EVOLUTIONARY BIOLOGY

Lineage-specific accelerated sequences underlying primate evolution

Xupeng Bi[1]†, Long Zhou[1]†, Jin-Jin Zhang[2]†, Shaohong Feng[1,3], Mei Hu[2], David N. Cooper[4], Jiangwei Lin[2], Jiali Li[2], Dong-Dong Wu[2,5,6]*, Guojie Zhang[1,2,3,7]*

Understanding the mechanisms underlying phenotypic innovation is a key goal of comparative genomic studies. Here, we investigated the evolutionary landscape of lineage-specific accelerated regions (LinARs) across 49 primate species. Genomic comparison with dense taxa sampling of primate species significantly improved LinAR detection accuracy and revealed many novel human LinARs associated with brain development or disease. Our study also yielded detailed maps of LinARs in other primate lineages that may have influenced lineage-specific phenotypic innovation and adaptation. Functional experimentation identified gibbon LinARs, which could have participated in the developmental regulation of their unique limb structures, whereas some LinARs in the Colobinae were associated with metabolite detoxification which may have been adaptive in relation to their leaf-eating diet. Overall, our study broadens knowledge of the functional roles of LinARs in primate evolution.

INTRODUCTION

The genetic alterations underlying major phenotypic changes in nature remain largely unknown. Previous comparisons between human and other primate genomes have highlighted the high degree of similarity in protein-coding genes (*1*, *2*), which is hard to reconcile with the marked differences so evident between human and nonhuman primates. Thus, King and Wilson (*2*) proposed that mutations in gene regulatory elements might serve to explain this paradox. Regulatory elements play a central role in many biological processes by controlling the spatial and temporal expression patterns of specific genes. Many heritable components of common disease risk fall within noncoding regions and appear to be especially enriched in regulatory regions (*3*, *4*). Alterations of regulatory regions have fewer pleiotropic effects than changes in coding sequences and, hence, may be expected to be more likely to give rise to significant phenotypic effects (*5*, *6*). This is also exemplified by the evolution of lineage-specific phenotypes in many species resulting from changes to regulatory elements (*7*, *8*).

Whole-genome comparisons have allowed the identification of noncoding sequences exhibiting an elevated substitution rate across the genome in particular lineages (*9*), as well as thousands of human-accelerated regions (HARs) many of which have turned out to be important for brain development and cognitive functions (*10*, *11*). The high degree of conservation of genomic elements over an extended period of evolutionary time is indicative of the functional importance of these genomic regions, whereas their accelerated evolution observed in a specific lineage suggests that these regions might have once been under strong positive selection with advantageous substitutions occurring specifically in that lineage. A previous paper referred to these lineage-specific accelerated regions as LinARs and identified the LinARs in ape genomes (*12*). The LinARs in other primate lineages might also have contributed mechanistically to the long-term evolution of primates and the vast phenotypic diversity apparent in extant primate lineages. However, the extent of the roles that these LinARs have played remains largely unexplored.

The search for LinARs is usually performed on genomic regions that have maintained a high level of conservation over a considerable period of evolutionary time and involves the application of a statistical test that compares the substitution rate observed in a given lineage with the expected rate in the rest of the tree (*9*). Therefore, the detection power and accuracy of LinARs identified are critically reliant on both the total branch length of the tree, which can affect the false-positive rate for the background evolutionary constraint signal, and the taxon sampling of the closely related lineages that influence estimates of the lineage-specific substitution rate. Here, taking advantage of the available high-quality reference genomes from 49 primate species, including 27 species that were newly sequenced and assembled with long-reads technology, we identified the LinARs in all extant primate lineages and all key evolutionary nodes across the primates and explored the functional implications of these LinARs for primate diversification.

RESULTS

Improved detection power for accelerated conserved noncoding sequences

To obtain accelerated regions for clades of interest, we first identified the highly conserved regions with multiple genome alignments across species, and then scanned for regions showing faster

[1]Centre for Evolutionary & Organismal Biology, and Women's Hospital, Zhejiang University School of Medicine, Hangzhou 310058, China. [2]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China. [3]Liangzhu Laboratory, Zhejiang University Medical Center, 1369 West Wenyi Road, Hangzhou 311121, China. [4]Institute of Medical Genetics, School of Medicine, Cardiff University, Heath Park, Cardiff CF14 4XN, UK. [5]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, 32 Jiaochang Donglu, Kunming 650223, China. [6]National Resource Center for Non-Human Primates, Kunming Primate Research Center, and National Research Facility for Phenotypic & Genetic Analysis of Model Animals (Primate Facility), Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650107, China. [7]Villum Center for Biodiversity Genomics, Section for Ecology and Evolution, Department of Biology, University of Copenhagen, Copenhagen, Denmark.
*Corresponding author. Email: wudongdong@mail.kiz.ac.cn (D.-D.W.); guojiezhang@zju.edu.cn (G.Z.)
†These authors contributed equally to this work.

evolutionary rates than neutral rates among these conserved regions using the phyloP tool (*9*, *13*). As a result, we obtained approximately one million highly conserved elements (HCEs), together spanning ~5.5% of the human genome (fig. S1). By means of further comparison with other vertebrate genomes, we categorized these HCEs into vertebrate HCEs (29%), amniote-specific HCEs (4%), mammalian-specific HCEs (3%), and primate-specific HCEs (64%) (fig. S2). To identify the potential regulatory elements that might have contributed to lineage-specific evolution, we excluded the coding regions from these HCEs and used phyloP to screen for conserved loci that had experienced lineage-specific accelerated evolution in each evolutionary node of our primate phylogeny. In total, we identified 6.6% of HCEs that showed accelerated evolution in at least one primate branch (Fig. 1 and data S1). More than 77% of these LinARs have been found to overlap with long noncoding RNAs (lncRNAs) or candidate cis-regulatory elements (cCREs) (*14*, *15*), suggesting the functional significance of LinARs in primates.

Our updated list of LinARs in humans included 1674 regions (Fig. 2A), of which 892 sequences have been highly conserved across all vertebrates, 43 across sarcopterygians, 63 across amniotes, and 55 across mammals, whereas 621 are conserved but only in a primate-specific setting. We found that about one-third of our human LinARs (620 of 1674) overlapped with 3168 potential HARs previously identified in various sets of mammalian conserved elements (*16*–*23*). Further investigation of the alignments suggested that the disparity was mainly due to the power difference in detecting the background constraint for sequence conservation or accelerated evolution signals in humans consequent to different numbers of primate genomes being used for the analyses. Using a maximum of six primate species in the previous comparison provided insufficient power to detect the accelerated signals of the 929 human LinARs that were detected in our current comparison of 49 primate genomes (Fig. 2B). An additional 125 human LinARs were missed in previous studies because they were located within primate-specific conserved regions which were undetectable by analyses with only a few primate genomes for comparison (Fig. 2C). Conversely, various supposed HARs identified previously did not show up in our current dataset either because they turned out not to be conserved across the primates or because the level of acceleration in the human lineage was not significant when additional primate genomes were used as controls (fig. S3). More specifically, among those previously identified HARs, 2100 were artificial because the human-specific variations that detected with fewer species are in fact also appeared in other closely related primate lineages and the accelerated signal was no longer significant in human lineage when adding more primate species in analyses (fig. S4). In addition, 445 previous HARs were not detected now because these regions were not evolutionarily conserved across primate lineages (fig. S5). Moreover, a total of 66 previously detected HARs were artificial and had experienced an increased rate of evolution in the ancestral nodes of primates. Because of the low sample density for primates, the accelerated evolution signals in other evolutionary nodes could not be detected previously, but they were revealed in our analyses (fig. S6).

Human LinARs associated with brain development

It has been previously shown that many rapidly evolving regions in the human genome act as neurodevelopmental enhancers and play roles in neuronal gene regulatory programs (*11*). We found that 30% of the human LinARs located in cis-regulatory regions are functional elements annotated by the Encyclopedia of DNA Elements (ENCODE) database based on biochemical signatures of the genome (*15*). These human LinARs overlapped significantly with cis-regulatory elements active in the spinal cord ($P = 3.9 \times 10^{-34}$), eye ($P = 3.8 \times 10^{-31}$), and brain ($P = 5.7 \times 10^{-17}$) tissues (fig. S7). Genes associated with these LinARs were also significantly enriched in midbrain-hindbrain boundary development [Gene Ontology (GO): 0030917], rostrocaudal neural tube patterning (GO:0021903), and neuron recognition (GO:0008038) (data S2). For example, we found that the human LinAR (chr2: 235,865,048 to 235,865,565, rank23, coordinates are based on hg38) is nearest to gene *GBX2* [distance to transcription start site: 303,062 base pairs (bp)], a homeobox gene involved in the normal development of rhombomeres, segments of the developing mid/hindbrain region (data S3). During gastrulation and later stages of embryogenesis, *GBX2* is necessary not only for the early establishment of A/P (anterior/posterior) patterning in the neural plate but also for the normal development of the anterior hindbrain and the proper formation of the mid/hindbrain organizer in brain morphogenesis (*24*). Another human LinAR (chr3:2,660,287 to 2,660,641, rank140) is closest to gene *CNTN4* (distance to transcription start site: 421,635 bp) which encodes a cell adhesion molecule that supports the developing nervous system by promoting neurite outgrowth and axon guidance (*25*).

