Inversion variants in human and primate genomes

Claudia Rita Catacchio,[1,4] Flavia Angela Maria Maggiolini,[1,4] Pietro D'Addabbo,[1]
Miriana Bitonto,[1] Oronzo Capozzi,[1] Martina Lepore Signorile,[1] Mattia Miroballo,[1]
Nicoletta Archidiacono,[1] Evan E. Eichler,[2,3] Mario Ventura,[1] and Francesca Antonacci[1]

[1]Dipartimento di Biologia, Università degli Studi di Bari "Aldo Moro," Bari 70125, Italy; [2]Department of Genome Sciences, University of Washington School of Medicine, Seattle, Washington 98195, USA; [3]Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA

For many years, inversions have been proposed to be a direct driving force in speciation since they suppress recombination when heterozygous. Inversions are the most common large-scale differences among humans and great apes. Nevertheless, they represent large events easily distinguishable by classical cytogenetics, whose resolution, however, is limited. Here, we performed a genome-wide comparison between human, great ape, and macaque genomes using the net alignments for the most recent releases of genome assemblies. We identified a total of 156 putative inversions, between 103 kb and 91 Mb, corresponding to 136 human loci. Combining literature, sequence, and experimental analyses, we analyzed 109 of these loci and found 67 regions inverted in one or multiple primates, including 28 newly identified inversions. These events overlap with 81 human genes at their breakpoints, and seven correspond to sites of recurrent rearrangements associated with human disease. This work doubles the number of validated primate inversions larger than 100 kb, beyond what was previously documented. We identified 74 sites of errors, where the sequence has been assembled in the wrong orientation, in the reference genomes analyzed. Our data serve two purposes: First, we generated a map of evolutionary inversions in these genomes representing a resource for interrogating differences among these species at a functional level; second, we provide a list of misassembled regions in these primate genomes, involving over 300 Mb of DNA and 1978 human genes. Accurately annotating these regions in the genome references has immediate applications for evolutionary and biomedical studies on primates.

[Supplemental material is available for this article.]

A long-standing question in evolutionary biology concerns the effect of inversions in shaping the genomic architecture of organisms. The most conspicuous differences between the human and chimpanzee karyotypes are nine pericentric inversions, suggesting that inversions occur quite frequently in primate chromosomal evolution (Yunis and Prakash 1982; Ventura et al. 2001, 2003, 2004, 2007, 2011; Carbone et al. 2002; Eder et al. 2003; Misceo et al. 2003, 2005; Kehrer-Sawatzki et al. 2005a,b; Cardone et al. 2006, 2007, 2008; Kehrer-Sawatzki and Cooper 2008; Stanyon et al. 2008; Capozzi et al. 2012). The main evolutionary importance of inversions comes from the fact that they suppress recombination in heterokaryotypes (Sturtevant 1917). As a consequence, inverted and noninverted segments can follow distinct evolutionary histories and accumulate variation independently, creating a genetic barrier to gene flux and contributing to speciation (Navarro et al. 1997; Farre et al. 2013). In addition to single-nucleotide variation, the two haplotypes can also harbor different segmental duplications (duplicated sequences ≥1 kb in length and showing ≥90% sequence identity) (Bailey et al. 2001). Differences in segmental duplication architecture can predispose one of the haplotypes to nonallelic homologous recombination (NAHR) leading to additional putative pathogenic rearrangements in subsequent generations (Lupski 1998). The best-known example of this effect in the human genome is the 900-kb polymorphic inversion on 17q21.31 (Stefansson et al. 2005). This locus occurs as two haplotypes, in direct and inverted orientation, not recombining over nearly 2 Mb, resulting in extended linkage disequilibrium (Skipper et al. 2004). The two haplotypes have different functional impacts: The direct haplotype is associated with neurological disorders such as Parkinson's disease (Tobin et al. 2008), while the inverted haplotype is enriched in European populations and predisposes to the 17q21.31 microdeletion in subsequent generations, as a result of NAHR between homologous segmental duplications (Koolen et al. 2008; Zody et al. 2008; Steinberg et al. 2012).

Despite their importance in human disease and genome evolution, inversions represent relatively unexplored forms of structural variation mainly due to the lack of high-throughput techniques for detecting them. Most inversions described to date between human and nonhuman primate genomes result from laborious and target-based chromosomal studies in cytogenetics that led to the identification of several large inversion variants (Ventura et al. 2001, 2003, 2004, 2007, 2011; Carbone et al. 2002; Eder et al. 2003; Misceo et al. 2003, 2005; Goidts et al. 2004; Kehrer-Sawatzki et al. 2005a,b; Cardone et al. 2006, 2007, 2008; Kehrer-Sawatzki and Cooper 2008; Stanyon et al. 2008; Capozzi et al. 2012).

A major breakthrough in the discovery of inversions came with the introduction of paired-end sequencing and mapping

910 **Genome Research**
www.genome.org
28:910–920 Published by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/18; www.genome.org

(Newman et al. 2005; Tuzun et al. 2005; Kidd et al. 2008). Nevertheless, a huge limitation of this method is related to the genome architecture associated with inversions. The majority of the inversions described in the human genome to date are flanked by high-identity segmental duplications (Kidd et al. 2008; Sanders et al. 2016) or inverted repeats (Vicente-Salvador et al. 2017), causing problems for inversion discovery using paired-end mapping (Alkan et al. 2011).

Exploiting new technologies for DNA sequencing, researchers have significantly improved the reference genome assemblies for a number of primates in the last decade (The Chimpanzee Sequencing and Analysis Consortium 2005; Rhesus Macaque Genome Sequencing and Analysis Consortium 2007; Locke et al. 2011; Scally et al. 2012; Gordon et al. 2016). Access to these genome sequences has increased the ability to carry out basic comparative and structural genomics analyses in these species. For instance, Feuk et al. (2005) identified 1576 submicroscopic inversions at a genome-wide level through computational comparisons of genome sequences between human and chimpanzee. However, only 27 predicted inversions were experimentally validated to distinguish real inversions from false positives. In this study, we took advantage of the recent reference genome releases and compared the net alignments for the most recent builds of human and nonhuman primate genome assemblies, including chimpanzee, gorilla, orangutan, and macaque genomes. Given the extensive karyotypic diversity compared to the other apes, gibbons were excluded from the analysis (Jauch et al. 1992; Muller et al. 2003; Capozzi et al. 2012; Carbone et al. 2014). Our study shows that comparison of independently assembled primate genomes with high-quality sequence is a good alternative to overcome some of the limitations of paired-end mapping. We identified, validated, and studied the evolutionary history of genomic inversions, and discovered regions that are misassembled in one or more reference genomes. Our study emphasizes the importance of improving the quality of primate assemblies to the current level of the human reference in order to facilitate additional comparative analyses and to fully enable the use of these organisms in biomedical research.

Results

Sequence alignments between human and primate genome assemblies

Net alignments between human (GRCh38/hg38) (Schneider et al. 2017), chimpanzee (Pan_tro 3.0/panTro5) (Kuderna et al. 2017), gorilla (gorGor4.1/gorGor4) (Scally et al. 2012), orangutan (WUGSC 2.0.2/ponAbe2) (Locke et al. 2011), and macaque (BCM Mmul_8.0.1/rheMac8) (Zimin et al. 2014) genomes were downloaded from the UCSC Genome Browser (University of California, Santa Cruz; http://genome.ucsc.edu/) and filtered for those longer than 100 kb and containing <90% of repeats or segmental duplications. After filtering, 38 alignments in inverted orientation were identified between human and chimpanzee, 28 between human and gorilla, 36 between human and orangutan, and 112 between human and macaque genomes (Supplemental Table S1). Multiple flanking alignments were manually inspected and merged into a single inversion call (Supplemental Table S2). After merging, 136 regions (156 inversion calls) were identified as potentially inverted in one or more species. These include 18 regions inverted between human and chimpanzee, 11 between human and gorilla, 13 between human and orangutan, and 78

between human and macaque. Sixteen regions were found to be inverted between human and more than one species: one in common between chimpanzee and orangutan; two in common between gorilla and orangutan; nine in common between orangutan and macaque; one in common between chimpanzee and macaque; one in common between chimpanzee, orangutan, and macaque; one in common between gorilla, orangutan, and macaque; and one in common between all four primates. These inversions range in size from 103 kb to 91 Mb and are distributed throughout all autosomes, with the highest number mapping on Chromosome 7.

Inversion maps of large known genomic inversions

All previously reported inversions larger than 5 Mb were used to draw ideograms for each chromosome in all the species analyzed (Fig. 1; Supplemental Fig. S1; Ventura et al. 2007, 2011; Cardone et al. 2008). Each colored block (synteny block) represents a region that is inverted in at least one of the species analyzed, but the order of the markers within the block is conserved in all the different species. Previous comparative studies focused on autosomal variants only, and the sex chromosomes were neglected. As a consequence, we excluded the X and Y Chromosomes from our analysis. We uploaded the inversion calls obtained from the net alignments comparison on the UCSC Genome Browser and highlighted the synteny blocks with the same color used to generate the ideograms (Fig. 1; Supplemental Fig. S1). This information was used in order to understand the relative orientation of the predicted inversions with respect to larger cytogenetic inversions.

