Whole-genome sequence of the Tibetan frog *Nanorana parkeri* and the comparative evolution of tetrapod genomes

Yan-Bo Sun[a,1], Zi-Jun Xiong[b,c,d,1], Xue-Yan Xiang[b,c,d,e,1], Shi-Ping Liu[b,c,d,f], Wei-Wei Zhou[a], Xiao-Long Tu[a,g], Li Zhong[h], Lu Wang[h], Dong-Dong Wu[a], Bao-Lin Zhang[a,h], Chun-Ling Zhu[a], Min-Min Yang[a], Hong-Man Chen[a], Fang Li[b,d], Long Zhou[b,d], Shao-Hong Feng[b,d], Chao Huang[b,d,f], Guo-Jie Zhang[b,d,i], David Irwin[a,j,k], David M. Hillis[l,2], Robert W. Murphy[a,m], Huan-Ming Yang[d,n,o], Jing Che[a,2], Jun Wang[d,n,p,q,r,2], and Ya-Ping Zhang[a,h,2]

[a]State Key Laboratory of Genetic Resources and Evolution, and Yunnan Laboratory of Molecular Biology of Domestic Animals, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China; [b]China National GeneBank and [c]Shenzhen Key Laboratory of Transomics Biotechnologies, [d]BGI-Shenzhen, Shenzhen 518083, China; [e]College of Life Sciences, Sichuan University, Chengdu 610064, China; [f]School of Bioscience and Biotechnology, South China University of Technology, Guangzhou 510641, China; [g]Kunming College of Life Science, Chinese Academy of Sciences, Kunming 650204, China; [h]Laboratory for Conservation and Utilization of Bio-resource, Yunnan University, Kunming 650091, China; [i]Centre for Social Evolution, Department of Biology, University of Copenhagen, DK-2100 Copenhagen, Denmark; [j]Department of Laboratory Medicine and Pathobiology and [k]Banting and Best Diabetes Centre, University of Toronto, Toronto, ON, M5S 1A8, Canada; [l]Department of Integrative Biology and Center for Computational Biology and Bioinformatics, University of Texas at Austin, Austin, TX 78712; [m]Centre for Biodiversity and Conservation Biology, Royal Ontario Museum, Toronto, ON, M5S 2C6, Canada; [n]Princess Al Jawhara Albrahim Center of Excellence in the Research of Hereditary Disorders, King Abdulaziz University, Jeddah 21589, Saudi Arabia; [o]James D. Watson Institute of Genome Science, Hangzhou 310008, China; [p]Department of Biology, University of Copenhagen, 2200 Copenhagen, Denmark; [q]Macau University of Science and Technology, Taipa, Macau 999078, China; and [r]Department of Medicine, University of Hong Kong, Hong Kong

Contributed by David M. Hillis, February 3, 2015 (sent for review July 14, 2014; reviewed by Peter D. Vize)

The development of efficient sequencing techniques has resulted in large numbers of genomes being available for evolutionary studies. However, only one genome is available for all amphibians, that of *Xenopus tropicalis*, which is distantly related from the majority of frogs. More than 96% of frogs belong to the Neobatrachia, and no genome exists for this group. This dearth of amphibian genomes greatly restricts genomic studies of amphibians and, more generally, our understanding of tetrapod genome evolution. To fill this gap, we provide the de novo genome of a Tibetan Plateau frog, *Nanorana parkeri*, and compare it to that of *X. tropicalis* and other vertebrates. This genome encodes more than 20,000 protein-coding genes, a number similar to that of *Xenopus*. Although the genome size of *Nanorana* is considerably larger than that of *Xenopus* (2.3 vs. 1.5 Gb), most of the difference is due to the respective number of transposable elements in the two genomes. The two frogs exhibit considerable conserved whole-genome synteny despite having diverged approximately 266 Ma, indicating a slow rate of DNA structural evolution in anurans. Multigenome synteny blocks further show that amphibians have fewer interchromosomal rearrangements than mammals but have a comparable rate of intrachromosomal rearrangements. Our analysis also identifies 11 Mb of anuran-specific highly conserved elements that will be useful for comparative genomic analyses of frogs. The *Nanorana* genome offers an improved understanding of evolution of tetrapod genomes and also provides a genomic reference for other evolutionary studies.

de novo genome | transposable elements | chromosome rearrangement | highly conserved element

The age of genomics has ushered in opportunities to decode the history of evolution in ways unimaginable only a decade ago. More than 100 complete genomes have been sequenced and released for vertebrates. Amphibians, however, are poorly represented among these genomes. Despite the existence of more than 7,000 living species of amphibians, only the genome of *Xenopus tropicalis* (1) has been published. *Xenopus tropicalis*, however, falls outside of the Neobatrachia, which contains more than 96% of the known frog species (2). As a result, no neobatrachian genome is available for comparative analyses. Thus, this dearth of amphibian genomes greatly restricts comparative genomic studies of amphibians, and more generally, our understanding of a critical portion of tetrapod genome evolution at the major aquatic to terrestrial transition of vertebrates.