Our collection of human LinARs included many previously detected HARs including several that have been shown to play functional roles in the human cerebral cortex (data S4) (*11*). In particular, it also contains many novel human LinARs with potentially important biological functions that were not found by previous studies. For instance, among the top 20 most significant human LinARs, only one (chr20: 63,102,113 to 63,102,273, rank3) overlapped with the previously detected HAR genes *HAR1A* and *HAR1B*, which have been shown to be specifically expressed in Cajal-Retzius neurons and are essential for neocortex development (*26*). Most of these LinARs are overlapping with the lncRNA. Thus, the top eight human-LinARs are located within seven lncRNAs (data S5), including *HAR1A* and *HAR1B*. We performed in situ hybridization (ISH) in the human motor cortex and showed that the other five of these lncRNAs were highly expressed in the human motor cortex (Fig. 2D). In addition, we found that 129 newly identified human LinARs were placed near genes known to be associated with human inherited disease, of which 105 were associated with neurological disorders (data S6). Thus, our study has expanded the list of LinARs in the human lineage that may have contributed to the evolution of uniquely human features.

Accelerated evolution in nonhuman primate lineages

In addition to the human LinARs, which were highly enriched within brain developmental gene regulatory elements and detection of chemical stimuli involved in sensory perception, we noted that several other evolutionary lineages close to humans, such as the Homininae, Ponginae, and Hominidae great ape, also harbor LinARs which were found to be significantly associated with genes involved in sensory perception (a neurological process; data S2), implying that some of those LinARs residing in ancestral nodes may have contributed to the evolution of primate olfaction and other chemical senses (*27*). Gibbon (hylobatidae) LinARs tend to occur near genes involved in growth plate cartilage chondrocyte

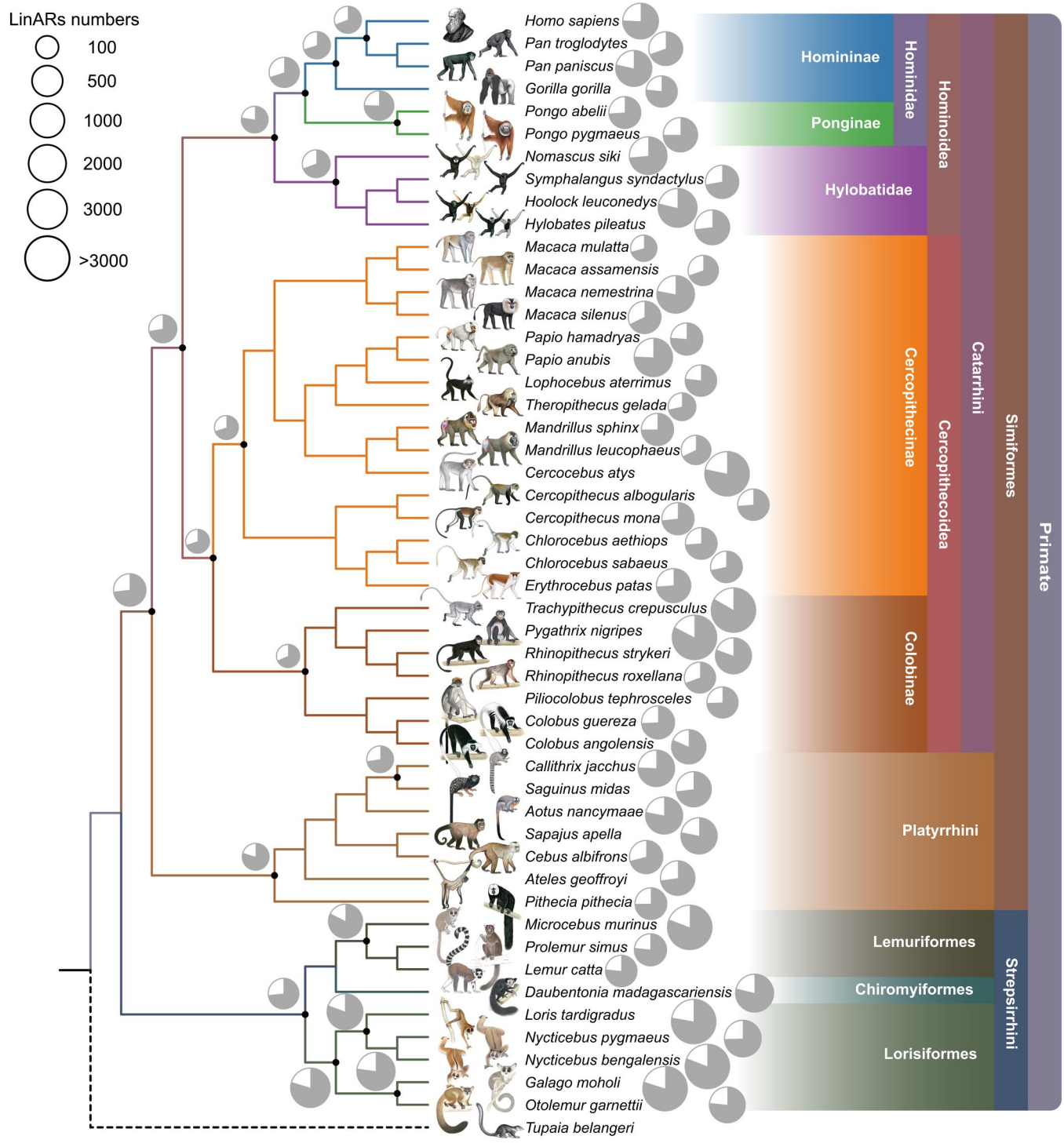

Fig. 1. Landscape and characteristics of lineage-specific accelerated regions (LinARs) in primates. The number distribution of LinARs in 18 major primate evolutionary nodes (suborders and families) and 49 extant species. More than 70% of LinARs in each lineage overlapped with long noncoding RNAs (lncRNAs) or human candidate cis-regulatory elements (cCREs). The size of the circles represents the number scale of the LinARs, whereas the gray area in the pie chart represents the overlap ratio with the lncRNA or cCREs. Species pictures are copyrighted by S. D. Nash/IUCN/SSC Primate Specialist Group and are used with their permission in this study.

