybrid mapping panel for the rhesus macaque, J. Hered. 92 (2001) 516–519.
[18] E. Eizirik, W.J. Murphy, M.S. Springer, S.J. O'Brien, Molecular phylogeny and dating of early primate divergences, in: R. Kay, C. Ross (Eds.), Anthropoid Origins, New Visions, Kluwer/Plenum, 2004, pp. 45–64.
[19] C.M. Moore, C. Janish, C.A. Eddy, G.B. Hubbard, M.M. Leland, J. Rogers, Cytogenetic and fertility studies of a rheboon, rhesus macaque (*Macaca mulatta*) x baboon (*Papio hamadryas*) cross: Further support for a single karyotype nomenclature, Am. J. Phys. Anthropol. 110 (1999) 119–127.
[20] J. Wienberg, R. Stanyon, A. Jauch, T. Cremer, Homologies in human and Macaca fuscata chromosomes revealed by in situ suppression hybridization with human chromosome specific DNA libraries, Chromosoma 101 (1992) 265–270.
[21] G. Tesler, GRIMM: Genome rearrangements web server, Bioinformatics 18 (2002) 492–493.
[22] R. Vallente Samonte, E.E. Eichler, Segmental duplications and the evolution of the primate genome, Nat. Rev. Genet. 3 (2002) 65–72.
[23] L. Armengol, M.A. Pujana, J. Cheung, S.W. Scherer, X. Estivil, Enrichment of segmental duplications in regions of breaks of synteny between the human and mouse genomes suggest their involvement in evolutionary rearrangements, Hum. Mol. Genet. 12 (2003) 2201–2208.
[24] J.A. Bailey, R. Baertsch, W.J. Kent, D. Haussler, E.E. Eichler, Hotspots of mammalian chromosomal evolution, Genome Biol. 5 (2004) R23.
[25] J.A. Bailey, Z. Gu, R.A. Clark, K. Reinert, R.V. Samonte, S. Schwartz, M.D. Adams, E.W. Myers, P.W. Li, E.E. Eichler, Recent segmental duplications in the human genome, Science 297 (2002) 1003–1007.
[26] P. Pevzner, G. Tesler, Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution, Proc. Natl. Acad. Sci. USA 100 (2003) 7672–7677.
[27] W.J. Kent, R. Baertsch, A. Hinrichs, W. Miller, D. Haussler, Evolution's cauldron: Duplication, deletion, and rearrangement in the mouse and human genomes, Proc. Natl. Acad. Sci. USA 100 (2003) 11484–11489.
[28] W.J. Murphy, S. Sun, Z.-Q. Chen, N. Yuhki, D. Hirschmann, M. Menotti-Raymond, S.J. O'Brien, A radiation hybrid map of the cat genome: implications for comparative mapping, Genome Res. 10 (2000) 691–702.
[29] R. Sarno, V.A. David, W.L. Franklin, S.J. O'Brien, W.E. Johnson, Development of microsatellite markers in the guanaco, Lama guanicoe: utility for South American camelids, Mol. Ecol. 9 (2000) 1922–1924.
[30] M. Boehnke, K. Lange, D.R. Cox, Statistical methods for multipoint radiation hybrid mapping, Am. J. Hum. Gen. 49 (1991) 1174–1188.
[31] A. Ben-Dor, B. Chor, On constructing radiation hybrid maps, J. Comput. Biol. 4 (1997) 517–533.
[32] R. Agarwala, D.L. Applegate, D. Maglott, G.D. Schuler, A.A. Schäffer, A fast and scalable radiation hybrid map construction and integration strategy, Genome Res. 10 (2000) 350–364.
[33] D. Applegate, R. Bixby, V. Chvatal, W. Cook, On the solution of traveling salesman problems. Documenta mathematica, extra volume International Congress of Mathematics III (1998) 645–656.
[34] W. Van Etten, R.G. Steen, H. Nguyen, A.B. Castle, D.K. Slonim, B. Ge, C. Nusbaum, G.D. Schuler, E.S. Lander, T.J. Hudson, Radiation hybrid map of the mouse genome, Nat. Genet. 22 (1999) 384–387.
[35] M. Menotti-Raymond, V.A. David, Z.Q. Chen, K.A. Menotti, S. Sun, A.A. Schäffer, R. Agarwala, J.F. Tomlin, S.J. O'Brien, W.J. Murphy, Second-generation integrated genetic linkage/radiation hybrid maps of the domestic cat (*Felis catus*), J. Hered. 94 (2003) 95–104.
[36] R. Guyon, T.D. Lorentzen, C. Hitte, L. Kim, E. Cadieu, H.G. Parker, P. Quignon, J.K. Lowe, C. Renier, B. Gelfenbeyn, et al., A 1-Mb resolution radiation hybrid map of the dog genome, Proc. Natl. Acad. Sci. USA 100 (2003) 5296–5301.
[37] G. Andelfinger, C. Hitte, L. Etter, R. Guyon, G. Bourque, G. Tesler, P. Pevzner, E. Kirkness, F. Galibert, D.W. Benson, Detailed four-way comparative mapping and gene order analysis of the canine *ctvm* locus reveals evolutionary chromosome rearrangements, Genomics 83 (2004) 1053–1062.
[38] G. Bourque, P.A. Pevzner, G. Tesler, Reconstructing the genomic architecture of ancestral mammals: lessons from human, mouse, and rat genomes, Genome Res. 14 (2004) 507–516.