Nanorana (Dicroglossidae) includes more than 20 species of frogs native to Asia (research.amnh.org/vz/herpetology/amphibia). In this genus, three species, *Nanorana parkeri*, *Nanorana pleskei*, and *Nanorana ventripunctata*, are endemic to the Qinghai-Tibetan Plateau (3). In contrast to *Xenopus*, which is a secondarily derived aquatic obligate, species of *Nanorana* exhibit the terrestrial adult lifestyle that is typical of most anurans. *N. parkeri* occurs at elevations ranging from 2,850 to 5,000 m. Because this

Significance

We provide a de novo genome of the Tibetan frog, *Nanorana parkeri*, and conduct a series of comparisons with other vertebrates. Approximately one-half of the genome of *Nanorana* consists of transposable elements (TEs). The frequencies and distributional patterns of TEs differ considerably between *Nanorana* and *Xenopus*, the only other amphibian for which a genome has been sequenced. The genomes of these two frogs exhibit substantial homologous synteny blocks with rare interchromosomal and intrachromosomal rearrangements. We also identify 11 Mb of amphibian-specific conserved elements comprising 217 genes. These highly conserved genes provide a basis for comparative genomic analyses throughout frogs.

Author contributions: G.-J.Z., R.W.M., H.-M.Y., J.C., J.W., and Y.-P.Z. designed research; Y.-B.S., W.-W.Z., X.-L.T., L. Zhong, L.W., B.-L.Z., C.-L.Z., M.-M.Y., and H.-M.C. performed research; Y.-B.S., Z.-J.X., X.-Y.X., S.-P.L., D.-D.W., F.L., L. Zhou, S.-H.F., and C.H. analyzed data; and Y.-B.S., Z.-J.X., X.-Y.X., D.I., D.M.H., R.W.M., J.C., and Y.-P.Z. wrote the paper.

Reviewers included: P.D.V., University of Calgary.

The authors declare no conflict of interest.

Data deposition: The sequencing data have been deposited in the National Center for Biotechnology Information BioProject database, www.ncbi.nlm.nih.gov/bioproject (project ID PRJNA243398), and the raw sequencing reads in this paper have been deposited in National Center for Biotechnology Information Sequence Read Archive, www.ncbi.nlm.nih.gov/sra (accession no. SRA151427). The assemblies, gene sets, transposable elements annotation, and other supporting data are available from the GigaScience database, dx.doi.org/10.5524/100132.

[1]Y.-B.S., Z.-J.X., and X.-Y.X. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: dhillis@austin.utexas.edu, chej@mail.kiz.ac.cn, wangjun30@gmail.com, or zhangyp@mail.kiz.ac.cn.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1501764112/-/DCSupplemental.

species lives at such high elevations, it experiences extreme environmental conditions, including hypoxia, high levels of UV radiation, and dramatic changes in temperature on a daily basis. Consequently, *Nanorana* provides an additional excellent biological model to study the frog's adaptations to extreme conditions (cf. refs. 4 and 5), a topic of considerable interest in studies of endothermic birds and mammals (6–9).

In this study, we report the sequencing and annotation of the de novo genome of a female *N. parkeri*. We fill the gap of missing genomic data for neobatrachians and compare the *Nanorana* genome to that of *Xenopus* and some other key vertebrates. To advance our understanding of structural evolution of tetrapod genomes, our comparative analyses consider whole-genome synteny and chromosomal rearrangements, transposable elements and their distribution, amphibian-specific highly conserved elements (HCEs), and changes in functionally important multigene families.

Results and Discussion

The Tibetan frog genome was sequenced by using the Illumina sequencing platform. High-quality reads were first assembled into contigs, which were subsequently merged into scaffolds. Our 2.0-Gb de novo assembly (National Center for Biotechnology Information Bioproject accession no. PRJNA243398) had contig and scaffold N50s of 8.1 Kb and 1.05 Mb, respectively (*SI Appendix*, Tables S1–S3). The average sequence depth was estimated to be 83-fold, and ∼94% of the assembly was covered more than 20× (*SI Appendix*, Fig. S1). Average guanine-cytosine (GC) content (42.5%) of *N. parkeri* was similar to that of *X. tropicalis* and *Homo sapiens* (*SI Appendix*, Fig. S2), indicating GC-biased nonrandom sampling did not strongly affect the assembly. In addition, nearly 170 million *N. parkeri* transcriptome reads from brain, liver, ovary, and skin were also generated to complement genomic annotations. Multiple approaches for gene prediction identified 23,408 protein-coding genes in the draft assembly (*SI Appendix*, Table S4). The average coding sequence (CDS) length was 1,382 bp (*SI Appendix*, Table S4), and 96% of these genes were functionally annotated according to SwissProt and TrEMBL databases (*SI Appendix*, Table S5).

The haploid genome size of *N. parkeri* was estimated by k-mer coverage evaluation to be 2.3 Gb (*SI Appendix*, Table S2 and Fig. S3). Thus, our assembly covered about 87% (2.0/2.3) of the total genome. Although relatively small for an amphibian, the genome of *N. parkeri* (2.3 Gb) is still much larger than that of *X. tropicalis* (1.5 Gb). To address this difference, we compared their repeated sequences. We identified 970 Mb of transposable elements (TEs) in the genome of *N. parkeri*, which occupied approximately 48% of the total assembly. These sequences, which amounted to 318 Mb more than in *X. tropicalis* (*SI Appendix*, Table S6), accounted for most of the difference in size between the genomes of these two species. The two frogs also differed in their dominant forms of TEs. Long terminal repeats (LTRs) predominated in *N. parkeri* but transposons prevailed in *X. tropicalis* (*SI Appendix*, Table S7 and Fig. 1*A*). LTRs have been shown to be a major contributor to genomic gigantism (10). Thus, we compared the sequence divergence and insertion times of the LTRs in *N. parkeri* and *X. tropicalis* to determine whether a recent expansion of LTRs in *N. parkeri* drove its larger genome size. LTRs appeared to have been inserted into the genome of *N. parkeri* much earlier than those in *X. tropicalis* (∼50 vs. ∼30 Ma; *SI Appendix*, Fig. S4); however, the LTRs of *N. parkeri* were less divergent than those of *X. tropicalis* (Fig. 1*B*).