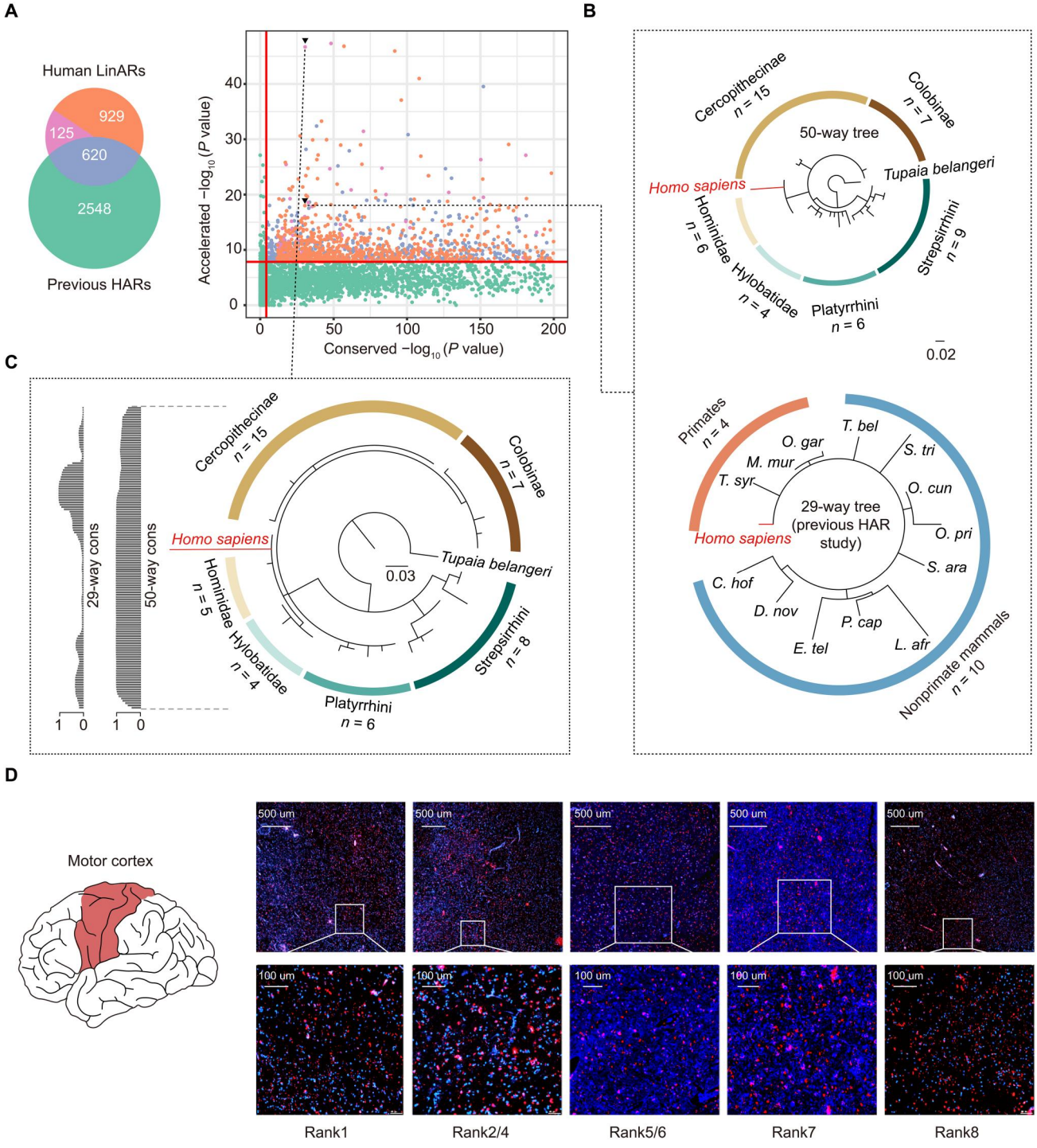

Fig. 2. Comparison between our identified human-accelerated regions and those of previous studies. (**A**) The Venn diagram depicts the overlap between the human LinARs identified in our study and previously detected HARs. The dot plot shows the distribution of conservation score *P* values across the primates and human-accelerated signal *P* values for human LinARs detected specifically in our study [orange corresponds to (**B**), pink to (**C**), green to previously identified HARs, and blue to shared regions]. The red lines represent the significance threshold of the *P* values (conserved: $P = 7.24 \times 10^{-5}$, accelerated: $P = 1.45 \times 10^{-8}$). (B) A case of our newly identified human LinAR (HLinAR-Rank114). The top panel shows the branch length of the neighbor-joining tree across primate species with humans having a significantly longer branch than other lineages. The bottom panel shows the branch length of the neighbor-joining tree across the mammals, the branch length in the human lineage being no different from those in other mammalian lineages (*20*). *n* represents the number of species. The scale bar denotes the mean number of nucleotide substitutions per site. (C) An example of our newly identified human LinAR (HLinAR-Rank18), which was not detected by previous studies because, although this region was specifically conserved in primates, it was not conserved across the mammals. (**D**) Results of in situ hybridization for the expression of five lncRNAs in the adult human cerebral cortex. The tissue was dissected from the motor cortex. The expression signals of all genes were detected and marked (red). The blue dot is 4′,6-diamidino-2-phenylindole staining, showing the location of the nucleus.

morphogenesis ($P = 1.7 \times 10{-}13$, data S2), which has been shown to be related to their characteristically long bone formation (*28*). Genes residing close to LinARs in Simiformes and Strepsirrhini are enriched in functions related to the regulation of dendrite morphogenesis in brain development (Simiformes: $P = 1.8 \times 10{-}10$, and Strepsirrhini: $P = 6.1 \times 10{-}5$). While the LinARs in Cercopithecinae are highly associated with genes in the immune system ($P = 2.1 \times 10{-}10$) and MHC class I protein complex in Hominidae ($P = 1.1 \times 10{-}12$). For example, human leukocyte antigen C (HLA-C) is an important major histocompatibility complex class I protein that has evolved only recently in the great ape lineage and exerts an influence on HIV disease progression (*29*). We identified a total of eight LinARs, located in the 4- to 17kb upstream regions of HLA-C genes (table S1). Polymorphic loci within these LinARs have been shown to affect innate immunity function in relation to viral infection, involving, for example, susceptibility to psoriasis in association with HIV-1 infection (*30*, *31*).

The role of two gibbon LinARs as regulatory elements in limb development

We next explored the potential importance of LinARs for phenotypic evolution and adaptation in other primate lineages. Gibbons are unique among all primate groups in terms of their special mode of locomotion involving skillful brachiation using their strong forelimbs to support more than 50% of their body weight during movement through the canopy (*32*). The advanced suspensory locomotion of gibbons has contributed to their specialized anatomical features, such as long forelimbs, well-developed scapular spine, and strong shoulder joints that allow a wide range of motion during rapid swinging (*33*). We detected several LinARs in gibbons that were associated with limb development genes, including *DLX5* and *EMX2*. Transgenic mice experiments have previously shown that the *EMX2* gene is involved in the formation of the scapula (*34*), while *DLX5* knockout mice are characterized by multiple developmental defects in their limb structures (*35*). The gibbon *DLX5* and *EMX2* genes both harbor a LinAR in their downstream regions (Fig. 3, A and B, and figs. S8 and S9). To test whether these LinARs have enhancer activity in vivo, we performed transgenic experiments using a mouse LacZ transgenic reporter assay. Notably, the transgene expression profiles demonstrated that these two elements alone were capable of activating lacZ activity in limbs and shoulder joints (Fig. 3C and fig. S10, A and B). Compared to the gibbon LinAR, we detected only weak signals of LacZ activity for its orthologs in humans and rhesus macaque (*Macaca mulatta*), suggesting that gibbon-specific changes in these regions do affect regulatory activity (fig. S10, C to F). Taxon-specific substitutions in these regulatory elements may have provided a mechanism to rewire the expression of the host genes and contribute to phenotypic innovation, although confirmation of this postulate will require further detailed examination of the functional consequences of the lineage-specific substitutions.

Influences of Colobines LinARs on diet trait

We further examined the functional role of the LinARs in another primate group, the Colobinae of the Old World monkeys, which have evolved highly specialized digestive systems as an adaptation to a predominantly leaf-eating diet; the Colobines represent the only primate group with foregut fermentation, allowing symbiotic bacteria to break down plant material in multichambered stomachs (*36*–*38*). Previous studies have suggested that the Colobinae have higher microbial diversity in their guts than their closely related sister group, the Cercopithecinae (*39*). The fecal microbiomes of *Rhinopithecus bieti* (a member of the Colobinae subfamily) have broad bacterial diversity and contain a large number of glycoside hydrolases responsible for lignocellulosic biomass degradation reflecting their adaptation to a diet rich in fibrous matter (*40*). We found that many of the LinARs that emerged in the Colobinae ancestral lineage were associated with genes related to metabolite detoxification (table S2). These include the *CYP4Z1* gene that encodes a member of the cytochrome P450 superfamily of enzymes that are important for metabolism and detoxification (*41*), and *ASL* that encodes argininosuccinate lyase, which catalyzes the reversible hydrolytic cleavage of argininosuccinate into arginine and fumarate, an essential step in the detoxification of ammonia in the liver (*42*).

Another example is *SOD1* which encodes an antioxidant enzyme that metabolizes superoxide radicals (*43*). *SOD1* also functions as an antimicrobial peptide that displays antibacterial, antifungal, and anti-methicillin-resistant *Staphylococcus aureus* (*MRSA*) activity (*44*). We identified a Colobinae LinAR located within an intronic region of *SOD1* (Fig. 4A). Although this element was highly conserved and under strong negative selection across other Simians, it accumulated multiple substitutions in the common ancestor of the Colobinae. It is of note that all these newly derived substitutions were subsequently fixed across all extant Colobinae species (Fig. 4A), implying that the neofunctionalization associated with these substitutions could have produced advantageous phenotypes that would have been rapidly subject to selective constraint in the Colobinae. The three-dimensional interaction data in six human tissues showed that this Colobinae LinAR and *SOD1* were located in the same topologically associating domain (fig. S11) (*45*). Transcriptome comparison indicated that the *SOD1* gene was widely expressed in many tissues but exhibits a lower level of activity specifically in the Colobine, black-and-white snub-nosed monkey (*R. bieti*), especially in the stomach, liver, and kidney, as compared to the rhesus macaque and humans (Fig. 4B). We further performed a dual-luciferase reporter gene experiment in human embryonic kidney (HEK) 293T cells and found that the *SOD1*-associated LinAR in *R. bieti* exhibited significantly lower promoter activity as compared with its counterparts in humans or rhesus macaque (Fig. 4C), suggesting that the low expression of *SOD1* in the Colobinae might have been driven by this LinAR. It is possible that the low *SOD1* activity in the gut of the Colobinae allows more microbes to survive, thereby assisting herbivorous digestion in this subfamily of primates. However, further confirmation in gastric cell lines will be useful in understanding the function of this LinAR.

DISCUSSION

Overall, our study has provided an evolutionary landscape of LinARs across the primates and broadened our understanding of the key roles played by evolutionarily accelerated regions in the acquisition of lineage-specific phenotypic innovations and adaptations. We identified hundreds of LinARs that were not found in previous studies. This suggests that increasing the phylogenetic density of genomic data by incorporating additional closely related primate species could significantly improve the detection efficiency and accuracy of LinARs, although other factors such as improving genome quality and sequence alignment quality may also