Inversion validations

First, we compared our inversion calls to previously reported inversions (Feuk et al. 2005; Ventura et al. 2007, 2011; Cardone et al. 2008; Antonacci et al. 2009; Nuttle et al. 2016) and confirmed the inverted orientation of 39 events (Fig. 1; Supplemental Fig. S1; Supplemental Table S3). These correspond to the large known inversions shown by the green arrows in the chromosome ideograms in Figure 1A and Supplemental Figure S1 and that were also reidentified in the current study. Then, we tested 30 inversions using fluorescence in situ hybridization (FISH) in multiple human HapMap cell lines and primate species cell lines where the regions were called to be inverted (Supplemental Tables S3, S4). Owing to limitations in resolution, FISH analysis allowed us to validate only inversions larger than 500 kb. In particular, we used interphase three-color FISH for inversions between 500 kb and 2 Mb in size and metaphase two-color FISH for inversions larger than 2 Mb (Supplemental Table S5; Supplemental Fig. S2). Testing the inversion in multiple individuals allowed us to investigate if the inversion was polymorphic in the species analyzed. However, we only tested two individuals (four chromosomes) per primate species, and therefore, we were unable to define if an inversion was polymorphic for allele frequencies lower than 25%. FISH experiments showed that 16 calls were not inverted and therefore represent an orientation error in the assembly of the primate genomes analyzed, while 19 inversion calls were confirmed and were further studied in order to understand the evolutionary history of the inversion events (see "Evolutionary analyses"). One region of 846 kb on human Chromosome 1 (Chr1_inv3, GRCh38/hg38 Chr 1: 147,079,442–147,925,603), called to be inverted in chimpanzee and orangutan (calls hg38_panTro5_22 and hg38_ponAbe2_22), was confirmed as inverted by FISH in 10 human cell lines compared to the reference genome (GRCh38/hg38), suggesting that this is
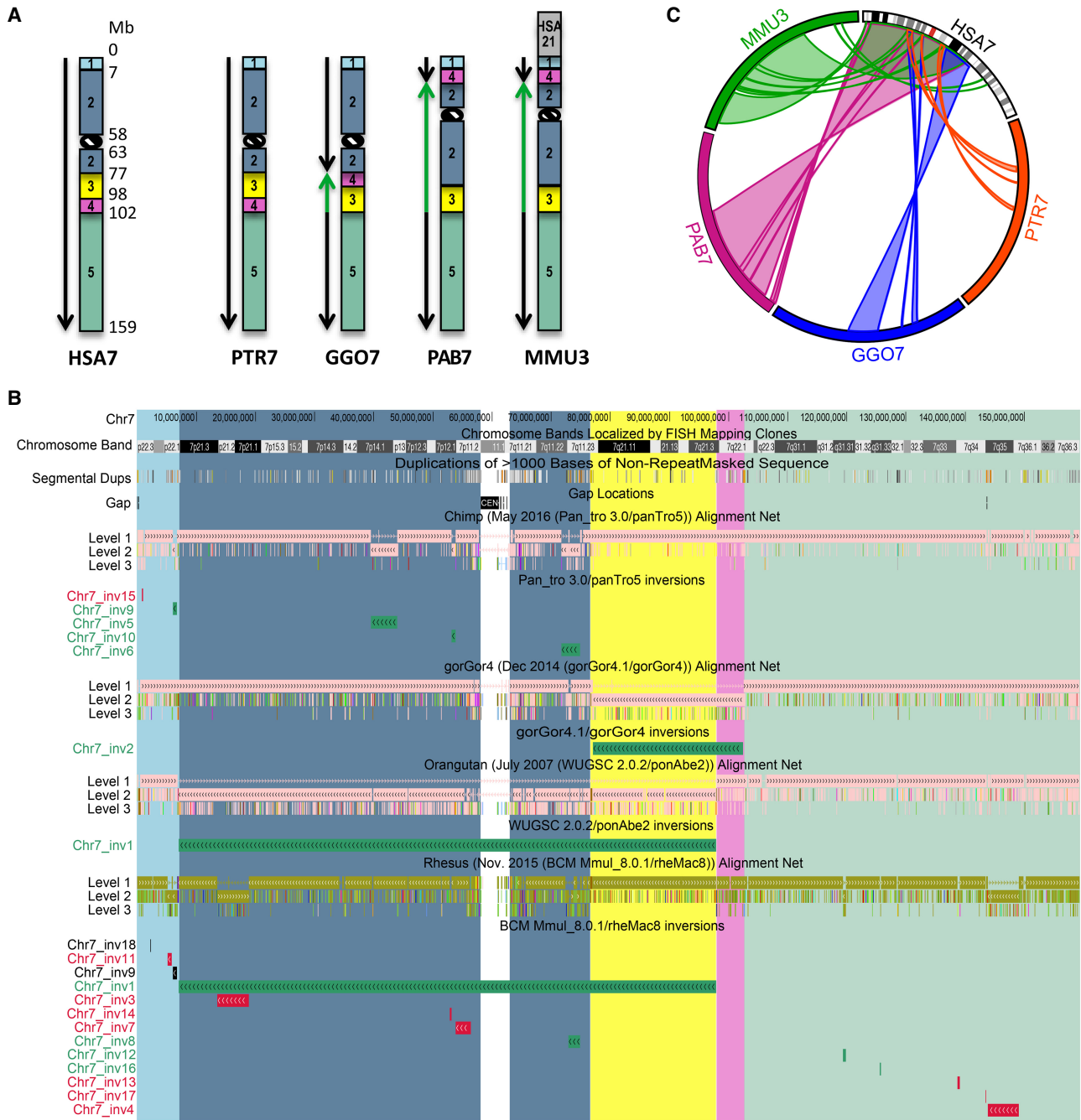
Figure 1. Inversions map of human Chromosome 7. (*A*) Ideograms for human Chromosome 7 (HSA7) and its chimpanzee (PTR7), gorilla (GGO7), orangutan (PAB7), and macaque (MMU3) homologs. Ideograms only show previously reported inversions larger than 5 Mb. Synteny blocks, distinguished by colors and numbers, represent regions that are inverted in at least one of the species analyzed, but the order of the markers within the block is conserved in all of the different species. Green arrows indicate inverted blocks with respect to human orientation. (*B*) UCSC Genome Browser view of Chromosome 7 net alignments and inversions predicted in this study between human and nonhuman primate genomes. Regions called to be inverted in this study are shown as green, red, and black horizontal bars and represent real, false, and not determined (ND) inversions, respectively. Synteny block colors are consistent with panel *A* and allow for comparison of regions called to be inverted in this study with previously identified inversions. For example, Chr7_inv1 corresponds to an inversion involving synteny blocks 2 (blue) and 3 (yellow) that was previously reported in orangutan and macaque (shown in panel *A*). (*C*) Circos diagram (Krzywinski et al. 2009) reporting all validated evolutionary inversions between human Chromosome 7 and its primate homologs.

either an error in the human assembly or the reference genome represents the minor allele. FISH results showed that all primate individuals tested are inverted compared to human except for gorilla; therefore, this region represents a real evolutionary inversion

event but assembled in the wrong orientation in chimpanzee, gorilla, and orangutan genomes (Supplemental Table S3). Four inversions were polymorphic in one or more primate individuals (Supplemental Table S4), while one inversion mapping on

10q11.22 was polymorphic just in humans with an inverted allele frequency of 37% (Supplemental Fig. S3).

Next, we investigated the BAC-end sequence (BES) pair mapping profiling of the predicted inversions by downloading the BES from all primate species analyzed and mapping them to the human reference GRCh38/hg38. BACs spanning inversions are discordant since they have end pairs that map abnormally far apart and have ends that are incorrectly oriented when mapped to the human reference genome sequence (Tuzun et al. 2005). Fifty-six of the putative inversion sites had BACs spanning at least one breakpoint. Of these, 24 showed discordant clones supporting the inversion and 31 events were false positives since only concordant clones from the primate species detected to be inverted mapped at the putative inversion breakpoints (Supplemental Tables S3, S6; Supplemental Fig. S4).

As a more direct means of validation, we selected 29 BAC clones for complete BAC-insert sequencing with Illumina sequencing as previously described (Supplemental Tables S3, S6; Supplemental Fig. S4; Steinberg et al. 2012). Sequencing was 100% consistent with our BES pair mapping profiling and FISH validation analysis (Supplemental Table S3).