N. parkeri has higher levels of divergence in most major TE families (e.g., LINE and DNA transposons) than in *X. tropicalis* (Fig. 1*B*), except for LTRs. This pattern is especially true for the DNA transposons, which constitute the most abundant TE families in *X. tropicalis* (1). For example, the transposon *Kolobok*, which was found for the first time in *X. tropicalis* (1), also occurs in *N. parkeri,* but in much smaller copy numbers (0.1 Mb, vs. 96.2 Mb in *X. tropicalis*; *SI Appendix*, Table S7). Sequences of *Kolobok* show higher divergence in *N. parkeri* (Fig. 1*B*) despite the differences in abundance. This result indicates the expansion of the TE family in *X. tropicalis* occurred after the two frogs split. Furthermore, the two frogs also differ in the distributional patterns of TE families along the total genome. As Fig. 1*C* shows, many TE families that cluster near one another in *N. parkeri* appear scattered across different clusters in *X. tropicalis*. The distribution of TEs in amphibians shows much greater differences than that observed between human and chicken genomes.

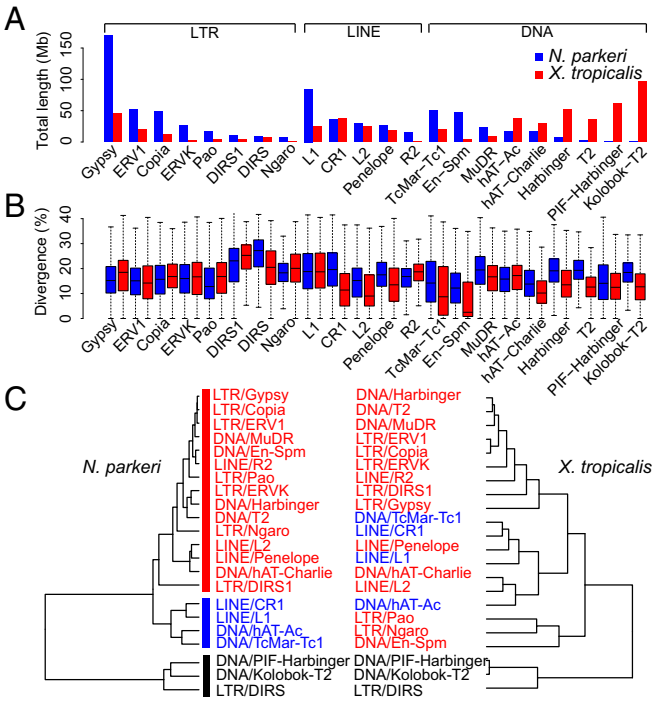

Fig. 1. Comparison of transposable elements (TEs) between *N. parkeri* and *X. tropicalis*. (*A*) Comparisons of the TEs in the two frogs. (*B*) Boxplot of divergence rates for different TE families for the two frogs. Divergence rate calculated from alignments of the TEs identified against entries in the Repbase library. Generally, the divergence of TEs in *Nanorana* is higher than in *Xenopus*. (*C*) Clustering results based on the distribution correlation within a 2-Mb sliding windows across genome; two TE families are more correlated (with same color) in distribution if they show more similar occurrences in amount. DNA, DNA transposon; LINE, long interspersed nuclear elements; LTR, long terminal repeat transposable elements.

Comparison of the Tibetan frog genome with those of *X. tropicalis*, *Anolis carolinensis*, *Gallus gallus*, *H. sapiens*, and *Danio rerio* from the ENSEMBL database yields insights into the divergence of vertebrates. The six species share 9,413 gene families (Fig. 2*A* and *SI Appendix*, Fig. S5), and we construct a maximum likelihood phylogeny from 4,279 single-copy orthologs (Fig. 2*A*). This phylogeny suggests the two frogs diverged at approximately 266 (134.2–311.2) Ma (Fig. 2*A*), approximately 40 million years earlier than previously proposed by TimeTree project (11).

Analysis of gene families among the six vertebrates identifies 832 and 161 gene families unique to *N. parkeri* and *X. tropicalis*, respectively (*SI Appendix*, Fig. S5). Dynamic evolutionary analysis further identifies 328 significantly ($P < 0.05$) expanded multigene families in *N. parkeri* compared with the other species (Fig. 2*A*). In comparison, *X. tropicalis* has 99 expanded families (Fig. 2*A*). The expanded gene families of *N. parkeri* and *X. tropicalis* differ in their enriched functional classifications. The former mainly functions in signaling receptor activities, such as olfactory receptor