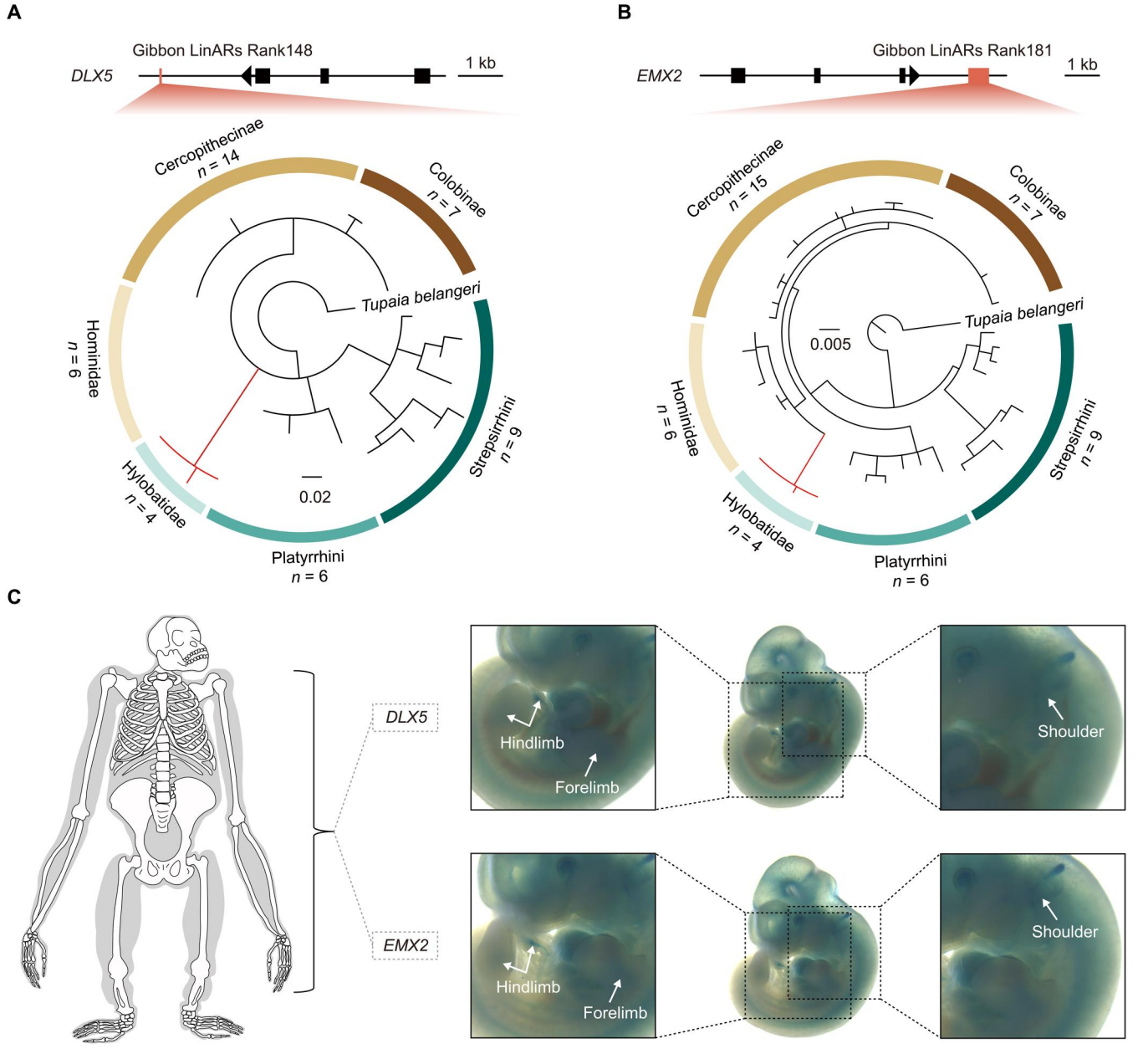

Fig. 3. Analysis of enhancer activity for two gibbon-specific accelerated regions in a lacZ reporter transgenic mouse assay. (**A** and **B**) Gibbon-specific accelerated regions downstream of two genes (*DLX5* and *EMX2*) associated with limb development. Black rectangles represent exons, orange rectangles represent rapidly evolving regions, and arrows represent the direction of transcription. The red lines in the phylogenetic tree represent the gibbon lineage. (**C**) Representative E11.5 transgenic embryos obtained for two gibbon LinAR (Rank148 and Rank181) reporters. Each construct shows three embryos resulting from independent transgene integration events. The images in the left-hand panels show close-up views of forelimb and hindlimb expressions in a representative embryo for each construct, with arrows indicating the positions where limb expression is present. Right-hand panels show zoom-in figures of the shoulder. See also fig. S10.

have contributed. In particular, our data allow the detection of rapidly evolving signals in a species or node within genomic regions that only experienced purifying selection constraints in the common ancestor of primates. This is important because these genomic regions may have contributed to the unique phenotypic features that have evolved in primates, such as the larger neocortex, advanced cognitive ability, and sensitive vision with depth perception (*46*). Many genes associated with the newly found human LinARs are functional in nervous system development. In a similar vein, the most critical LinARs function as lncRNAs which are expressed in the human brain, suggesting that these regions could have played a role in human-specific brain development. Thus, the updated list of human LinARs may contain loci associated with human-specific characteristics. Similarly, the LinARs we have described here in the gibbon and Colobinae lineages serve to demonstrate that the LinARs detected in other evolutionary

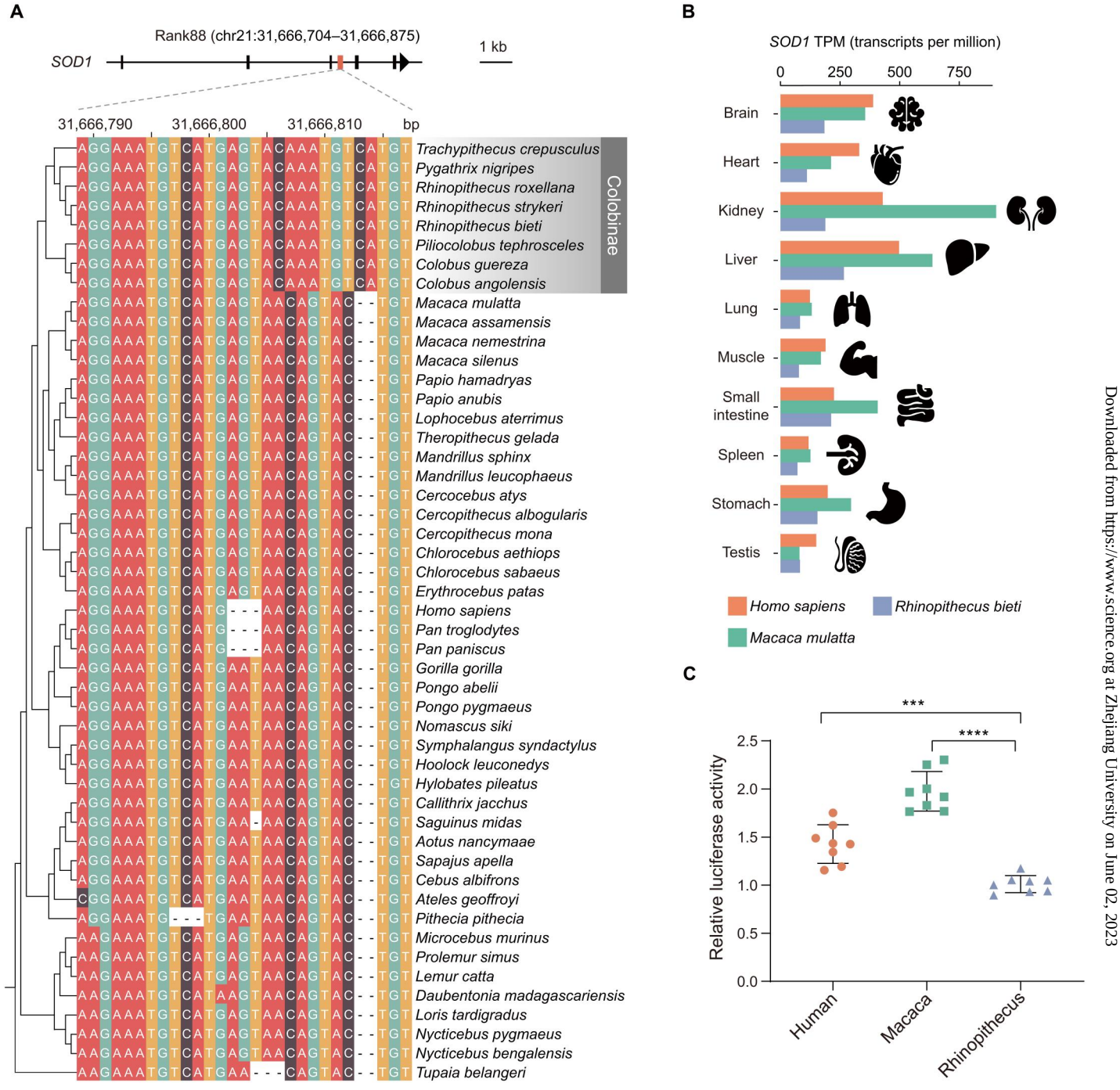

Fig. 4. Dual-luciferase reporter gene assay for the assessment of LinAR promoter activity. (**A**) Sequence alignment shows a Colobinae LinAR (Rank88) located within a *SOD1* intron. Black rectangles denote the exon-intron structure of *SOD1*, the orange rectangle represents the LinAR, and the arrow denotes the direction of transcription. (**B**) RNA sequencing data indicated that *SOD1* gene expression was decreased in many tissues from the black-and-white snub-nosed monkey (*R. bieti*) as compared with humans and macaques. The expression level was normalized with quantile normalization. (**C**) Relative luciferase activity of reporter vectors containing orthologous sequences of this LinAR in humans, *Macaca*, and *Rhinopithecus* was measured in human embryonic kidney cells. Eight replications were performed for each experimental group. ***$P < 0.001$; ****$P < 0.0001$ (*t* test, calculated by GraphPad Prism software).

nodes or primate species might also help to explain some of the phenotypic innovations and adaptations that have evolved in different primate lineages.

Our study has provided a valuable resource of LinARs across primate lineages. However, more in-depth experiments are now required to investigate in detail the regulatory functions of these elements. So far, a comprehensive annotation of regulatory elements has only been performed for the human genome, which has limited our understanding of the roles of LinARs in other primate species. The initiation of an ENCODE-like project, involving, at the very least, those primate species that have been widely used as research models, such as macaques and marmosets (*47*, *48*), will be extremely valuable in expanding our knowledge of the evolutionary roles of LinARs in these lineages. A high-throughput reporter assay system in cell lines would also be useful to investigate the functional types of these LinARs, i.e., to determine whether they play a role as enhancers or promoters. Although genetic modification remains challenging in most primate species, either due to technical obstacles or ethical reasons, the success in developing transgenic marmosets and rhesus monkeys has offered hope for further investigation of the biological roles of these LinARs during the developmental process of primates (*49*, *50*). Last, our study has only covered 10% of extant primate species. The sequencing of high-quality reference genomes from more primate species will ultimately provide a complete picture of LinARs during primate diversification and should shed new light on how their alteration has helped to shape primate biodiversity.