In total, we investigated 126 inversion calls (109 human loci) through the literature (33/109 loci), experimental analyses (70/109 loci), and a combination of both (6/109 loci) and found that 83 calls (67/109 loci) represent real inversions. These include 14 calls between human and chimpanzee, nine between human and gorilla, 22 between human and orangutan, and 38 between human and macaque (Supplemental Table S3). Almost half of these loci (28/67 human regions) have been identified as inverted for the first time in this study. The remaining 39 events (39/126 inversion calls, 31%) were determined to be errors in the primate reference genomes, where the sequence was likely assembled in the wrong orientation (Supplemental Tables S3, S7).

Evolutionary analyses

We determined whether the 67 regions inverted in one or more primate species represent the derived or ancestral state based on comparisons with outgroup nonhuman primate species. To do so, we used previously published data (Feuk et al. 2005; Ventura et al. 2007, 2011; Cardone et al. 2008; Antonacci et al. 2009; Nuttle et al. 2016) for 33 regions, experimental analyses for 28 newly identified loci, and a combination of both for six regions for which published data was available for just a subset of primates.

We tested for the presence of 30 larger inversions by FISH analysis of cell lines from multiple individuals of chimpanzee (*Pan troglodytes*), gorilla (*Gorilla gorilla*), orangutan (*Pongo pygmaeus*), and macaque (*Macaca mulatta*). When necessary, marmoset (*Callithrix jacchus*) was used as an outgroup species (Supplemental Tables S3, S4).

For all these larger inversions and to resolve the status of the smaller inversions, we analyzed the BES pair mapping profiling in the different primate species analyzed and used marmoset as an outgroup when necessary. Of the 142 clones spanning the breakpoints of the validated inversions, 59 were fully sequenced with Illumina (Fig. 2; Supplemental Fig. S4; Supplemental Tables S3, S6).

In total, we were able to successfully determine the lineage specificity of 60 out of 67 inversions (Supplemental Table S3). Of these, three inversions occurred in the great ape ancestor, six in the human and chimpanzee ancestor, six are human-specific, seven occurred in the chimpanzee lineage, six in the gorilla lineage, three are orangutan-specific, 15 occurred in the macaque lineage,

and two in the Old World monkey ancestor. The remaining 12 inversions occurred in the human–African great ape ancestor, with two (Chr1_inv3; Chr15_inv1) in direct orientation (compared to the ancestral orientation) in chimpanzee. Here, either the region flipped back to the ancestral orientation in the chimpanzee lineage or the chimpanzee configuration may represent a case of incomplete lineage sorting (Fig. 3).

Since duplications play a pivotal role in origin of inversions, we compared the extent of segmental duplication at the breakpoints of the inversions among the different primate species. Of the great ape inversions, 82% (32 of 39) map to segmentally duplicated regions of the genome, compared to the inversion events identified in macaque that overlap with segmental duplications at their breakpoints in only 41% of the cases (seven out of 17) (Supplemental Table S3). Sequence analyses of the remaining macaque inversions show that 47% (eight out of 17) contain SINEs or LINEs at their breakpoints (Supplemental Table S8). We annotated the breakpoint ranges of each real inversion, based on the different validation methods used (i.e., previous studies, FISH, BES pair mapping, Illumina sequencing of BAC clones) (Supplemental Table S9).

Of the 67 inverted regions, 36 have breakpoints overlapping with 81 human (RefSeq curated) genes (Supplemental Table S10). We performed a Gene Ontology analysis and found that these genes belong to functional groups related to drug metabolism (cytochrome P450), receptors (G protein-coupled receptors and olfactory receptors), DNA-binding proteins (*ZFP14* zinc finger protein), and transport proteins.

Identification of human polymorphic inversions through comparison with previous studies

We compared our predicted inversion calls with previously identified human polymorphic inversions and found a match for seven regions. These include a 5.1-Mb inversion on Chromosome 4 and a 3.8-Mb inversion on Chromosome 8, both of which occurred in the human–chimpanzee common ancestor and were previously shown to be predisposing to further rearrangements associated with complex neurological disorders (Giglio et al. 2001, 2002; Antonacci et al. 2009; Sanders et al. 2016); a 2-Mb inversion on Chromosome 7 occurring in the great ape common ancestor, predisposing to the deletion in Williams-Beuren syndrome (Osborne et al. 2001; Schubert 2009; Sanders et al. 2016); a 735-kb inversion on Chromosome 7 occurring specifically in the human lineage (Feuk et al. 2005; Sanders et al. 2016); and two inversions of 287 kb and 1.3 Mb mapping on Chromosome 10 and Chromosome 16 (Sanders et al. 2016), respectively, that we were not able to validate. However, these last two inversions were predicted just in chimpanzee and therefore might have occurred in the human–chimpanzee ancestor (Table 1).

Errors in human and nonhuman primate reference genome assemblies

Our validations by FISH, BES pair mapping, and BAC clone Illumina sequencing (see "Inversion validations") identified 39 regions that represent genome assembly errors in one or more primate species (Supplemental Tables S3, S7). Moreover, by studying the evolutionary history of the 67 validated inversions (see "Evolutionary analyses"), we were able to identify 14 regions that are in direct orientation in some primate assemblies but were actually inverted in the species analyzed (Supplemental Table S7). An example is an inversion on Chromosome 15q13.1
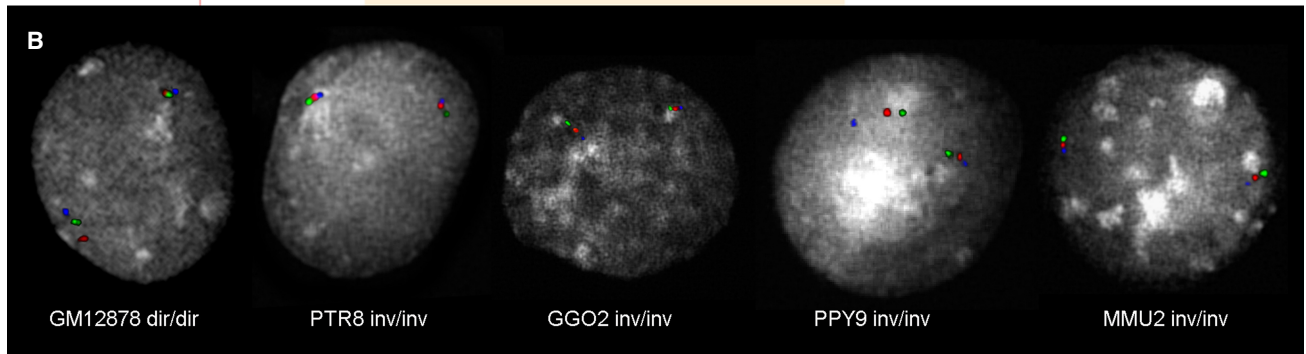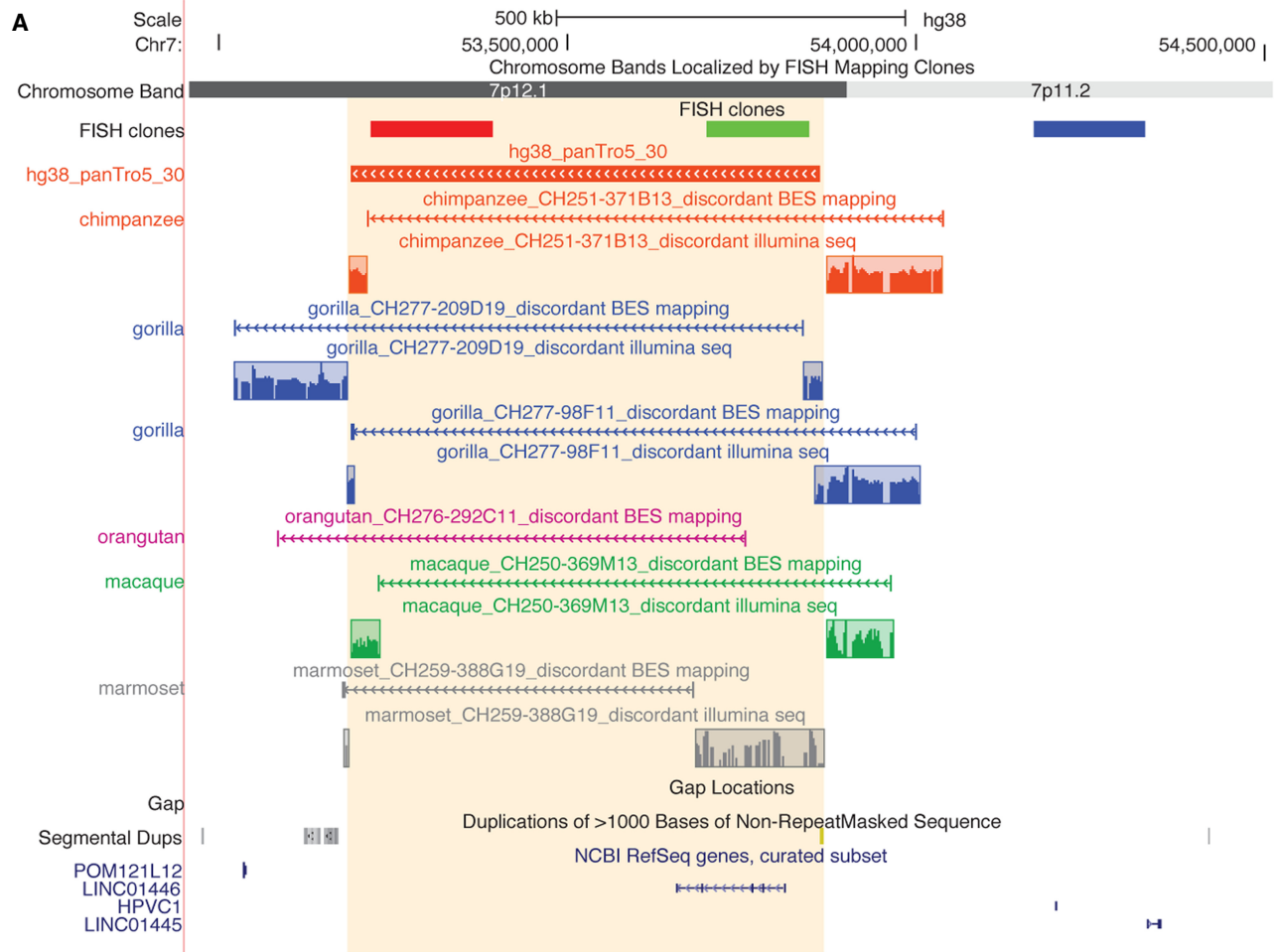
Figure 2. Experimental validation of Chr7_inv10 inversion. (*A*) UCSC Genome Browser view of Chr7_inv10 (hg38_panTro5_30), exclusively predicted in the chimpanzee lineage. BES pair mapping of primate clones and their Illumina sequencing (reads in the colored frames) show that all primates analyzed carry the inverted orientation. Discordant clones spanning the inversion breakpoints appear to be discontinuous due to the presence of the inversion. (*B*) The same inversion has been further validated by FISH in human (GM12878), chimpanzee (PTR8), gorilla (GGO2), orangutan (PPY9), and macaque (MMU2) individuals using the FISH clones shown in panel *A*.