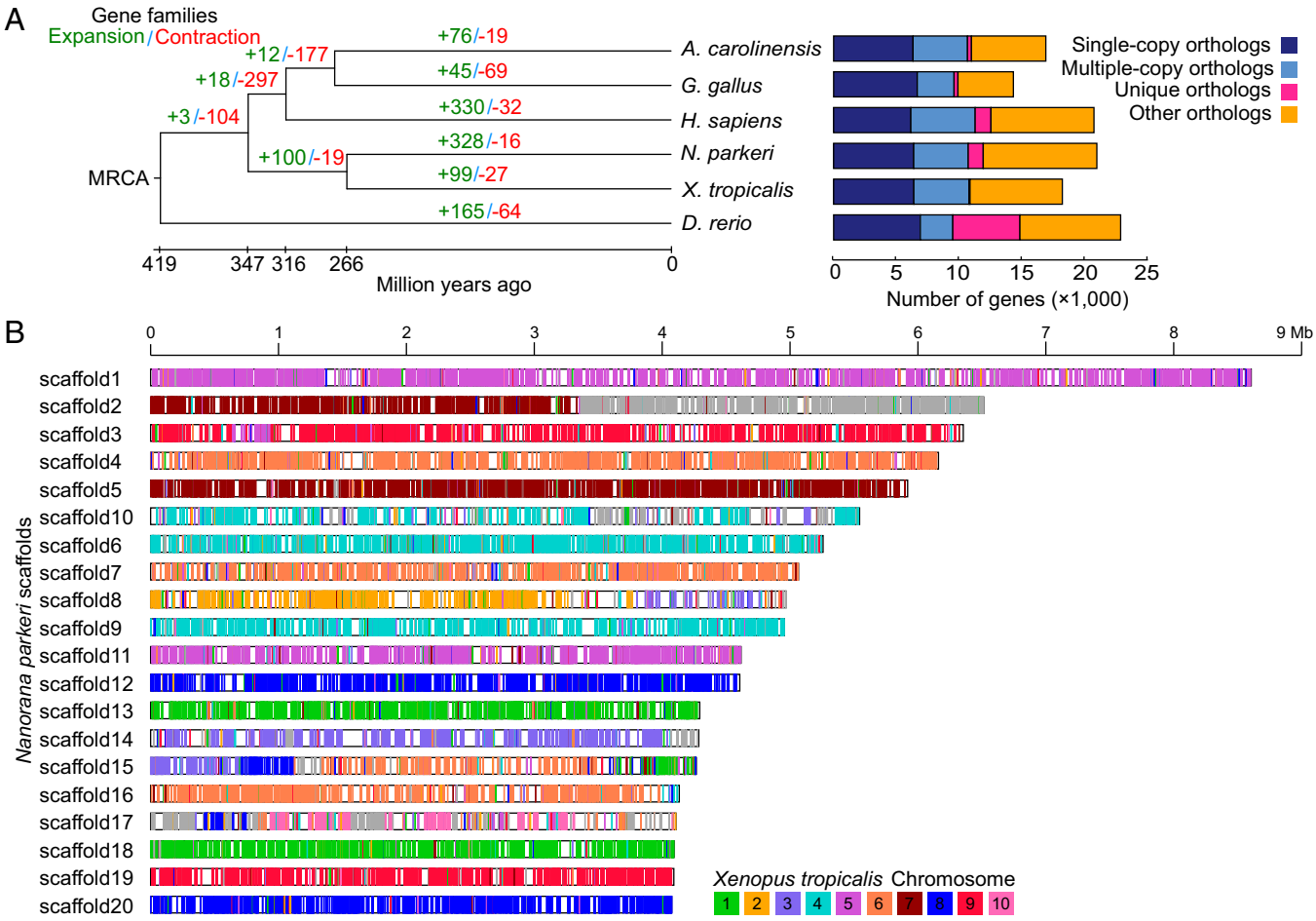

Fig. 2. Expansion and contraction in gene families (*A*) and whole-genome synteny between *N. parkeri* and *X. tropicalis* (*B*). (*A*) The phylogeny using one representative genome for each major vertebrate lineage and the dynamic evolution of gene families along each lineage (MRCA, most recent common ancestor). Divergence time between the two frog species was estimated to be 266 Ma. (*B*) Syntenic map between the large (>4 M) scaffolds of the genome of *N. parkeri* and the chromosomal map of *X. tropicalis*. Colors refer to the different chromosomes of *Xenopus* that hold scaffolds of *N. parkeri*. Gray bars indicate aligned scaffolds of *X. tropicalis* not incorporated into chromosomes of *Xenopus*.

activity (*SI Appendix*, Table S8.1), whereas the later primarily functions in endopeptidase activity (*SI Appendix*, Table S8.2). The difference likely reflects local adaptations imposed by evolutionary constraints that differ between the two frogs. Further comparisons of gene content reveal that 44 genes (*SI Appendix*, Table S9.1) in *Xenopus* are not present in the genome of *N. parkeri*. These genes mainly function as regulation of smooth muscle contraction, cytokine activity, and cytokine receptor binding (*SI Appendix*, Table S9.2). Our analysis also identified 148 pseudogenes found in *N. parkeri* that are still functional genes in *X. tropicalis* (*SI Appendix*, Table S10). Possible reasons for these missing genes in *N. parkeri* involve different natural selective constraints acting on the two frogs after their split. Note, however, that the difference of genome completeness between *N. parkeri* (~87%) and *X. tropicalis* (~95%) may also account for some of the missing genes.

Mapping of scaffolds from *N. parkeri* to the chromosomes of *X. tropicalis* (12) identifies patterns of structural evolution in the amphibian genomes. The genome-scale alignments reveal that although the two frogs diverged 266 Ma, amphibian chromosomes retain a large amount of conserved synteny. Only a few interchromosome rearrangements exist (Fig. 2*B*). Thus, amphibian genomes have a slow rate of structural evolution. Nonetheless, numerous small gaps exist in one large block (Fig. 2*B*), denoting the occurrence of segmental rearrangements. A self-versus-self comparison of the scaffolds of *N. parkeri* identifies ~41 Mb of segmental duplication (SD) regions with 1,268 genes locating within these regions. Unexpectedly, although *X. tropicalis* has a much smaller genome size, its genome contains a larger component of SD regions (125 Mb). Within each range of copy numbers (2~5, 6~10, 11~20, 21~50, and 51~100), *X. tropicalis* shows a much larger number of SD clusters (*SI Appendix*, Table S11). Furthermore, 328 kb of SD regions in *X. tropicalis* hold a syntenic relationship with 170 kb in *N. parkeri*, indicating these duplications occurred in their common ancestor. However, the majority of SD regions evolved after the two species diverged.