MATERIALS AND METHODS

Whole-genome alignments

The soft masked versions of 50 mammalian genome assemblies, including 49 primates and the northern tree shrew (*Tupaia belangeri*), were used to construct pairwise genome alignments with the human genome as a reference using Lastz v.1.04.00 (*51*) (the parameter set '--step=19 --hspthresh=2200 -inner=2000 --ydrop=3400 --gappedthresh=10000 --format=axt' and a score matrix for the comparison of distantly related species). Twenty-seven primate genomes were newly generated in our accompanying paper. The multiple whole-genome alignments (WGA) of 50 mammals (50-way WGA) were constructed using the program Multiz v.11.2 (*52*).

Detection of evolutionarily conserved elements

Before identifying the conserved elements, we first filtered the above multiple alignments: At least 90% of the species were required to be present in the 50-way WGA. The hidden Markov model–based method implemented in phastCons (a package of PHAST v.1.5, '--msa-format MAF --most-conserved –score') (*53*) was used to predict evolutionarily conserved elements in the 50-way multiple alignments. A phylogenetic tree model containing the evolutionary distance was constructed at neutrally evolving sites and a substitution rate matrix was drawn up using phyloFit (*54*) in the PHAST package v.1.5 with the general reversible substitution model from fourfold degenerate sites. As input, phastCons uses multiple alignments and a phylogenetic tree to predict HCEs. It produces base-by-base conservation scores and discrete conserved elements. The basic idea of the program is to scan along the alignment for regions that are better "explained" by the conserved model than by the nonconserved model. These genomic regions would then be exported as conserved elements. On the basis of the human annotation (Ensembl release 97), we classified these HCEs into coding sequences, 5′ and 3′ untranslated regions, RNA genes, pseudogenes, introns, and intergenic regions (fig. S1).

Lineage-specific accelerated regions

The statistical methods used to identify LinARs follow those of previous articles (*12*, *20*). LinARs were measured using the phyloP program (also from the PHAST package, '--msa-format MAF --branch --method LRT --mode CONACC') in 18 major primate evolutionary nodes (orders and families) and 49 extant species. The acceleration scores were generated by mode "ACC" (--mode CONACC) and were corrected to allow for multiple hypotheses using the false discovery rate (FDR; Benjamini and Hochberg) method (*55*). The significant acceleration at FDR-adjusted $P \leq 0.05$ was considered in further analyses (data S1). LinARs on the Y chromosome were removed. Annotation information for the lncRNAs and cCREs was obtained from RNAcentral (https://rnacentral.org/, release 19.0) (*14*) and the ENCODE database (https://screen.wenglab.org/) (*15*).

A lower total branch length of the phylogenetic tree may reduce the single base detection power, resulting in longer conserved elements. We found that some of the previously identified HARs were located in our HCE regions. However, in our phyloP test, owing to the longer conserved sequences used as input, these sites did not attain the threshold of accelerated evolution relative to that of a longer sequence background. Thus, we used phyloP to score these previous HARs in our 50-way WGA and found that 285 previous HARs were able to meet the threshold of significance (part of 620 overlapping human LinARs).

Classification of HCEs and human LinARs

We classified our HCEs and human LinARs in relation to vertebrates, sarcopterygians, amniotes, mammals, and primates using 100-way vertebrate alignment. First, phastCons was used to identify HCEs in vertebrates, sarcopterygians, amniotes, and mammals. Then, we overlapped our HCEs and human LinARs with 100-way HCEs from these four groups. The proportion of overlap was required to be more than 30%. Where our HCEs and human LinARs matched HCEs in only one taxon, we categorized these regions to that taxon. If they existed in two or more taxa, then we assigned them to their ancestral node (fig. S2). The set of 100-way vertebrate alignments for the human genome hg38 was downloaded from the UCSC website (http://hgdownload.cse.ucsc.edu/goldenPath/hg38/multiz100way/).

Human LinAR–related disease analysis

We analyzed our newly identified 1054 human LinARs in the context of human cCREs based on chromatin accessibility information generated by the ENCODE database (http://screen.encodeproject.org). By using BEDTools v2.29.2 to find the intersections of two data sets (*56*), we identified 199 human LinARs overlapping marks of cCREs that were activated in human brain samples. Then, we mapped these LinARs to chromatin conformation data (HiC) from human fetal cortex (GEO accession: GSM2054568, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM2054568) (*57*) yielding 167 LinARs with chromatin interactions for 154 genes within the same topologically associating domains in the fetal cortex. We used the UCSC liftOver tool to

convert genomic coordinates from Hg19 to Hg38 (https://genome.ucsc.edu/cgi-bin/hgLiftOver). Of the 122 (129 human LinARs) target genes corresponding to human inherited disease, 101 (105 human LinARs) genes were related to neurodevelopmental disorders recorded in the Human Gene Mutation Database (https://www.hgmd.cf.ac.uk).

Gene and tissue enrichment

To determine potential functions for LinARs in 15 major evolutionary nodes, we performed GO enrichment analysis on the Genomic Regions Enrichment of Annotations Tool (GREAT) website (http://great.stanford.edu/public/html/). The GREAT algorithm associates each human gene with a regulatory domain in the human genome (hg38) and calculates the total fraction of the genome annotated with GO terms. Statistical significance was assessed by means of the binomial test and hypergeometric test.

For tissue-type enrichment analysis, we classified human LinARs into specifically conserved primate ancestors or conserved mammalian ancestors according to whether the regions were conserved or not in the ancestral species. Regulatory elements were mapped to these accelerated regions, followed by a significant analysis of a number of overlaps. Cis-regulatory elements of humans were downloaded from the ENCODE database (https://screen.wenglab.org/) (*15*). *P* values were calculated by Fisher′s exact test and were adjusted for multiple comparisons using the FDR-controlling method of Benjamini and Hochberg in R software (p.adjust function, method = "fdr") (*55*).

In situ hybridization

Human RNA probes were constructed so as to correspond to five lncRNA sequences associated with human LinARs, and the RNA probes were labeled with biotin. First, the RNA probe sequence was designed. Then, the RNA probe and T7 transcript sequence were packaged into the plasmid, and the T7 transcript was used to reverse transcribe the RNA sequence in vitro, lastly combining the RNA sequence with biotin. The design and synthesis of RNA probes were completed by ShangHai MEIXUAN Biological Science and Technology Ltd. See table S3 for RNA probe sequences.

The human brain samples were derived from paraffin sections from the Li Jiali laboratory [Kunming Institute of Zoology, Chinese Academy of Sciences (CAS)], and all samples were taken from the cerebral cortices of ~70-year-old persons. Human autopsy paraffin-embedded 10-μm brain sections (tissue dissected from the motor cortex) from three individuals with no history of neurological or psychiatric disorders were used in this study. Tissues were graciously provided by the University of Pittsburgh Alzheimer′s Disease Research Center (ADRC) Brain Bank with approval from the internal review board of the Kunming Institute of Zoology, CAS. Additional frozen tissues were the generous gift of the ADRC at Washington University in St. Louis (Grant P50-AG-05681) with approval from the Ethics Committee of Kunming Institute of Zoology. Brain slices were used for ISH, and the work was performed at the Kunming Institute of Zoology, CAS being conducted under the regulations of the Human Biomedical Research Ethics Guidelines (regulated by the National Health Commission of the People′s Republic of China on 1 December 2016; approval number: 20170301). See table S4 for details of three case patients.

For hybridization, the brain slices were deparaffinized in xylene, hydrated with an alcohol gradient, and then washed in 0.1 M phosphate-buffered saline (PBS) for 10 min, 0.01 M glycine-PBS for 5 min, and 0.3% Triton X-100–PBS for 15 min, washed three times in 0.1 M PBS, and treated with proteinase K (1 μg/ml) for 30 min at 37°C. The slices were refixed in 4% paraformaldehyde (PFA) in PBS for 5 min, and then washed twice in 0.1 M PBS. The brain slices were then transferred to the prehybridization solution at 42°C for 30 min and incubated with the corresponding RNA probes at 42°C overnight. On the second day, the brain slices were first washed with hybridization buffer, and then with 4XSSC, 2XSSC, 1XSSC, 0.5XSSC, and 0.2XSSC at 37°C (20 min each time), and then washed with 0.1 M PBS: 0.2XSSC = 1:1 for 10 min at room temperature. After washing twice with 0.1 M PBS, the brain slices were incubated with streptavidin-Cy3 (Sigma-Aldrich, S6402) at a concentration of 1:200 in 0.3% Triton X-100–PBS for 3 hours at room temperature. Then, they were washed three times in 0.1 M PBS before adding 4′,6-diamidino-2-phenylindole (Invitrogen, P36931) and mounting the slide. Images were taken with a TissueFaxs cell analysis system (TissueGnostics GmbH, Austria).