(Chr15_inv1 inversion) identified in chimpanzee through alignments in inverted orientation between human and chimpanzee. Further evolutionary analyses by FISH showed that this region is also inverted in orangutan and macaque, although the genome assemblies for this region are in the same orientation as in human (Supplemental Tables S3, S4).

Finally, we searched for previously reported inversions in human and nonhuman primates and identified five regions inverted in one or more species (Antonacci et al. 2010; Ventura et al. 2011) but in direct orientation in their reference genome assemblies (Supplemental Table S7). Additional misassemblies identified through FISH experiments, BES pair mapping, and Illumina sequencing of BAC clones include 13 breakpoint errors of inversion calls and three inversions made up of two or more calls interrupted by an interval of sequence in direct orientation (Supplemental Table S7; Supplemental Fig. S5).
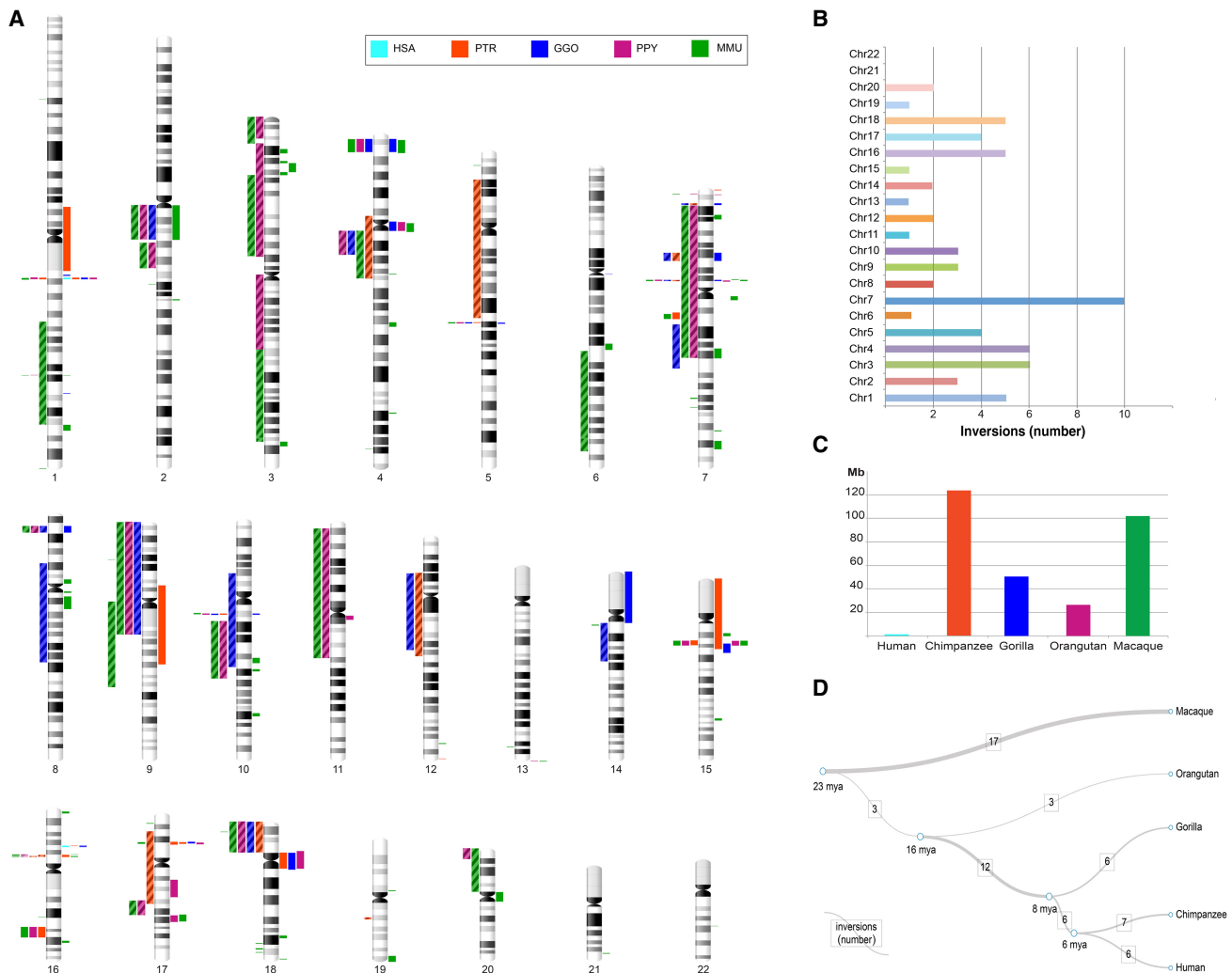
Figure 3. Map of primate inversions and assemblies errors. (*A*) All inversions discovered and validated between human, chimpanzee, gorilla, orangutan, and macaque chromosomes are shown on the *left* side of the chromosome ideograms. In particular, previously reported inversions reidentified in the current study are represented by colored blocks with a diagonal pattern, while novel inversions are depicted with solid color blocks. Errors in human and non-human primate assemblies are shown on the *right* side of the chromosome ideograms. (*B*) The horizontal bar chart shows the number of inversions per human chromosome. (*C*) Megabases (Mb) of assembly errors are shown for each species. (*D*) All inversions for which the lineage specificity has been determined are mapped on a phylogenetic tree (Sudmant et al. 2013) in which the branch thickness is proportional to the number of inversions. (HSA) *Homo sapiens*, (PTR) *Pan troglodytes*, (GGO) *Gorilla gorilla*, (PPY) *Pongo pygmaeus*, (MMU) *Macaca mulatta*.

In total, we detected 74 regions (Fig. 3), including over 300 Mb of sequence within human, great ape, and macaque reference genome assemblies involved in sequence orientation errors, overlapping with 1978 human RefSeq curated genes.

Discussion

We have generated the first genome-wide map of intermediate-scale inversion variants between human, great ape, and macaque genomes by comparing the net alignments for the most recent builds of these primate genome assemblies. We initially identified 136 human regions potentially inverted between human and either chimpanzee, gorilla, orangutan, or macaque genomes. Combining the literature, FISH analyses, BES pair mapping, and complete sequencing of primate clones for 120 inversion calls showed that 83 (69%: corresponding to 67 human loci) are real events. Each verified inversion was additionally tested in multiple

chimpanzee, gorilla, orangutan, and macaque individuals, allowing us to determine the ancestral state of each event and to also verify the inversion status in species where the assembly did not show the presence of the inversion. This allowed us to extend the number of validated primate inversions from 83 to 105 (Fig. 3). With our experimentally verified inversions, we more than doubled the number of events known between human, great ape, and macaque genomes. We identified 28 new regions between 103 kb and 5 Mb that are inverted in one or more primate species (Table 1; Fig. 3) and determined the ancestral orientation for 23 of these loci (Table 1; Supplemental Table S3).