Multiple genome alignments, including human, chicken, and lizard, allow the identification and comparison of other rearrangement events such as indels, translocations, and reversed blocks. The amphibians (*Nanorana* and *Xenopus*), reptile, bird, and mammal genomes have 237, 290, 326, and 528 genomic rearrangement blocks, respectively. Frogs appear to have 0.043 rearrangements per 100 Mb per million years. This rate is comparable to that of ectothermic reptiles (0.039), but much smaller than endothermic birds (0.128) or mammals (0.101) (*SI Appendix*, Table S12). This analysis represents the first whole-genome comparison to our knowledge of rearrangements within amphibians, and it provides evidence that ectothermic vertebrates may have a slower evolutionary rate of genomic rearrangements than do endothermic vertebrates. This slower evolutionary rate

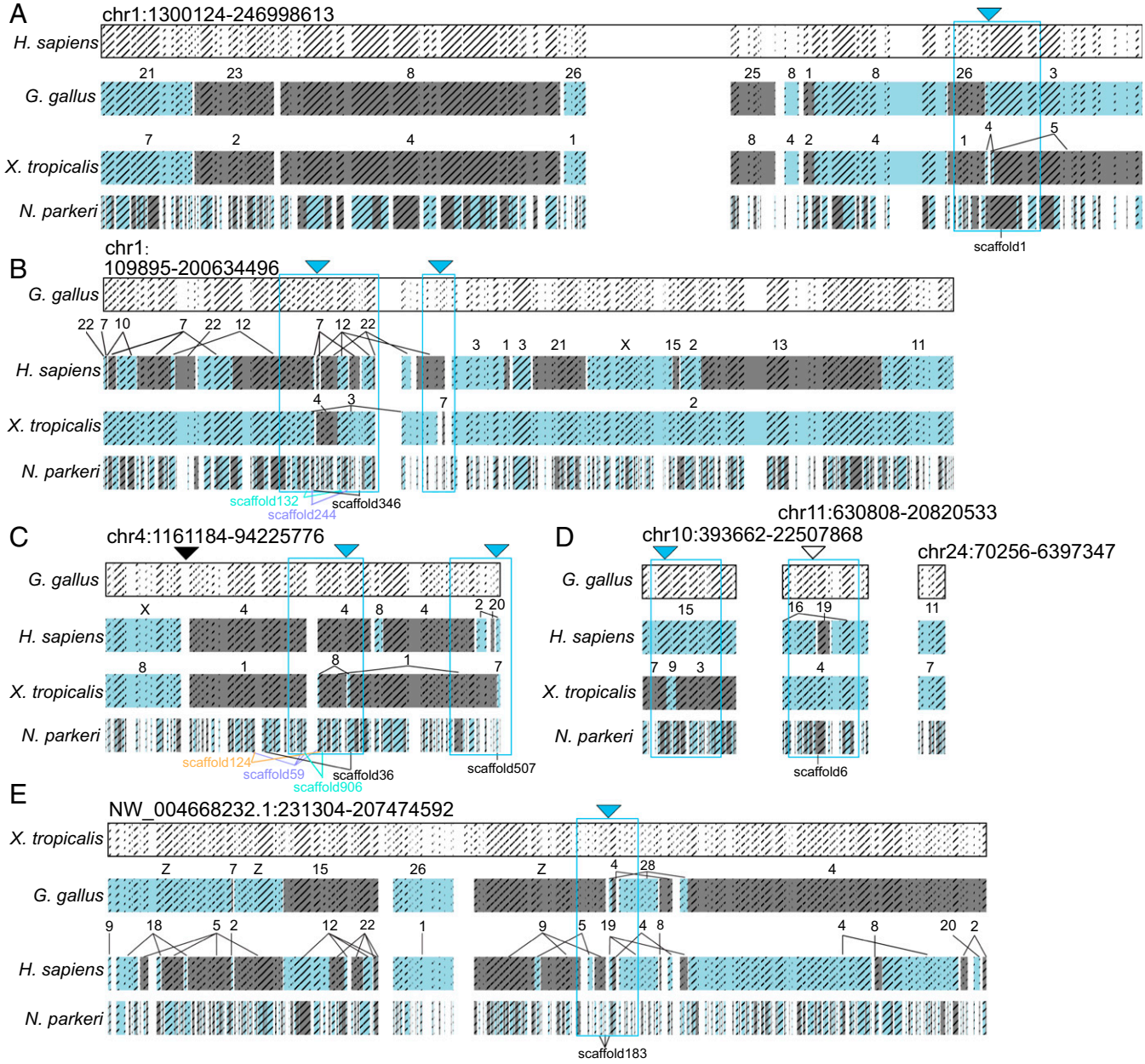

Fig. 3. Chromosome synteny blocks among human, chicken, *N. parkeri*, and *X. tropicalis*. (*A*) Homologous synteny blocks between human chr1 and other genomes. The homologous synteny blocks (HSBs; shaded areas) occur over the whole human chr1; gray and blue shadings are used to indicate different chromosomes/scaffolds. A light-blue triangle points to a *Xenopus*-specific break. (*B*) Homologous synteny blocks between chicken chr1 and other genomes. HSBs (shaded areas) occur across chicken chr1. Light blue triangles point to two breaks in *Xenopus*, one of which also occurs in *Nanorana*, indicating a common break in amphibians. (*C*) Homologous synteny blocks between chicken chr4 and other genomes. Light blue triangles point to two *Xenopus*-specific breaks, and the black triangle points to a chicken-specific fusion. (*D*) Homologous synteny blocks between chicken microchromosomes and other genomes. Light blue and white triangles point to amphibian- and human-specific breaks, respectively. (*E*) Homologous synteny blocks between the longest scaffold of *Xenopus* and the other genomes. The light blue triangle indicates an amniote fission.