LacZ in vivo reporter assay

BDF1 mouse (a cross between female C57BL/6NCrl and male DBA/2NCrl mice) care and experimental procedures were performed in compliance with the guidelines of the Institutional Animal Care and Use Committee (IACUC) of the Kunming Institute of Zoology, CAS (approval number: IACUC-RE-2021-12-001). Two gibbon LinARs were detected, which may be related to the development of bones, limbs, and tail. The two LinAR sequences (Rank148 and Rank181) of a gibbon were cloned into the Hsp68-lacZ reporter vector that was completed by the VectorBuilder platform. Transgenic mouse embryos were generated by pronuclear injection, and F0 embryos were collected at E11.5 and stained for LacZ activity (*58*). Before injection, plasmid DNA was linearized with *Xmn*I or *Ahd*I, followed by purification with the Gel Extraction Kit (Omega, D2500-01). BDF1 and Institute of Cancer Research (ICR) mouse strains were used as embryo donors and foster mothers, respectively. Superovulated female BDF1 mice (7 to 8 weeks old) were mated to BDF1 stud males, and fertilized embryos were collected from oviducts. The DNA was diluted to a final concentration of 2 ng/μl and injected into BDF1 zygotes, and HCZB medium was used for injection. The numbers of injected zygotes were 117 and 144, respectively. The injected zygotes were cultured in KSOM (four Yamanaka factors—*OCT4*, *KLF4*, *SOX2*, and *MYC*) culture medium at 37°C under 5% CO_2 and 5% O_2 in the air for approximately 15 hours. Thereafter, zygotes were transferred into the uteruses of ICR females. Embryos were harvested at embryonic day 11.5 in cold PBS, followed by 30 min of incubation with 4% PFA. The embryos were washed three times for 30 min with embryo wash buffer [2 mM $MgCl_2$, 0.01% deoxycholate, 0.02% NP-40, and 0.1 M PBS buffer (pH 7.3)]. LacZ activity was detected by incubating with freshly made staining solution [1 mg/ml X-gal (Sigma-Aldrich, V900468), 5 mM potassium ferrocyanide, 5 mM potassium ferricyanide, and 5 mM EGTA (pH 7.5) in wash buffer] at 37°C for a few hours to overnight until the desired staining was achieved. Following staining, embryos were washed in lacZ buffer and stored at 4°C in 4% PFA in PBS. To be considered reproducible, bone, limb, and tail expression patterns had to be observed in at least three embryos (*59*). Embryos were imaged using a Nikon SMZ18 stereo microscope.

SOD1 gene expression

We downloaded RNA sequencing (RNA-seq) data from different tissues of *R. bieti* from the National Center for Biotechnology Information (NCBI) website (PRJNA248058). The genome and gene set data derived from *R. bieti* were also downloaded from NCBI (PRJNA339282, RefSeq assembly accession: GCF_001698545.2). We mapped the RNA-seq reads for each tissue to the reference genome using HISAT2 (http://github.com/infphilo/hisat2, version 2.1.0; hisat2 -x species.index −1 read1−2 read2 -S read.sam -q -p 4 --phred33 --dta-cufflinks --rna-strandness RF) (*60*). Then, the SAM files were converted to BAM files and sorted by SAMTools (http://samtools.sourceforge.net, version 1.4.1; samtools view -bhS read.sam -o read.bam; samtools sort read.bam -o read.sort.bam). These BAM files were used as input to calculate transcripts per million (TPM) for each tissue by TPMCalculator (https://github.com/ncbi/TPMCalculator, version 0.0.3; TPMCalculator -g species.gtf -b read.sort.bam -p -e). TPM data for humans and *M. mulatta* were downloaded from Gtex (https://gtexportal.org/home/datasets) and the EBI database (https://www.ebi.ac.uk/gxa/experiments/E-MTAB-2799/Downloads). Stomach TPM data for *M. mulatta* were obtained from the accompanying article. All gene expression data were subjected to quantile normalization using the R package preprocessCore {normalize.quantiles [as.matrix (TPM)]}. The Colobinae LinAR (chr21:31,666,703-31,666,875, rank88) and *SOD1* were located in the same topologically associating domain, which was supported by six tissues in human HiC data from PENNSTATE website (http://3dgenome.fsm.northwestern.edu/view.php) (*45*).

Dual-luciferase reporter experiment

HEK293T cell lines were cultured in Dulbecco′s modified Eagle′s medium. All media were supplemented with 10% heat-inactivated fetal bovine serum (Life Technologies BRL), and the cell lines were maintained in a 5% CO_2-humidified atmosphere at 37°C.

HEK293T cells were seeded in the wells of a 24-well plate at a density of 1×10^4 cells per well 1 day before transfection. The transfection reagent ViaFect (E4981, Promega) was used according to the manufacturer′s instructions. A Renilla-luciferase plasmid (PRL-TK) was cotransfected to control for transfection efficiency. SOD1-pGL3-Basic (SOD1 promoter sequences from humans, *Macaca*, and *Rhinopithecus*), and PRL-TK were then cotransfected at a total concentration of 500 ng per well (4:1 molar ratio of SOD1-pGL3-Basic and PRL-TK). Luciferase assays were performed 48 hours after transfection using the Dual-Glo Luciferase Assay System (E2920, Promega) according to the manufacturer′s instructions.

Supplementary Materials

This PDF file includes:
Figs. S1 to S11
Tables S1 to S4
Legends for data S1 to S6

Other Supplementary Material for this manuscript includes the following:
Data file S1 to S6

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. The Chimpanzee Sequencing and Analysis Consortium, Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87 (2005).
2. M. C. King, A. C. Wilson, Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
3. M. T. Maurano, R. Humbert, E. Rynes, R. E. Thurman, E. Haugen, H. Wang, A. P. Reynolds, R. Sandstrom, H. Qu, J. Brody, A. Shafer, F. Neri, K. Lee, T. Kutyavin, S. Stehling-Sun, A. K. Johnson, T. K. Canfield, E. Giste, M. Diegel, D. Bates, R. S. Hansen, S. Neph, P. J. Sabo, S. Heimfeld, A. Raubitschek, S. Ziegler, C. Cotsapas, N. Sotoodehnia, I. Glass, S. R. Sunyaev, R. Kaul, J. A. Stamatoyannopoulos, Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
4. K. K.-H. Farh, A. Marson, J. Zhu, M. Kleinewietfeld, W. J. Housley, S. Beik, N. Shoresh, H. Whitton, R. J. H. Ryan, A. A. Shishkin, M. Hatan, M. J. Carrasco-Alfonso, D. Mayer, C. J. Luckey, N. A. Patsopoulos, P. L. de Jager, V. K. Kuchroo, C. B. Epstein, M. J. Daly, D. A. Hafler, B. E. Bernstein, Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2015).
5. G. A. Wray, The evolutionary significance of *cis*-regulatory mutations. *Nat. Rev. Genet.* **8**, 206–216 (2007).
6. D. L. Stern, V. Orgogozo, The loci of evolution: How predictable is genetic evolution? *Evol. Int. J. Org. Evol.* **62**, 2155–2177 (2008).
7. R. Bachner-Melman, C. Dina, A. H. Zohar, N. Constantini, E. Lerer, S. Hoch, S. Sella, L. Nemanov, I. Gritsenko, P. Lichtenberg, R. Granot, R. P. Ebstein, *AVPR1a* and *SLC6A4* gene polymorphisms are associated with creative dance performance. *PLOS Genet.* **1**, e42 (2005).
8. E. A. Hammock, L. J. Young, Microsatellite instability generates diversity in brain and sociobehavioral traits. *Science* **308**, 1630–1634 (2005).
9. K. S. Pollard, M. J. Hubisz, K. R. Rosenbloom, A. Siepel, Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).
10. R. N. Doan, B. I. Bae, B. Cubelos, C. Chang, A. A. Hossain, S. al-Saad, N. M. Mukaddes, O. Oner, M. al-Saffar, S. Balkhy, G. G. Gascon; Homozygosity Mapping Consortium for Autism, M. Nieto, C. A. Walsh, Mutations in human accelerated regions disrupt cognition and social behavior. *Cell* **167**, 341–354.e12 (2016).
11. K. M. Girskis, A. B. Stergachis, E. M. DeGennaro, R. N. Doan, X. Qian, M. B. Johnson, P. P. Wang, G. M. Sejourne, M. A. Nagy, E. A. Pollina, A. M. M. Sousa, T. Shin, C. J. Kenny, J. L. Scotellaro, B. M. Debo, D. M. Gonzalez, L. M. Rento, R. C. Yeh, J. H. T. Song, M. Beaudin, J. Fan, P. V. Kharchenko, N. Sestan, M. E. Greenberg, C. A. Walsh, Rewiring of human neurodevelopmental gene regulatory programs by human accelerated regions. *Neuron* **109**, 3239–3251.e7 (2021).
12. D. Kostka, A. K. Holloway, K. S. Pollard, Developmental loci harbor clusters of accelerated regions that evolved independently in ape lineages. *Mol. Biol. Evol.* **35**, 2034–2045 (2018).
13. M. J. Hubisz, K. S. Pollard, A. Siepel, PHAST and RPHAST: Phylogenetic analysis with space/time models. *Brief. Bioinform.* **12**, 41–51 (2011).
14. R. Consortium, RNAcentral 2021: Secondary structure integration, improved sequence search and new member databases. *Nucleic Acids Res.* **49**, D212–D220 (2020).
15. The ENCODE Project Consortium, J. E. Moore, M. J. Purcaro, H. E. Pratt, C. B. Epstein, N. Shoresh, J. Adrian, T. Kawli, C. A. Davis, A. Dobin, R. Kaul, J. Halow, E. L. Van Nostrand, P. Freese, D. U. Gorkin, Y. Shen, Y. He, M. Mackiewicz, F. Pauli-Behn, B. A. Williams, A. Mortazavi, C. A. Keller, X.-O. Zhang, S. I. Elhajjajy, J. Huey, D. E. Dickel, V. Snetkova, X. Wei, X. Wang, J. C. Rivera-Mulia, J. Rozowsky, J. Zhang, S. B. Chhetri, J. Zhang, A. Victorsen, K. P. White, A. Visel, G. W. Yeo, C. B. Burge, E. Lécuyer, D. M. Gilbert, J. Dekker, J. Rinn, E. M. Mendenhall, J. R. Ecker, M. Kellis, R. J. Klein, W. S. Noble, A. Kundaje, R. Guigó, P. J. Farnham, J. M. Cherry, R. M. Myers, B. Ren, B. R. Graveley, M. B. Gerstein, L. A. Pennacchio, M. P. Snyder, B. E. Bernstein, B. Wold, R. C. Hardison, T. R. Gingeras, J. A. Stamatoyannopoulos, Z. Weng, Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).
16. K. S. Pollard, S. R. Salama, B. King, A. D. Kern, T. Dreszer, S. Katzman, A. Siepel, J. S. Pedersen, G. Bejerano, R. Baertsch, K. R. Rosenbloom, J. Kent, D. Haussler, Forces shaping the fastest evolving regions in the human genome. *PLOS Genet.* **2**, e168 (2006).
17. S. Prabhakar, J. P. Noonan, S. Pääbo, E. M. Rubin, Accelerated evolution of conserved noncoding sequences in humans. *Science* **314**, 786–786 (2006).
18. C. P. Bird, B. E. Stranger, M. Liu, D. J. Thomas, C. E. Ingle, C. Beazley, W. Miller, M. E. Hurles, E. T. Dermitzakis, Fast-evolving noncoding sequences in the human genome. *Genome Biol.* **8**, 1–12 (2007).
19. E. C. Bush, B. T. Lahn, A genome-wide screen for noncoding elements important in primate evolution. *BMC Evol. Biol.* **8**, 1–10 (2008).
20. K. Lindblad-Toh, M. Garber, O. Zuk, M. F. Lin, B. J. Parker, S. Washietl, P. Kheradpour, J. Ernst, G. Jordan, E. Mauceli, L. D. Ward, C. B. Lowe, A. K. Holloway, M. Clamp, S. Gnerre, J. Alföldi, K. Beal, J. Chang, H. Clawson, J. Cuff, F. di Palma, S. Fitzgerald, P. Flicek, M. Guttman, M. J. Hubisz, D. B. Jaffe, I. Jungreis, W. J. Kent, D. Kostka, M. Lara, A. L. Martins,