In total, our inversions range in size from 103 kb to 91 Mb and are randomly distributed, with the highest number of events on Chromosome 7 and no inversions found on Chromosomes 21 and 22 (Fig. 3). We successfully resolved the evolutionary history of 60 inversions and found that macaque has the highest number of lineage-specific inversions (*n* = 17), followed by chimpanzee

Table 1. Inversions polymorphic in human and/or corresponding to sites of recurrent rearrangements associated with human disease

Inversion	Mapping (GRCh38/hg38)	Size (bp)	In the ancestor of	Disease association	References
Chr1_inv3	Chr 1: 147,079,442–147,925,603	846,161	HSA-PTR-GGO; ILS or recurrent inv in PTR	1q21.1-q21.2 deletion and duplication	Cooper et al. 2011
Chr4_inv5*	Chr 4: 4,182,444–9,339,607	5,157,163	HSA-PTR	Recurrent t(4;8)(p16;p23) translocation	Giglio et al. 2002
Chr7_inv8*	Chr 7: 72,893,718–74,869,947	1,976,229	Great apes	Williams-Beuren syndrome	Osborne et al. 2001; Schubert 2009; Sanders et al. 2016
Chr7_inv9*	Chr 7: 5,997,690– 6,732,324	734,634	HSA		Feuk et al. 2005; Sanders et al. 2016
Chr8_inv2*	Chr 8: 8,242,347–12,174,746	3,932,399	HSA-PTR	inv dup(8p)	Giglio et al. 2001; Antonacci et al. 2009; Sanders et al. 2016
Chr10_inv6*	Chr 10: 46,870,207–47,457,081	586,874	HSA	10q11.22 deletion and duplication	Cooper et al. 2011
Chr10_inv7*	Chr 10: 79,923,086– 80,209,920	286,834	ND (predicted HSA-PTR)		Sanders et al. 2016
Chr15_inv1	Chr 15: 28,852,754–30,406,229	1,553,475	HSA-PTR-GGO; ILS or recurrent inv in PTR	15q13.1-q13.2 deletion and duplication	Cooper et al. 2011
Chr16_inv5*	Chr 16: 34,962,662– 36,253,738	1,291,076	ND (predicted HSA-PTR)		Sanders et al. 2016
Chr16_inv7	Chr 16: 28,378,167–29,034,255	656,088	HSA-PTR	16p11.2 deletion and duplication	Cooper et al. 2011
Chr16_inv8	Chr 16: 29,640,910–30,210,335	569,425	HSA-PTR	16p11.2 deletion and duplication	Cooper et al. 2011

Shown are 11 regions that are polymorphic in humans and/or associated with human disease. The asterisk (*) indicates inversions found to be polymorphic in human.
(HSA) *Homo sapiens*, (PTR) *Pan troglodytes*, (GGO) *Gorilla gorilla*, (ND) not determined, (ILS) incomplete lineage sorting.

(n = 7), human (n = 6), gorilla (n = 6), and orangutan (n = 3). A high number (n = 12) of inversions seems to have occurred in the African great ape ancestor. The remaining inversions occurred in the human–chimpanzee ancestor (n = 6) and in the great ape ancestor (n = 3). Our sequence analyses suggest that NAHR, mediated by segmental duplications, is the predominant mechanism underlying these events (Supplemental Table S3) in humans and great apes (82%). Only 41% of macaque inversion breakpoints contain segmental duplications, while LINEs and SINEs were found at the remaining breakpoints (47%) (Supplemental Tables S3, S8). This is expected, since most (80%) high-identity segmental duplications arose after the divergence of the Old World and hominoid lineages (Marques-Bonet et al. 2009).

We identified a very low number of inversion calls generated by the comparison of the net alignments of human (GRCh38/hg38) and gorilla (gorGor4.1/gorGor4) genomes. Experimental validations show that only previously reported large events are real (Ventura et al. 2011). The remaining smaller inversions are false calls or have not been verified because they are too small for FISH validation and BES mapping profiling results are unclear (Supplemental Table S3). The published gorilla genome assembly (gorGor4.1/gorGor4) used in this study was generated by a mix of capillary sequence and whole-genome shotgun Illumina sequencing, resulting in an assembly containing more than 400,000 gaps (Scally et al. 2012). The human genome was used to help guide the gorilla assembly, therefore generating an artificially low number of inversions. Unfortunately, the absence of the chromosome information for the latest gorilla assembly, GSMRT3/gorGor5, made it impossible to use it for our analysis.

In searching for genes that could be altered by the presence of the inversions, we found that 36 out of 67 loci are inverted in one or more primates and have breakpoints overlapping 81 human genes (Supplemental Table S10), prioritizing them as candidates for biological and evolutionary studies. These genes include members of several functional groups, including those related to drug metabolism, receptors, DNA-binding proteins, and transport proteins. Future experimental studies are required to demonstrate the functional significance of these genes in contributing to phenotypic differences among humans, great apes, and macaque.

Many previously identified inversion variants have been linked to susceptibility to further genomic rearrangements in offspring. We therefore investigated all inversion regions under 10 Mb in size in which humans carry the inverted orientation compared to the ancestral state and found seven inversions corresponding to sites of recurrent rearrangements associated with human disease, with four being polymorphic in humans (Table 1; Giglio et al. 2001, 2002; Osborne et al. 2001; Feuk et al. 2005; Antonacci et al. 2009; Schubert 2009; Cooper et al. 2011; Sanders et al. 2016).

This included four previously described events: a 5.1-Mb inversion on 4p16.3-p16.1 (Chr4_inv5) (Supplemental Fig. S3) and a 3.9-Mb inversion on 8p23.1 (Chr8_inv2), both known to be polymorphic in the human population and predisposing to recurrent t(4;8)(p16;p23) translocations and inv dup(8p) rearrangements in the offspring (Giglio et al. 2001, 2002; Antonacci et al. 2009; Sanders et al. 2016); a 2-Mb polymorphic inversion on 7q11.23 (Chr7_inv8) predisposing to Williams-Beuren syndrome (Osborne et al. 2001; Schubert 2009; Sanders et al. 2016); and a 569-kb inversion (Chr16_inv8) mapping to the 16p11.2 disease-associated region (Cooper et al. 2011; Nuttle et al. 2016). We also detected three novel smaller inversions, including a 1.5-Mb inversion on 15q13.1-q13.2 (Chr15_inv1), a 846-kb inversion on 1q21.1-q21.2 (Chr1_inv3), and a 580-kb inversion on 10q11.22 (Chr10_inv6) (Table 2; Supplemental Fig. S3). These three regions correspond to sites of recurrent deletions and duplications associated with intellectual disability and developmental delay (Cooper et al. 2011), providing further support for a link between human inversions and genomic disorders. The 10q11.22 region was recently resequenced and assembled using single-molecule, real-time (SMRT) sequencing of BAC clones (Chaisson et al. 2015) since the previous build of the human reference genome (GRCh37/

Table 2. Detected and validated novel inversion events

Inversion	Mapping (build38)	Size (bp)	In the ancestor of
Chr1_inv3	Chr 1: 147,079,442–147,925,603	846,161	HSA-PTR-GGO; ILS or recurrent inv in PTR
Chr1_inv6	Chr 1: 248,121,676–248,295,646	173,970	ND
Chr1_inv7	Chr 1: 194,242,717–194,399,262	156,545	HSA-PTR-GGO
Chr1_inv8	Chr 1: 46,948,073–47,050,839	102,766	ND
Chr2_inv4	Chr 2: 137,588,644–137,750,208	161,564	MMU
Chr4_inv5	Chr 4: 4,182,444–9,339,607	5,157,163	HSA-PTR
Chr4_inv9	Chr 4: 178,205,683–178,375,658	169,975	Great apes
Chr5_inv2	Chr 5: 99,582,578–100,374,690	792,112	ND
Chr5_inv3	Chr 5: 99,584,037–100,069,297	485,260	ND
Chr5_inv4	Chr 5: 8,767,190–8,984,256	217,066	MMU
Chr7_inv6	Chr 7: 71,693,970–74,869,950	3,175,980	PTR
Chr7_inv8	Chr 7: 72,893,718–74,869,947	1,976,229	Great apes
Chr7_inv10*	Chr 7: 53,188,941–53,862,225	673,284	HSA
Chr7_inv12	Chr 7: 119,289,791–119,798,781	508,990	MMU
Chr7_inv16	Chr 7: 125,524,041–125,732,150	208,109	MMU
Chr7_inv18	Chr 7: 2,531,298–2,671,789	140,491	HSA-PTR-GGO
Chr9_inv4	Chr 9: 24,424,656–24,553,856	129,200	MMU
Chr10_inv6	Chr 10: 46,870,207–47,457,081	586,874	HSA
Chr13_inv2	Chr 13: 104,797,611–105,108,126	310,515	MMU
Chr14_inv2	Chr 14: 19,714,886–20,148,024	433,138	ND
Chr15_inv1	Chr 15: 28,852,754–30,406,229	1,553,475	HSA-PTR-GGO; ILS or recurrent inv in PTR
Chr16_inv1	Chr 16: 70,075,634–74,327,699	4,252,065	HSA
Chr16_inv3*	Chr 16: 28,781,228–30,210,335	1,429,107	PTR
Chr16_inv7*	Chr 16: 28,378,167–29,034,255	656,088	HSA-PTR
Chr16_inv8*	Chr 16: 29,640,910–30,210,335	569,425	HSA-PTR
Chr16_inv11	Chr 16: 65,673,523–65,868,382	194,859	MMU
Chr17_inv4	Chr 17: 15,812,020–16,664,213	852,193	Great apes
Chr17_inv7	Chr 17: 3,064,892–3,217,165	152,273	ND
Chr18_inv3	Chr 18: 70,885,697–71,666,858	781,161	MMU
Chr18_inv6	Chr 18: 68,257,880–68,546,524	288,644	MMU
Chr18_inv7	Chr 18: 73,572,681–73,849,640	276,959	MMU
Chr18_inv8	Chr 18: 5,677,162–5,815,084	137,922	MMU
Chr19_inv2*	Chr 19: 36,331,795–37,251,831	920,036	PTR