also appears to extend to the evolutionary rate of protein-coding genes in ectothermic vertebrates (compared with endothermic birds and mammals). Among the 9,964 single-copy orthologs shared among *N. parkeri*, *X. tropicalis*, and human, frogs appear to have a substitution rate of 0.776×10^{-9} substitutions per site per year ($\pm 1.34 \times 10^{-12}$ SE), which falls within the range of a previous estimate based only on a small part of nuDNAs (13). The coelacanth also showed a much slower evolutionary rate (14). Rates for these ectotherms are much lower than those of endothermic birds (15) ($\sim 2 \times 10^{-9}$) and mammals (16) ($\sim 2.2 \times 10^{-9}$). Further analysis on the whole-genome alignments (without indels) produces a mean divergence estimate of 39.8% (number of mismatches/total length of genome alignment) between the two frog genomes, in contrast with 3.7% for the primate genomes (human versus chimpanzee). Given previously estimated divergence times of these species (~266 and ~6 MYs for frog and primate, respectively; ref. 17), frogs still appear to have a much slower substitution rate (0.749×10^{-9} versus 3.12×10^{-9} substitutions per site per year). Thus, among tetrapods, the genomes of ectotherms appear to evolve more slowly than do those of endotherms. This lower evolutionary rate may correspond to the lower metabolic rate of ectotherms (18).

Within the multigenome alignments, human, chicken, and the two frogs have 114 ancestral homologous synteny blocks (aHSBs). These aHSBs cover 71% (overlap with 1,482 scaffolds) and 89% (overlap with all 10 chromosomes) of *N. parkeri* and *X. tropicalis* genomes, respectively (*SI Appendix*, Table S13). Fig. 3*A* shows the tetrapod synteny of human chromosome 1 (chr1). Several intact regions along human chr1 have detailed synteny with other animals, such as the chicken chr21 (collinear with *Xenopus* chr7), chr23 (*Xenopus* chr2), chr8 (*Xenopus* chr4), chr26 (*Xenopus* chr1), chr25 (*Xenopus* chr8), and chr3 (*Xenopus* chr5). The synteny is remarkable given the ∼360 MYs since their divergence (Fig. 3*A*). Similarly, several intact regions in the chicken chr1 are collinear with chromosomes of some other animals (Fig. 3*B*).

Tetrapod ancestral synteny of human chr1 shows one *Xenopus*-specific breakpoint (Fig. 3*A*), and the synteny of chicken chr1 also exhibits a common breakpoint within amphibians (Fig. 3*B*). Thus, the rearrangement predates the divergence of the two frogs. To further identify lineage-specific fusion and breakage events within these animals, we extended the analysis to all human and chicken chromosomes (1). Our analyses identify six chromosomal/scaffold fissions in *X. tropicalis*, one of which also occurs in *N. parkeri* (Fig. 3*B*). In addition, we confirmed three *X. tropicalis*-specific rearrangements in comparison with *N. parkeri* (Fig. 3 *A* and *C*). Generally, the amphibians and chicken have comparable rates of intrachromosome rearrangements (*SI Appendix*, Table S14), yet they exhibit relatively fewer interchromosomal rearrangements than human (Fig. 3 *A* and *B*). For example, human has 35 fusion and 29 fission events, versus 7 fusions and 3 fissions in chicken. The results again show that mammals have undergone considerably more rearrangements than other tetrapods (1). In agreement with the previous estimates (1), we confirm more human- and chicken-specific chromosomal rearrangements. Analyses also show that most chicken microchromosomes do not have interchromosomal rearrangements (Fig. 3*D*). This finding is consistent with the previous hypothesis (19) that many chicken microchromosomes might correspond to ancestral tetrapod chromosomes. Furthermore, the multigenome alignments also identify an amniote fission event (Fig. 3*E*).

Highly conserved DNA sequences among distantly related species are always under purifying selection on essential functions (20). By estimating PhastCons scores (the genome conservation index) (20) among these animals, the genome of *N. parkeri* also shows sequence conservation along with other vertebrates. Approximately 12 Mb of highly conserved elements (HCEs) occur among human, chicken, *X. tropicalis,* and *N. parkeri*. The majority of these HCEs are located within the protein-coding regions, which corresponds to 43.08% of the HCEs. Furthermore, 3.15% of HCEs are located in 3,088 long noncoding RNAs (lncRNAs). The remaining HCEs occur in intergenic regions (26.68%), introns (20.40%), and untranslated regions (UTRs, 6.67%) (*SI Appendix*, Fig. S6). By mapping the HCEs to the human genome, we identify 2,466 genes in these regions. These genes show enrichment in many fundamental biological functions. The top three terms from the Gene Ontology (GO) database (21) associate with these genes are hydrolase activity, ATPase activity, and nucleosome (*SI Appendix*, Table S15).