T. Massingham, I. Moltke, B. J. Raney, M. D. Rasmussen, J. Robinson, A. Stark, A. J. Vilella, J. Wen, X. Xie, M. C. Zody; Broad Institute Sequencing Platform and Whole Genome Assembly Team, J. Baldwin, T. Bloom, C. W. Chin, D. Heiman, R. Nicol, C. Nusbaum, S. Young, J. Wilkinson, K. C. Worley, C. L. Kovar, D. M. Muzny, R. A. Gibbs; Baylor College of Medicine Human Genome Sequencing Center Sequencing Team, A. Cree, H. H. Dihn, G. Fowler, S. Jhangiani, V. Joshi, S. Lee, L. R. Lewis, L. V. Nazareth, G. Okwuonu, J. Santibanez, W. C. Warren, E. R. Mardis, G. M. Weinstock, R. K. Wilson; Genome Institute at Washington University, K. Delehaunty, D. Dooling, C. Fronik, L. Fulton, B. Fulton, T. Graves, P. Minx, E. Sodergren, E. Birney, E. H. Margulies, J. Herrero, E. D. Green, D. Haussler, A. Siepel, N. Goldman, K. S. Pollard, J. S. Pedersen, E. S. Lander, M. Kellis, A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476–482 (2011).

21. J. A. Capra, G. D. Erwin, G. McKinsey, J. L. Rubenstein, K. S. Pollard, Many human accelerated regions are developmental enhancers. *Philos. Trans. R. Soc. B Biol. Sci.* **368**, 20130025 (2013).
22. M. J. Hubisz, K. S. Pollard, Exploring the genesis and functions of human accelerated regions sheds light on their role in human evolution. *Curr. Opin. Genet. Dev.* **29**, 15–21 (2014).
23. R. M. Gittelman, E. Hun, F. Ay, J. Madeoy, L. Pennacchio, W. S. Noble, R. D. Hawkins, J. M. Akey, Comprehensive identification and analysis of human accelerated regulatory DNA. *Genome Res.* **25**, 1245–1255 (2015).
24. S. T. Waters, M. Lewandoski, A threshold requirement for Gbx2 levels in hindbrain development. *Development* **133**, 1991–2000 (2006).
25. T. Kaneko-Goto, S. Yoshihara, H. Miyazaki, Y. Yoshihara, BIG-2 mediates olfactory axon convergence to target glomeruli. *Neuron* **57**, 834–846 (2008).
26. K. S. Pollard, S. R. Salama, N. Lambert, M. A. Lambot, S. Coppens, J. S. Pedersen, S. Katzman, B. King, C. Onodera, A. Siepel, A. D. Kern, C. Dehay, H. Igel, M. Ares Jr., P. Vanderhaeghen, D. Haussler, An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* **443**, 167–172 (2006).
27. N. J. Dominy, C. F. Ross, T. D. Smith, Evolution of the special senses in primates: Past, present, and future. *Anat. Rec. A Discov. Mol. Cell. Evol. Biol.* **281**, 1078–1082 (2004).
28. Y. Li, V. Trivedi, T. V. Truong, D. S. Koos, R. Lansford, C. M. Chuong, D. Warburton, R. A. Moats, S. E. Fraser, Dynamic imaging of the growth plate cartilage reveals multiple contributors to skeletal morphogenesis. *Nat. Commun.* **6**, 6798 (2015).
29. C. M. Heijmans, N. G. de Groot, R. E. Bontrop, Comparative genetics of the major histocompatibility complex in humans and nonhuman primates. *Int. J. Immunogenet.* **47**, 243–260 (2020).
30. K.-Y. Lee, K.-S. Leung, N. L. Tang, M.-H. Wong, Discovering genetic factors for psoriasis through exhaustively searching for significant second order SNP-SNP interactions. *Sci. Rep.* **8**, 1–11 (2018).
31. International HIV Controllers Study, The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science* **330**, 1551–1557 (2010).
32. F. Michilsens, E. E. Vereecke, K. D'août, P. Aerts, Functional anatomy of the gibbon forelimb: Adaptations to a brachiating lifestyle. *J. Anat.* **215**, 335–354 (2009).
33. M. S. Selby, C. O. Lovejoy, Evolution of the hominoid scapula and its implications for earliest hominid locomotion. *Am. J. Phys. Anthropol.* **162**, 682–700 (2017).
34. F. Pröls, F. Ehehalt, M. Rodriguez-Niedenführ, L. He, R. Huang, B. Christ, The role of Emx2 during scapula formation. *Dev. Biol.* **275**, 315–324 (2004).
35. R. F. Robledo, L. Rajan, X. Li, T. Lufkin, The Dlx5 and Dlx6 homeobox genes are essential for craniofacial, axial, and appendicular skeletal development. *Genes Dev.* **16**, 1089–1101 (2002).
36. D. J. Chivers, *Functional Anatomy of the Gastrointestinal Tract. Colobine Monkeys: Their Ecology, Behaviour and Evolution* (Cambridge Univ. Press, 1994), pp. 205–227.
37. J. E. Lambert, Primate digestion: Interactions among anatomy, physiology, and feeding ecology. *Evol. Anthropol.* **7**, 8–20 (1998).
38. T. Hayakawa, S. K. S. S. Nathan, D. J. Stark, D. A. R. Saldivar, R. Sipangkui, B. Goossens, A. Tuuga, M. Clauss, A. Sawada, S. Fukuda, H. Imai, I. Matsuda, First report of foregut microbial community in proboscis monkeys: Are diverse forests a reservoir for diverse microbiomes? *Environ. Microbiol. Rep.* **10**, 655–662 (2018).
39. Z. Huan, Y. Yao, J. Yu, H. Chen, M. Li, C. Yang, B. Zhao, Q. Ni, M. Zhang, M. Xie, H. Xu, Differences in the gut microbiota between Cercopithecinae and Colobinae in captivity. *J. Microbiol.* **58**, 367–376 (2020).
40. B. Xu, W. Xu, J. Li, L. Dai, C. Xiong, X. Tang, Y. Yang, Y. Mu, J. Zhou, J. Ding, Q. Wu, Z. Huang, Metagenomic analysis of the *Rhinopithecus bieti* fecal microbiome reveals a broad diversity of bacterial and glycoside hydrolase profiles related to lignocellulose degradation. *BMC Genomics* **16**, 1–11 (2015).
41. J. H. Thomas, Rapid birth—Death evolution specific to xenobiotic cytochrome P450 genes in vertebrates. *PLOS Genet.* **3**, e67 (2007).
42. C. Balmer, A. V. Pandey, V. Rüfenacht, J. M. Nuoffer, P. Fang, L. J. Wong, J. Häberle, Mutations and polymorphisms in the human *Argininosuccinate Lyase* (ASL) gene. *Gene* **35**, 27–35 (2014).
43. Y. Nojima, K. Ito, H. Ono, T. Nakazato, H. Bono, T. Yokoyama, R. Sato, Y. Suetsugu, Y. Nakamura, K. Yamamoto, J. I. Satoh, H. Tabunoki, H. Fugo, Superoxide dismutases, SOD1 and SOD2, play a distinct role in the fat body during pupation in silkworm *Bombyx mori*. *PLOS ONE* **10**, e0116007 (2015).
44. M. Azkargorta, M. Bregón-Villahoz, I. Escobes, J. Ibáñez-Pérez, I. Iloro, M. Iglesias, M. Diez-Zapirain, A. Rabanal, B. Prieto, M. D. Moragues, R. Matorras, F. Elortza, In-depth proteomics and natural peptidomics analyses reveal antibacterial peptides in human endometrial fluid. *J. Proteomics* **216**, 103652 (2020).
45. Y. Wang, F. Song, B. Zhang, L. Zhang, J. Xu, D. Kuang, D. Li, M. N. K. Choudhary, Y. Li, M. Hu, R. Hardison, T. Wang, F. Yue, The 3D Genome Browser: A web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol.* **19**, 151 (2018).
46. R. A. Barton, Visual specialization and brain evolution in primates. *Proc. Royal Soc. Lond. Ser. B Biol. Sci.* **265**, 1933–1937 (1998).