Shown are 28 new inversions detected and validated in this study and five inversions, labeled with an asterisk (*), that were previously studied (Feuk et al. 2005; Nuttle et al. 2016) in a subset of primates and for which we extended the evolutionary analyses. The evolutionary history was resolved for 27 events.
(HSA) *Homo sapiens*, (PTR) *Pan troglodytes*, (GGO) *Gorilla gorilla* (MMU) *Macaca mulatta*, (ND) not determined, (ILS) incomplete lineage sorting.

hg19) contained seven gaps. The region is particularly enriched in segmental duplications in all the primate genomes analyzed (Marques-Bonet et al. 2009; Sudmant et al. 2013) and is polymorphic in human with a frequency of the minor allele in inverted orientation of 37%. All the other nonhuman primates tested are in opposite orientation compared to most humans, suggesting that the inversion occurred in the human lineage and is still polymorphic (Supplemental Fig. S3; Supplemental Table S4). It is possible that the inverted haplotype carries a genomic architecture that now predisposes this region to deletions and duplications associated with neuropsychiatric disease.

Notably, FISH results for the 1q21.1-q21.1 region (Chr1_inv3), inverted in the chimpanzee and orangutan genome references (calls hg38_panTro5_22 and hg38_ponAbe2_22), showed that 20 of 20 human chromosomes tested were inverted relative to GRCh38/hg38, suggesting a potential error in the orientation of the reference genome assembly involving 10 genes (Supplemental Table S3). Alternatively, the human reference at this complex region of the genome may represent a rare haplotype. Therefore, this disease-associated region of 846 kb represents a real evolutionary inversion event that is misassembled in chimpanzee, gorilla, and orangutan reference genomes.

Of note, our analysis yielded a total of 74 regions that are not properly assembled in the published versions of one or more primate genomes, containing over 300 Mb of DNA and 1978 human (RefSeq curated) genes. Using BES pair mapping analyses and

Illumina sequencing of BAC inserts, we identified species-specific clones for future further characterization through long-read single-molecule sequencing (Chaisson et al. 2015) in order to refine the sequence and orientation at these complex genomic regions. Further efforts should be devoted toward producing primate genome sequences with greater accuracy and completeness. This is critical to detect subtle differences in genes or their regulatory regions that could lead to hominid physiological and behavioral differences and, importantly, to know that such differences are due to biology rather than poor sequence quality. In addition, many primate species are significant biomedical research models, and high-quality sequencing of primate genomes is needed to provide comparative sequence information that has implications for the understanding of the genetic basis for human disease.

Methods

Sequence alignments between human and primate genome assemblies

Tables of net alignments between a human (GRCh38/hg38) and a primate genome, i.e., chimpanzee (Pan_tro 3.0/panTro5), gorilla (gorGor4.1/gorGor4), orangutan (WUGSC 2.0.2/ponAbe2), and macaque (BCM Mmul_8.0.1/rheMac8), were downloaded from the University of California, Santa Cruz "goldenPath" website (http://hgdownload-test.cse.ucsc.edu/goldenPath/). For every

primate genome, the net-alignment table shows the best human/primate genome chain alignment, with a gap scoring system allowing longer gaps than traditional affine gap scoring systems. This analysis is useful for finding orthologous regions and for studying genomic rearrangements (Chiaromonte et al. 2002; Schwartz et al. 2003). All alignments labeled as inversions were filtered for size (longer than 100 kb) and percentage of repeats or segmental duplications (<90% of the inversion size). To reduce the number of false positives, we filtered out the nested inversions, considering only alignments better than level four of the net-alignment output, as previously described by Feuk et al. (2005). We repeated the analyses using the complementary net alignments (primate species versus human genome) and manually inspected inversion calls that were missed in the first analysis. This step allowed us to add three additional calls (rheMac8_hg38_98; gorGor4_hg38_5; panTro5_hg38_31). Alignments on sex chromosomes were excluded from the analysis.

Inversion maps of large known genomic inversions

All previously reported inversions larger than 5 Mb were used to draw ideograms for each chromosome in all the species analyzed (Fig. 1; Supplemental Fig. S1; Ventura et al. 2007, 2011; Cardone et al. 2008). Each colored block represents a synteny block—a region previously described as inverted in at least one of the species analyzed—but the order of the markers within the block is conserved in all the different species.

BES pair mapping

During the sequencing of the chimpanzee, gorilla, orangutan, and macaque genomes, BAC libraries (CHORI-251, CHORI-277, CHORI-276, CHORI-250, and CHORI-259) were constructed from the same individuals used to generate the sequence assemblies. We obtained the sequence for all traces from the NIH trace repository (http://www.ncbi.nlm.nih.gov/Traces/trace.cgi?). We aligned BES against the human reference genome sequence (GRCh38/hg38) as part of a three-step process (recruitment, quality rescoring, and pairing) optimized and published by Tuzun et al. (2005). Since the original protocol was applied using fosmid paired ends, we adapted it based on longer expected BAC insert sizes (range: 160–180 kb). BACs spanning inversions are *discordant* since they have end pairs that map abnormally far apart and have ends that are incorrectly oriented when mapped to the human reference genome sequence. Clones spanning regions in the same orientation as in human are *concordant* in size and for orientation of the ends.

FISH analysis

Metaphase spreads and interphase nuclei were obtained from lymphoblast and fibroblast cell lines from 10 human HapMap individuals (Coriell Cell Repository), four chimpanzees, three gorillas, two orangutans, two macaques, and one marmoset (Supplemental Table S4). FISH experiments were performed using human fosmid ($n = 46$) or BAC ($n = 41$) clones (Supplemental Table S5; Supplemental Fig. S2) directly labeled by nick-translation with Cy3-dUTP (PerkinElmer), Cy5-dUTP (PerkinElmer), and fluorescein-dUTP (Enzo) as described by Lichter et al. (1990), with minor modifications. Briefly, 300 ng of labeled probe were used for the FISH experiments; hybridization was performed at 37°C in 2×SSC, 50% (v/v) formamide, 10% (w/v) dextran sulphate, and 3 mg sonicated salmon sperm DNA, in a volume of 10 mL. Post-hybridization washing was at 60°C in 0.1×SSC (three times, high stringency, for hybridizations on human, chimpanzee, gorilla, and orangutan), or at 37°C in 2×SSC and 42°C in 2×SSC, 50% form-

amide (three times each, low stringency, for hybridizations on macaque and marmoset). Nuclei were simultaneously DAPI-stained. Digital images were obtained using a Leica DMRXA2 epi-fluorescence microscope equipped with a cooled CCD camera (Princeton Instruments). DAPI, Cy3, Cy5, and fluorescein fluorescence signals, detected with specific filters, were recorded separately as grayscale images. Pseudocoloring and merging of images were performed using Adobe Photoshop software. Interphase three-color FISH was used to validate inversions between 500 kb and 2 Mb in size, and metaphase two-color FISH was used for inversions larger than 2 Mb (Supplemental Table S5). In the case of interphase three-color FISH, each region was interrogated using two probes within the putative inversion region and a reference probe outside. A change in the order of the probes mapping within the inversion was indicative of the presence of the inversion. Examining inversions in interphase, distance, position, and order between FISH dots of different colors has to be measured in order to reveal their spatial pattern. Therefore, a minimum of 50 interphase cells were scored for each region in order to determine if the pattern observed was casual or due to a real inversion. For this reason, only patterns where the probes were positioned on a straight line and whose distance was consistent with their mapping position were considered. Probes overlapping or whose pattern resembled a triangle were excluded. A region was considered homozygously inverted if scoring of the probes in inverted orientation exceeded 80% of the total count and heterozygously inverted if probes in direct and inverted orientation were equally scored. A minimum of 10 cells were scored instead for FISH on metaphase chromosomes. In this case, the orientation of the region was more easily assessed by simply visualizing a switch in the order of the probes mapping within the inversion.