All organisms experience specific evolutionary conditions that result in diverse adaptations. Lineage-specific HCEs should indicate some of the genetic bases of lineage-specific adaptations. Amphibians have a number of HCEs that are not found in other tetrapods. To detect these potentially amphibian-specific HCEs, we remove regions with high similarities to human and chicken genomes from the 22 Mb of amphibian HCEs, which results in a list of 217 genes (*SI Appendix*, Table S16.1). Functional classifications of these potentially frog-specific highly conserved genes show that RNA processes (metabolic, catabolic, and regulation of translation) constitute the majority of enriched GO terms (*SI Appendix*, Table S16.2). In addition, some of these conserved genes associate with tongue development (GO:0043586 with genes *HAND2*, *PRDM16*, and *WNT10A*) (*SI Appendix*, Table S16). The tongue of most frogs can be flipped out quickly to catch insects and other prey, and their tongue is morphologically specialized to accomplish this function. *HAND2* plays important roles in tongue morphogenesis by regulating expression levels of other genes, including *Dlx5* and *Dlx6* (22). *PRDM16* is also necessary for normal palatogenesis (23). We expect that these highly conserved genes play important roles in the evolution of amphibians, but their role probably extends far beyond tongue development. These genes are still conserved in *X. tropicalis*, although this species does not have a tongue.

Antimicrobial peptides (AMPs) have been identified in various species ranging from bacteria to mammals (24). They form the first line of host–defense against pathogenic infections and are a key component of the ancient innate immune system. By mapping the 2,195 published AMPs (24) to the frog genomes, we identify more than 40 kinds of AMPs in both frogs. Frogs share 14 of these with the genomes of human, chicken, and lizards. Three kinds of AMPs (AP01357, AP00097, and AP01583) occur only in *N. parkeri*. AP01583 in *N. parkeri* serves multiple functions in defending against environmental oxidative stress and pathogenic microorganisms (25). All of the *N. parkeri*-specific AMPs may partially associate with ecological adaptations to its extreme environment.

We can also infer some aspects of the population history of *N. parkeri* from its genome. The genome contains 765,172 heterozygous single-nucleotide polymorphisms (SNPs). Thus, 0.43% of its total nucleotide sites exhibit heterozygosity. This value is much lower than that seen in other vertebrates (i.e., human: 0.69%; naked mole rat: 0.70%; panda: 1.32%; chicken: 4.5%). The difference could result from high levels of inbreeding within the Tibetan frog. Using pairwise sequential Markovian coalescent analysis (26), we can detect a major increase in the effective population size (N_e) of *N. parkeri*, which dates to just after the last glacial maximum (*SI Appendix*, Fig. S7). This result suggests that the species had a small population size at times of maximum glaciation.

Genomes are now available for two species of amphibians that diverged approximately 266 Ma. The genome of *N. parkeri* is the first complete genome to our knowledge from the species-rich and widespread Neobatrachia, which contains the vast majority of amphibian taxa. This genome greatly broadens our understanding of the evolution of tetrapod genomes by providing additional comparative resources for the water-to-land transitional vertebrates. *N. parkeri* also provides materials for deciphering amphibian high-elevation adaptations, and genomic changes that accompanied the vertebrate transition from water to land.

Methods

SI Appendix, SI Methods has additional information relating to the methodologies described below.

Genome Sequencing and Assembly. We isolated genomic DNA from muscle tissue of an adult female *N. parkeri*. We collected this frog from the Qinghai-Tibetan Plateau at an elevation of 4,900 m. We constructed paired-end DNA libraries with different insert-size lengths (170 bp to 20 kb) and sequenced these libraries on the Illumina HiSeq 2000 sequencer. After performing a series of strict filtering steps to remove artificial duplication, adapter contamination, and low-quality reads, we obtained 190 Gbp of high-quality data (83× genome coverage) for assembly. We assembled the genome by using SOAPdenovo (27) and SSPACE (28). To support subsequent annotation, we collected one additional individual of the same species from the same region of the Tibetan Plateau and extracted RNA from fresh liver, brain, skin, and ovary tissues for the generation of transcriptome data.

Gene Models. We used all available transcriptome data from multiple tissues to develop gene-model predictions. The assemblies, gene sets, transposable elements annotation, and other supporting data are available from the GigaScience database (dx.doi.org/10.5524/100132). Gene functions were

assigned according to the best match to the SwissProt and TrEMBL databases. Based on the gene predictions, ortholog sets among the major vertebrate lineages (human, chicken, lizard, *Xenopus*, and zebrafish) were assembled with the TreeFam method (29). Analyses included estimation of gene family expansion and contraction and identification of lineage-specific genes, during which gene family expansion analysis was performed by CAFE (30).

Transposable Element Identification. We used repeat elements libraries from Repbase database (version 16.10) to predict tandem repeats and the LTRs in the Tibetan frog genome. We also constructed a de novo repeat library to identify the frog's repeat elements. To compare the distribution pattern of different TE families between *N. parkeri* and *X. tropicalis*, we used a non-overlapping sliding window analysis (window = 2 Mb) to count the number of TEs and finally calculated the correlation coefficients for pair-wise TEs. For the segmental duplication identification, we performed a self-versus-self analysis on the repeat-masked genome by using Lastz (31) with parameters set to T = 2, C = 2, H = 2,000, Y = 3,400, L = 6,000, and K = 2,200.

Genome Rearrangement and Conservation. To compare the rate of genomic rearrangement in *N. parkeri* with other vertebrates, we first generated pairwise whole-genome alignments to determine synteny blocks. We then used a dynamic programming script to estimate the total numbers of indels, translocations, and reversals of the blocks. We constructed ancestral homologous synteny blocks (aHSBs) for the common ancestor of human, chicken, *N. parkeri*, and *X. tropicalis* and identified lineage-specific chromosome fusion and/or fission events. PhastCons (20) was used to identify conserved elements in these vertebrate genomes.