47. Rhesus Macaque Genome Sequencing and Analysis Consortium, R. A. Gibbs, J. Rogers, M. G. Katze, R. Bumgarner, G. M. Weinstock, E. R. Mardis, K. A. Remington, R. L. Strausberg, J. C. Venter, R. K. Wilson, M. A. Batzer, C. D. Bustamante, E. E. Eichler, M. W. Hahn, R. C. Hardison, K. D. Makova, W. Miller, A. Milosavljevic, R. E. Palermo, A. Siepel, J. M. Sikela, T. Attaway, S. Bell, K. E. Bernard, C. J. Buhay, M. N. Chandrabose, M. Dao, C. Davis, K. D. Delehaunty, Y. Ding, H. H. Dinh, S. Dugan-Rocha, L. A. Fulton, R. A. Gabisi, T. T. Garner, J. Godfrey, A. C. Hawes, J. Hernandez, S. Hines, M. Holder, J. Hume, S. N. Jhangiani, V. Joshi, Z. M. Khan, E. F. Kirkness, A. Cree, R. G. Fowler, S. Lee, L. R. Lewis, Z. Li, Y.-S. Liu, S. M. Moore, D. Muzny, L. V. Nazareth, D. N. Ngo, G. O. Okwuonu, G. Pai, D. Parker, H. A. Paul, C. Pfannkoch, C. S. Pohl, Y.-H. Rogers, S. J. Ruiz, A. Sabo, J. Santibanez, B. W. Schneider, S. M. Smith, E. Sodergren, A. F. Svatek, T. R. Utterback, S. Vattathil, W. Warren, C. S. White, A. T. Chinwalla, Y. Feng, A. L. Halpern, L. W. Hillier, X. Huang, P. Minx, J. O. Nelson, K. H. Pepin, X. Qin, G. G. Sutton, E. Venter, B. P. Walenz, J. W. Wallis, K. C. Worley, S.-P. Yang, S. M. Jones, M. A. Marra, M. Rocchi, J. E. Schein, R. Baertsch, L. Clarke, M. Csürös, J. Glasscock, R. A. Harris, P. Havlak, A. R. Jackson, H. Jiang, Y. Liu, D. N. Messina, Y. Shen, H. X.-Z. Song, T. Wylie, L. Zhang, E. Birney, K. Han, M. K. Konkel, J. Lee, A. F. A. Smit, B. Ullmer, H. Wang, J. Xing, R. Burhans, Z. Cheng, J. E. Karro, J. Ma, B. Raney, X. She, M. J. Cox, J. P. Demuth, L. J. Dumas, S.-G. Han, J. Hopkins, A. Karimpour-Fard, Y. H. Kim, J. R. Pollack, T. Vinar, C. Addo-Quaye, J. Degenhardt, A. Denby, M. J. Hubisz, A. Indap, C. Kosiol, B. T. Lahn, H. A. Lawson, A. Marklein, R. Nielsen, E. J. Vallender, A. G. Clark, B. Ferguson, R. D. Hernandez, K. Hirani, H. Kehrer-Sawatzki, J. Kolb, S. Patil, L.-L. Pu, Y. Ren, D. G. Smith, D. A. Wheeler, I. Schenck, E. V. Ball, R. Chen, D. N. Cooper, B. Giardine, F. Hsu, W. J. Kent, A. Lesk, D. L. Nelson, W. E. O'brien, K. Prüfer, P. D. Stenson, J. C. Wallace, H. Ke, X.-M. Liu, P. Wang, A. P. Xiang, F. Yang, G. P. Barber, D. Haussler, D. Karolchik, A. D. Kern, R. M. Kuhn, K. E. Smith, A. S. Zwieg, Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316**, 222–234 (2007).
48. C. Yang, Y. Zhou, S. Marcus, G. Formenti, L. A. Bergeron, Z. Song, X. Bi, J. Bergman, M. M. C. Rousselle, C. Zhou, L. Zhou, Y. Deng, M. Fang, D. Xie, Y. Zhu, S. Tan, J. Mountcastle, B. Haase, J. Balacco, J. Wood, W. Chow, A. Rhie, M. Pippel, M. M. Fabiszak, S. Koren, O. Fedrigo, W. A. Freiwald, K. Howe, H. Yang, A. M. Phillippy, M. H. Schierup, E. D. Jarvis, G. Zhang, Evolutionary and biomedical insights from a marmoset diploid genome assembly. *Nature* **594**, 227–233 (2021).
49. Y. Niu, Y. Yu, A. Bernat, S. Yang, X. He, X. Guo, D. Chen, Y. Chen, S. Ji, W. Si, Y. Lv, T. Tan, Q. Wei, H. Wang, L. Shi, J. Guan, X. Zhu, M. Afanassieff, P. Savatier, K. Zhang, Q. Zhou, W. Ji, Transgenic rhesus monkeys produced by gene transfer into early-cleavage–stage embryos using a simian immunodeficiency virus-based vector. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 17663–17667 (2010).
50. E. Sasaki, H. Suemizu, A. Shimada, K. Hanazawa, R. Oiwa, M. Kamioka, I. Tomioka, Y. Sotomaru, R. Hirakawa, T. Eto, S. Shiozawa, T. Maeda, M. Ito, R. Ito, C. Kito, C. Yagihashi, K. Kawai, H. Miyoshi, Y. Tanioka, N. Tamaoki, S. Habu, H. Okano, T. Nomura, Generation of transgenic non-human primates with germline transmission. *Nature* **459**, 523–527 (2009).
51. R. S. Harris, *Improved Pairwise Alignment of Genomic DNA* (The Pennsylvania State Univ., 2007).
52. M. Blanchette, W. J. Kent, C. Riemer, L. Elnitski, A. F. A. Smit, K. M. Roskin, R. Baertsch, K. Rosenbloom, H. Clawson, E. D. Green, D. Haussler, W. Miller, Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* **14**, 708–715 (2004).
53. A. Siepel, G. Bejerano, J. S. Pedersen, A. S. Hinrichs, M. Hou, K. Rosenbloom, H. Clawson, J. Spieth, L. D. W. Hillier, S. Richards, G. M. Weinstock, R. K. Wilson, R. A. Gibbs, W. J. Kent, W. Miller, D. Haussler, Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
54. A. Siepel, D. Haussler, Phylogenetic estimation of context-dependent substitution rates by maximum likelihood. *Mol. Biol.* **21**, 468–488 (2003).

55. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B. Methodol.* **57**, 289–300 (1995).
56. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
57. H. Won, L. de la Torre-Ubieta, J. L. Stein, N. N. Parikshak, J. Huang, C. K. Opland, M. J. Gandal, G. J. Sutton, F. Hormozdiari, D. Lu, C. Lee, E. Eskin, I. Voineagu, J. Ernst, D. H. Geschwind, Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523–527 (2016).
58. R. Kothary, S. Clapoff, S. Darling, M. D. Perry, L. A. Moran, J. Rossant, Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice. *Development* **105**, 707–714 (1989).
59. A. Visel, S. Minovitsky, I. Dubchak, L. A. Pennacchio, VISTA Enhancer Browser—A database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88–D92 (2007).
60. D. Kim, B. Langmead, S. L. Salzberg, HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).

Acknowledgments: We thank all supervisors, collaborators, data analysts, and anyone else involved with the collection and processing of the primary datasets. **Funding:** This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB31020000), International Partnership Program of Chinese Academy of Sciences (no. 152453KYSB20170002), and a Villum Investigator grant (no. 25900) to G.Z. This work was also supported by the National Natural Science Foundation of China (31822048), Yunnan Fundamental Research Project (2019FI010), and the Animal Branch of the Germplasm Bank of Wild Species of Chinese Academy of Science (the Large Research Infrastructure Funding) to D.-D.W. **Author contributions:** G.Z. and D.-D.W. conceived the project and designed the experiments. X.B., L.Z., and S.F. analyzed the data. J.-J.Z., M.H., J.Lin, and J.Li performed the experiments. D.N.C. provided HGMD data and revised the manuscript. G.Z., D.-D.W., X.B., and L.Z. wrote the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 11 May 2022
Accepted 5 May 2023
Published 1 June 2023
10.1126/sciadv.adc9507

ScienceAdvances

Lineage-specific accelerated sequences underlying primate evolution

Xupeng Bi, Long Zhou, Jin-Jin Zhang, Shaohong Feng, Mei Hu, David N. Cooper, Jiangwei Lin, Jiali Li, Dong-Dong Wu, and Guojie Zhang

Sci. Adv., **9** (22), eadc9507.
DOI: 10.1126/sciadv.adc9507

Science Advances (ISSN) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.