Illumina sequencing of BAC clones

DNA from 11 CH251, nine CH277, eight CH276, 26 CH250, and five CH259 BAC clones (Supplemental Table S6) was isolated, prepped into genomic libraries, and sequenced (PE250) on an Illumina MiSeq using a Nextera protocol (Antonacci et al. 2010). DNA from 37 clones was barcoded before library preparation, while DNA from 22 clones mapping to different chromosomes and free of segmental duplications was pooled two at a time before library preparation, and then barcoded and sequenced. Sequencing data were mapped with mrsFAST (Alkan et al. 2009) to the human reference genome and singly unique nucleotide (SUN) identifiers were used to discriminate between highly identical segmental duplications (Sudmant et al. 2010).

Gene Ontology analysis

Genes at the inversions breakpoints were analyzed using InterPro (http://www.ebi.ac.uk/interpro/), which is a database that integrates diverse information about protein families, domains, and functional sites and makes it freely available to the public via web-based interfaces and services (Hunter et al. 2012). Gene Ontology codes were extracted from the InterPro output and manually clusterized.

Data access

Raw sequencing data from Illumina sequencing experiments from this study have been submitted to the Sequence Read Archive (SRA; https://www.ncbi.nlm.nih.gov/sra/) under BioProject PRJNA429373.

Acknowledgments

References

Alkan C, Kidd JM, Marques-Bonet T, Aksay G, Antonacci F, Hormozdiari F, Kitzman JO, Baker C, Malig M, Mutlu O, et al. 2009. Personalized copy number and segmental duplication maps using next-generation sequencing. *Nat Genet* **41:** 1061–1067.

Alkan C, Coe BP, Eichler EE. 2011. Genome structural variation discovery and genotyping. *Nat Rev Genet* **12:** 363–376.

Antonacci F, Kidd JM, Marques-Bonet T, Ventura M, Siswara P, Jiang Z, Eichler EE. 2009. Characterization of six human disease-associated inversion polymorphisms. *Hum Mol Genet* **18:** 2555–2566.

Antonacci F, Kidd JM, Marques-Bonet T, Teague B, Ventura M, Girirajan S, Alkan C, Campbell CD, Vives L, Malig M, et al. 2010. A large and complex structural polymorphism at 16p12.1 underlies microdeletion disease risk. *Nat Genet* **42:** 745–750.

Bailey JA, Yavor AM, Massa HF, Trask BJ, Eichler EE. 2001. Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res* **11:** 1005–1017.

Capozzi O, Carbone L, Stanyon RR, Marra A, Yang F, Whelan CW, de Jong PJ, Rocchi M, Archidiacono N. 2012. A comprehensive molecular cytogenetic analysis of chromosome rearrangements in gibbons. *Genome Res* **22:** 2520–2528.

Carbone L, Ventura M, Tempesta S, Rocchi M, Archidiacono N. 2002. Evolutionary history of chromosome 10 in primates. *Chromosoma* **111:** 267–272.

Carbone L, Harris RA, Gnerre S, Veeramah KR, Lorente-Galdos B, Huddleston J, Meyer TJ, Herrero J, Roos C, Aken B, et al. 2014. Gibbon genome and the fast karyotype evolution of small apes. *Nature* **513:** 195–201.

Cardone MF, Alonso A, Pazienza M, Ventura M, Montemurro G, Carbone L, de Jong PJ, Stanyon R, D'Addabbo P, Archidiacono N, et al. 2006. Independent centromere formation in a capricious, gene-free domain of chromosome 13q21 in Old World monkeys and pigs. *Genome Biol* **7:** R91.

Cardone MF, Lomiento M, Teti MG, Misceo D, Roberto R, Capozzi O, D'Addabbo P, Ventura M, Rocchi M, Archidiacono N. 2007. Evolutionary history of chromosome 11 featuring four distinct centromere repositioning events in Catarrhini. *Genomics* **90:** 35–43.

Cardone MF, Jiang Z, D'Addabbo P, Archidiacono N, Rocchi M, Eichler EE, Ventura M. 2008. Hominoid chromosomal rearrangements on 17q map to complex regions of segmental duplication. *Genome Biol* **9:** R28.

Chaisson MJ, Huddleston J, Dennis MY, Sudmant PH, Malig M, Hormozdiari F, Antonacci F, Surti U, Sandstrom R, Boitano M, et al. 2015. Resolving the complexity of the human genome using single-molecule sequencing. *Nature* **517:** 608–611.

Chiaromonte F, Yap VB, Miller W. 2002. Scoring pairwise genomic sequence alignments. *Pac Symp Biocomput* **2002:** 115–126.

The Chimpanzee Sequencing and Analysis Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437:** 69–87.

Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, Williams C, Stalker H, Hamid R, Hannig V, et al. 2011. A copy number variation morbidity map of developmental delay. *Nat Genet* **43:** 838–846.

Eder V, Ventura M, Ianigro M, Teti M, Rocchi M, Archidiacono N. 2003. Chromosome 6 phylogeny in primates and centromere repositioning. *Mol Biol Evol* **20:** 1506–1512.

Farre M, Micheletti D, Ruiz-Herrera A. 2013. Recombination rates and genomic shuffling in human and chimpanzee—a new twist in the chromosomal speciation theory. *Mol Biol Evol* **30:** 853–864.

Feuk L, MacDonald JR, Tang T, Carson AR, Li M, Rao G, Khaja R, Scherer SW. 2005. Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee DNA sequence assemblies. *PLoS Genet* **1:** e56.

Giglio S, Broman KW, Matsumoto N, Calvari V, Gimelli G, Neumann T, Ohashi H, Voullaire L, Larizza D, Giorda R, et al. 2001. Olfactory receptor–gene clusters, genomic-inversion polymorphisms, and common chromosome rearrangements. *Am J Hum Genet* **68:** 874–883.

Giglio S, Calvari V, Gregato G, Gimelli G, Camanini S, Giorda R, Ragusa A, Guerneri S, Selicorni A, Stumm M, et al. 2002. Heterozygous submicroscopic inversions involving olfactory receptor–gene clusters mediate the recurrent t(4;8)(p16;p23) translocation. *Am J Hum Genet* **71:** 276–285.

Goidts V, Szamalek JM, Hameister H, Kehrer-Sawatzki H. 2004. Segmental duplication associated with the human-specific inversion of chromosome 18: a further example of the impact of segmental duplications on karyotype and genome evolution in primates. *Hum Genet* **115:** 116–122.

Gordon D, Huddleston J, Chaisson MJ, Hill CM, Kronenberg ZN, Munson KM, Malig M, Raja A, Fiddes I, Hillier LW, et al. 2016. Long-read sequence assembly of the gorilla genome. *Science* **352:** aae0344.

Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, Bernard T, Binns D, Bork P, Burge S, et al. 2012. InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res* **40:** D306–D312.

Jauch A, Wienberg J, Stanyon R, Arnold N, Tofanelli S, Ishida T, Cremer T. 1992. Reconstruction of genomic rearrangements in great apes and gibbons by chromosome painting. *Proc Natl Acad Sci* **89:** 8611–8615.

Kehrer-Sawatzki H, Cooper DN. 2008. Molecular mechanisms of chromosomal rearrangement during primate evolution. *Chromosome Res* **16:** 41–56.

Kehrer-Sawatzki H, Sandig C, Chuzhanova N, Goidts V, Szamalek JM, Tanzer S, Muller S, Platzer M, Cooper DN, Hameister H. 2005a. Breakpoint analysis of the pericentric inversion distinguishing human chromosome 4 from the homologous chromosome in the chimpanzee (*Pan troglodytes*). *Hum Mutat* **25:** 45–55.

Kehrer-Sawatzki H, Sandig CA, Goidts V, Hameister H. 2005b. Breakpoint analysis of the pericentric inversion between chimpanzee chromosome 10 and the homologous chromosome 12 in humans. *Cytogenet Genome Res* **108:** 91–97.

Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, Graves T, Hansen N, Teague B, Alkan C, Antonacci F, et al. 2008. Mapping and sequencing of structural variation from eight human genomes. *Nature* **453:** 56–64.

Koolen DA, Sharp AJ, Hurst JA, Firth HV, Knight SJ, Goldenberg A, Saugier-Veber P, Pfundt R, Vissers LE, Destree A, et al. 2008. Clinical and molecular delineation of the 17q21.31 microdeletion syndrome. *J Med Genet* **45:** 710–720.

Krzywinski M, Schein J, Birol İ, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res* **19:** 1639–1645.