ACKNOWLEDGMENTS. We thank David Wake, Peng Shi, Guo-Dong Wang, and Min-Sheng Peng for helpful discussions and comments. Jun-Xiao Yang, Jie-Qiong Jin, Fang Yan, Ke Jiang, Kai Wang, Yun Gao, and Shi-Fang Wu assisted with sample collection and technical help in the laboratory, and Amy Lathrop prepared part of the figures. This work was supported by Strategic Priority Research Program (B) Grant XDB13020200 of the Chinese Academy of Sciences (CAS), National Natural Science Foundation of China Grant 91431105, Key Research Program of CAS Grant KJZD-EW-L07, the Animal Branch of the Germplasm Bank of Wild Species of CAS (the Large Research Infrastructure Funding), and Shenzhen Municipal Government of China Grant CXB201108250096A.

1. Hellsten U, et al. (2010) The genome of the Western clawed frog Xenopus tropicalis. *Science* 328(5978):633–636.
2. University of California, Berkeley (2014) AmphibiaWeb: Information on amphibian biology and conservation. Available at amphibiaweb.org/. Accessed December 14, 2014.
3. Che J, et al. (2010) Spiny frogs (Paini) illuminate the history of the Himalayan region and Southeast Asia. *Proc Natl Acad Sci USA* 107(31):13765–13770.
4. Ludwig G, Sinsch U, Pelster B (2013) Migratory behaviour during autumn and hibernation site selection in common frogs (Rana temporaria) at high altitude. *Herpetol J* 23(3):121–124.
5. Chen W, Wang X, Fan X (2013) Do anurans living in higher altitudes have higher prehibernation energy storage? Investigations from a high-altitude frog. *Herpetol J* 23(1):45–49.
6. Qu Y, et al. (2013) Ground tit genome reveals avian adaptation to living at high altitudes in the Tibetan plateau. *Nat Commun* 4:2071.
7. Ge RL, et al. (2013) Draft genome sequence of the Tibetan antelope. *Nat Commun* 4: 1858.
8. Cai Q, et al. (2013) Genome sequence of ground tit Pseudopodoces humilis and its adaptation to high altitude. *Genome Biol* 14(3):R29.
9. Qiu Q, et al. (2012) The yak genome and adaptation to life at high altitude. *Nat Genet* 44(8):946–949.
10. Sun C, et al. (2012) LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. *Genome Biol Evol* 4(2):168–183.
11. Hedges SB, Dudley J, Kumar S (2006) TimeTree: A public knowledge-base of divergence times among organisms. *Bioinformatics* 22(23):2971–2972.
12. Wells DE, et al. (2011) A genetic map of Xenopus tropicalis. *Dev Biol* 354(1):1–8.
13. Crawford AJ (2003) Relative rates of nucleotide substitution in frogs. *J Mol Evol* 57(6): 636–641.
14. Amemiya CT, et al. (2013) The African coelacanth genome provides insights into tetrapod evolution. *Nature* 496(7445):311–316.
15. Nam K, et al. (2010) Molecular evolution of genes in avian genomes. *Genome Biol* 11(6):R68.
16. Kumar S, Subramanian S (2002) Mutation rates in mammalian genomes. *Proc Natl Acad Sci USA* 99(2):803–808.
17. Kumar S, Filipski A, Swarna V, Walker A, Hedges SB (2005) Placing confidence limits on the molecular age of the human-chimpanzee divergence. *Proc Natl Acad Sci USA* 102(52):18842–18847.
18. Martin AP, Palumbi SR (1993) Body size, metabolic rate, generation time, and the molecular clock. *Proc Natl Acad Sci USA* 90(9):4087–4091.
19. Burt DW (2002) Origin and evolution of avian microchromosomes. *Cytogenet Genome Res* 96(1-4):97–112.
20. Siepel A, et al. (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 15(8):1034–1050.
21. Ashburner M, et al.; The Gene Ontology Consortium (2000) Gene ontology: Tool for the unification of biology. *Nat Genet* 25(1):25–29.
22. Barron F, et al. (2011) Downregulation of Dlx5 and Dlx6 expression by Hand2 is essential for initiation of tongue morphogenesis. *Development* 138(11):2249–2259.
23. Bjork BC, Turbe-Doan A, Prysak M, Herron BJ, Beier DR (2010) Prdm16 is required for normal palatogenesis in mice. *Hum Mol Genet* 19(5):774–789.
24. Wang Z, Wang G (2004) APD: The Antimicrobial Peptide Database. *Nucleic Acids Res* 32(Database issue):D590–D592.
25. Lu Z, et al. (2010) Novel families of antimicrobial peptides with multiple functions from skin of Xizang plateau frog, Nanorana parkeri. *Biochimie* 92(5):475–481.
26. Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* 475(7357):493–496.
27. Li R, Li Y, Kristiansen K, Wang J (2008) SOAP: Short oligonucleotide alignment program. *Bioinformatics* 24(5):713–714.
28. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W (2011) Scaffolding preassembled contigs using SSPACE. *Bioinformatics* 27(4):578–579.
29. Li H, et al. (2006) TreeFam: A curated database of phylogenetic trees of animal gene families. *Nucleic Acids Res* 34(Database issue):D572–D580.
30. De Bie T, Cristianini N, Demuth JP, Hahn MW (2006) CAFE: A computational tool for the study of gene family evolution. *Bioinformatics* 22(10):1269–1271.
31. Brudno M, et al. (2004) Automated whole-genome multiple alignment of rat, mouse, and human. *Genome Res* 14(4):685–692.