Kuderna LFK, Tomlinson C, Hillier LW, Tran A, Fiddes IT, Armstrong J, Laayouni H, Gordon D, Huddleston J, Garcia Perez R, et al. 2017. A 3-way hybrid approach to generate a new high-quality chimpanzee reference genome (Pan_tro_3.0). *Gigascience* **6:** 1–6.

Lichter P, Tang CJ, Call K, Hermanson G, Evans GA, Housman D, Ward DC. 1990. High-resolution mapping of human chromosome 11 by in situ hybridization with cosmid clones. *Science* **247:** 64–69.

Locke DP, Hillier LW, Warren WC, Worley KC, Nazareth LV, Muzny DM, Yang SP, Wang Z, Chinwalla AT, Minx P, et al. 2011. Comparative and demographic analysis of orang-utan genomes. *Nature* **469:** 529–533.

Lupski JR. 1998. Genomic disorders: Structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet* **14:** 417–422.

Marques-Bonet T, Kidd JM, Ventura M, Graves TA, Cheng Z, Hillier LW, Jiang Z, Baker C, Malfavon-Borja R, Fulton LA, et al. 2009. A burst of segmental duplications in the genome of the African great ape ancestor. *Nature* **457:** 877–881.

Misceo D, Ventura M, Eder V, Rocchi M, Archidiacono N. 2003. Human chromosome 16 conservation in primates. *Chromosome Res* **11:** 323–326.

Misceo D, Cardone MF, Carbone L, D'Addabbo P, de Jong PJ, Rocchi M, Archidiacono N. 2005. Evolutionary history of chromosome 20. *Mol Biol Evol* **22:** 360–366.

Muller S, Hollatz M, Wienberg J. 2003. Chromosomal phylogeny and evolution of gibbons (Hylobatidae). *Hum Genet* **113:** 493–501.

Navarro A, Betran E, Barbadilla A, Ruiz A. 1997. Recombination and gene flux caused by gene conversion and crossing over in inversion heterokaryotypes. *Genetics* **146:** 695–709.

Newman TL, Tuzun E, Morrison VA, Hayden KE, Ventura M, McGrath SD, Rocchi M, Eichler EE. 2005. A genome-wide survey of structural variation between human and chimpanzee. *Genome Res* **15:** 1344–1356.

Nuttle X, Giannuzzi G, Duyzend MH, Schraiber JG, Narvaiza I, Sudmant PH, Penn O, Chiatante G, Malig M, Huddleston J, et al. 2016. Emergence of a *Homo sapiens*-specific gene family and chromosome 16p11.2 CNV susceptibility. *Nature* **536:** 205–209.

Osborne LR, Li M, Pober B, Chitayat D, Bodurtha J, Mandel A, Costa T, Grebe T, Cox S, Tsui LC, et al. 2001. A 1.5 million–base pair inversion polymorphism in families with Williams-Beuren syndrome. *Nat Genet* **29:** 321–325.

Rhesus Macaque Genome Sequencing and Analysis Consortium. 2007. Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316:** 222–234.

Sanders AD, Hills M, Porubsky D, Guryev V, Falconer E, Lansdorp PM. 2016. Characterizing polymorphic inversions in human genomes by single-cell sequencing. *Genome Res* **26:** 1575–1587.

Scally A, Dutheil JY, Hillier LW, Jordan GE, Goodhead I, Herrero J, Hobolth A, Lappalainen T, Mailund T, Marques-Bonet T, et al. 2012. Insights into hominid evolution from the gorilla genome sequence. *Nature* **483:** 169–175.

Schneider VA, Graves-Lindsay T, Howe K, Bouk N, Chen HC, Kitts PA, Murphy TD, Pruitt KD, Thibaud-Nissen F, Albracht D, et al. 2017. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res* **27:** 849–864.

Schubert C. 2009. The genomic basis of the Williams-Beuren syndrome. *Cell Mol Life Sci* **66:** 1178–1197.

Schwartz S, Kent WJ, Smit A, Zhang Z, Baertsch R, Hardison RC, Haussler D, Miller W. 2003. Human-mouse alignments with BLASTZ. *Genome Res* **13:** 103–107.

Skipper L, Wilkes K, Toft M, Baker M, Lincoln S, Hulihan M, Ross OA, Hutton M, Aasly J, Farrer M. 2004. Linkage disequilibrium and association of *MAPT* H1 in Parkinson disease. *Am J Hum Genet* **75:** 669–677.

Stanyon R, Rocchi M, Capozzi O, Roberto R, Misceo D, Ventura M, Cardone MF, Bigoni F, Archidiacono N. 2008. Primate chromosome evolution: ancestral karyotypes, marker order and neocentromeres. *Chromosome Res* **16:** 17–39.

Stefansson H, Helgason A, Thorleifsson G, Steinthorsdottir V, Masson G, Barnard J, Baker A, Jonasdottir A, Ingason A, Gudnadottir VG, et al. 2005. A common inversion under selection in Europeans. *Nat Genet* **37:** 129–137.

Steinberg KM, Antonacci F, Sudmant PH, Kidd JM, Campbell CD, Vives L, Malig M, Scheinfeldt L, Beggs W, Ibrahim M, et al. 2012. Structural diversity and African origin of the 17q21.31 inversion polymorphism. *Nat Genet* **44:** 872–880.

Sturtevant AH. 1917. Genetic factors affecting the strength of linkage in *Drosophila*. *Proc Natl Acad Sci* **3:** 555–558.

Sudmant PH, Kitzman JO, Antonacci F, Alkan C, Malig M, Tsalenko A, Sampas N, Bruhn L, Shendure J, Genomes P, et al. 2010. Diversity of human copy number variation and multicopy genes. *Science* **330:** 641–646.

Sudmant PH, Huddleston J, Catacchio CR, Malig M, Hillier LW, Baker C, Mohajeri K, Kondova I, Bontrop RE, Persengiev S, et al. 2013. Evolution and diversity of copy number variation in the great ape lineage. *Genome Res* **23:** 1373–1382.

Tobin JE, Latourelle JC, Lew MF, Klein C, Suchowersky O, Shill HA, Golbe LI, Mark MH, Growdon JH, Wooten GF, et al. 2008. Haplotypes and gene expression implicate the *MAPT* region for Parkinson disease: the *Gene*PD study. *Neurology* **71:** 28–34.

Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, Haugen E, Hayden H, Albertson D, Pinkel D, et al. 2005. Fine-scale structural variation of the human genome. *Nat Genet* **37:** 727–732.

Ventura M, Archidiacono N, Rocchi M. 2001. Centromere emergence in evolution. *Genome Res* **11:** 595–599.

Ventura M, Mudge JM, Palumbo V, Burn S, Blennow E, Pierluigi M, Giorda R, Zuffardi O, Archidiacono N, Jackson MS, et al. 2003. Neocentromeres in 15q24-26 map to duplicons which flanked an ancestral centromere in 15q25. *Genome Res* **13:** 2059–2068.

Ventura M, Weigl S, Carbone L, Cardone MF, Misceo D, Teti M, D'Addabbo P, Wandall A, Bjorck E, de Jong PJ, et al. 2004. Recurrent sites for new centromere seeding. *Genome Res* **14:** 1696–1703.

Ventura M, Antonacci F, Cardone MF, Stanyon R, D'Addabbo P, Cellamare A, Sprague LJ, Eichler EE, Archidiacono N, Rocchi M. 2007. Evolutionary formation of new centromeres in macaque. *Science* **316:** 243–246.

Ventura M, Catacchio CR, Alkan C, Marques-Bonet T, Sajjadian S, Graves TA, Hormozdiari F, Navarro A, Malig M, Baker C, et al. 2011. Gorilla genome structural variation reveals evolutionary parallelisms with chimpanzee. *Genome Res* **21:** 1640–1649.

Vicente-Salvador D, Puig M, Gaya-Vidal M, Pacheco S, Giner-Delgado C, Noguera I, Izquierdo D, Martinez-Fundichely A, Ruiz-Herrera A, Estivill X, et al. 2017. Detailed analysis of inversions predicted between two human genomes: errors, real polymorphisms, and their origin and population distribution. *Hum Mol Genet* **26:** 567–581.

Yunis JJ, Prakash O. 1982. The origin of man: a chromosomal pictorial legacy. *Science* **215:** 1525–1530.

Zimin AV, Cornish AS, Maudhoo MD, Gibbs RM, Zhang X, Pandey S, Meehan DT, Wipfler K, Bosinger SE, Johnson ZP, et al. 2014. A new rhesus macaque assembly and annotation for next-generation sequencing analyses. *Biol Direct* **9:** 20.

Zody MC, Jiang Z, Fung HC, Antonacci F, Hillier LW, Cardone MF, Graves TA, Kidd JM, Cheng Z, Abouelleil A, et al. 2008. Evolutionary toggling of the *MAPT* 17q21.31 inversion region. *Nat Genet* **40:** 1076–1083.