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.

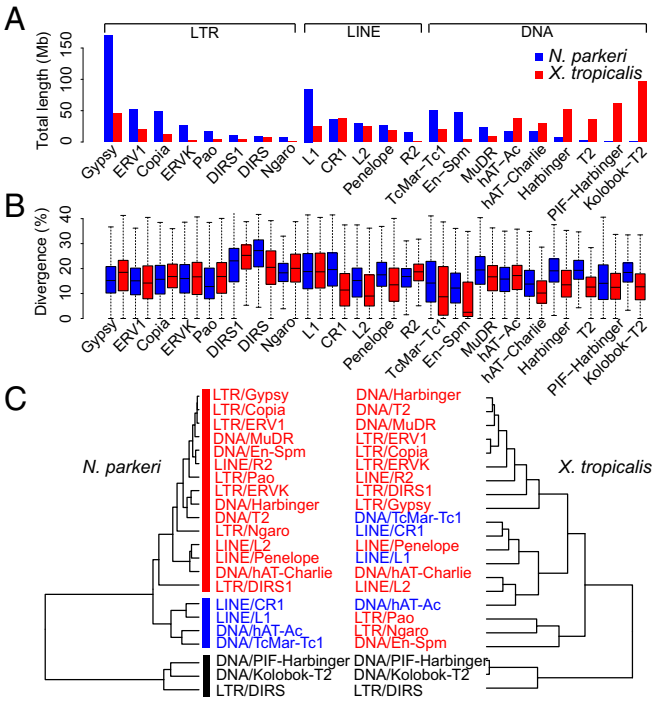

Fig. 1. Comparison of transposable elements (TEs) between *N. parkeri* and *X. tropicalis*. (*A*) Comparisons of the TEs in the two frogs. (*B*) Boxplot of divergence rates for different TE families for the two frogs. Divergence rate calculated from alignments of the TEs identified against entries in 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

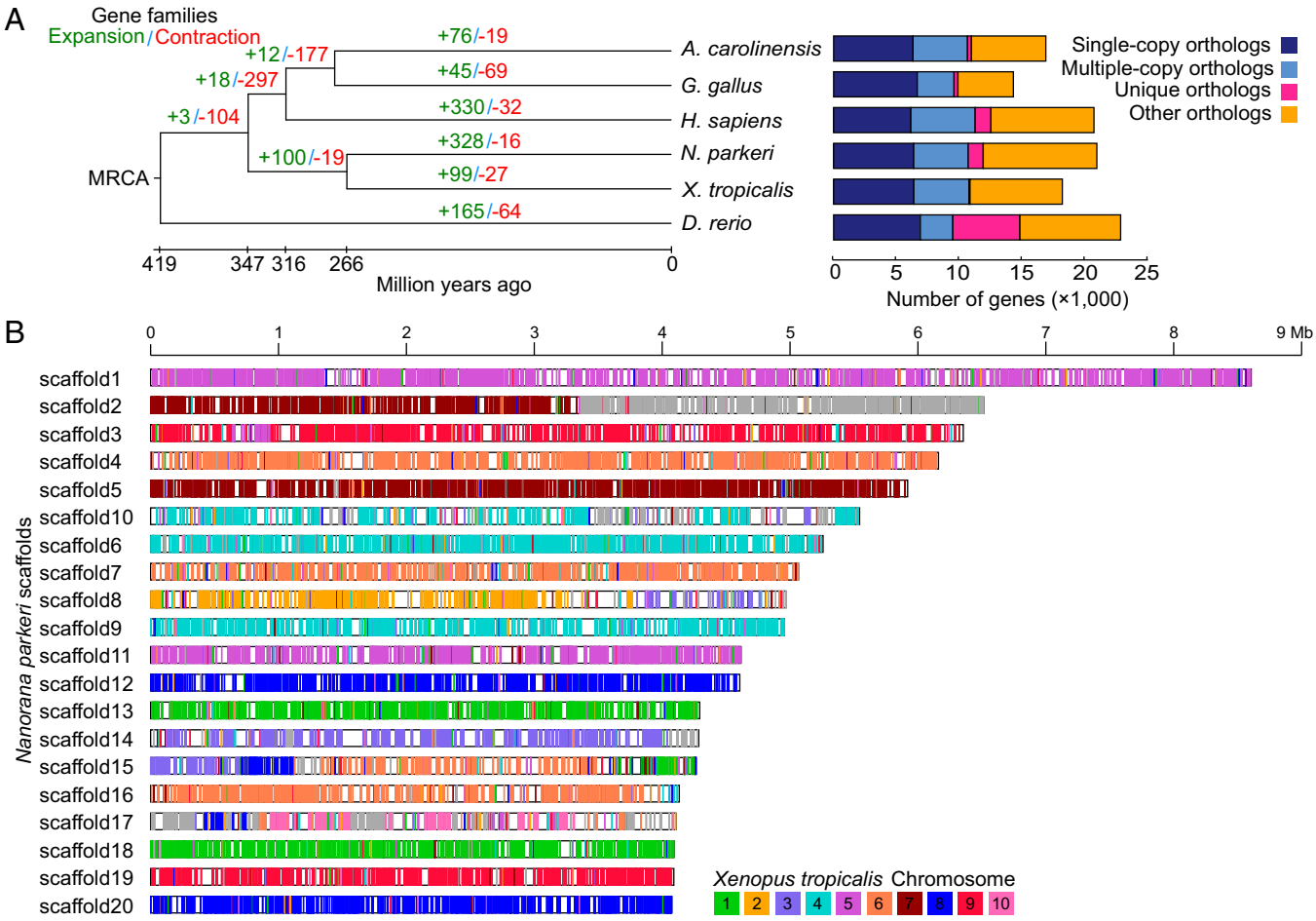

Fig. 2. Expansion and contraction in gene families (*A*) and whole-genome synteny between *N. parkeri* and *X. tropicalis* (*B*). (*A*) The phylogeny using one representative genome for each major vertebrate lineage and the dynamic evolution of gene families along each lineage (MRCA, most recent common 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

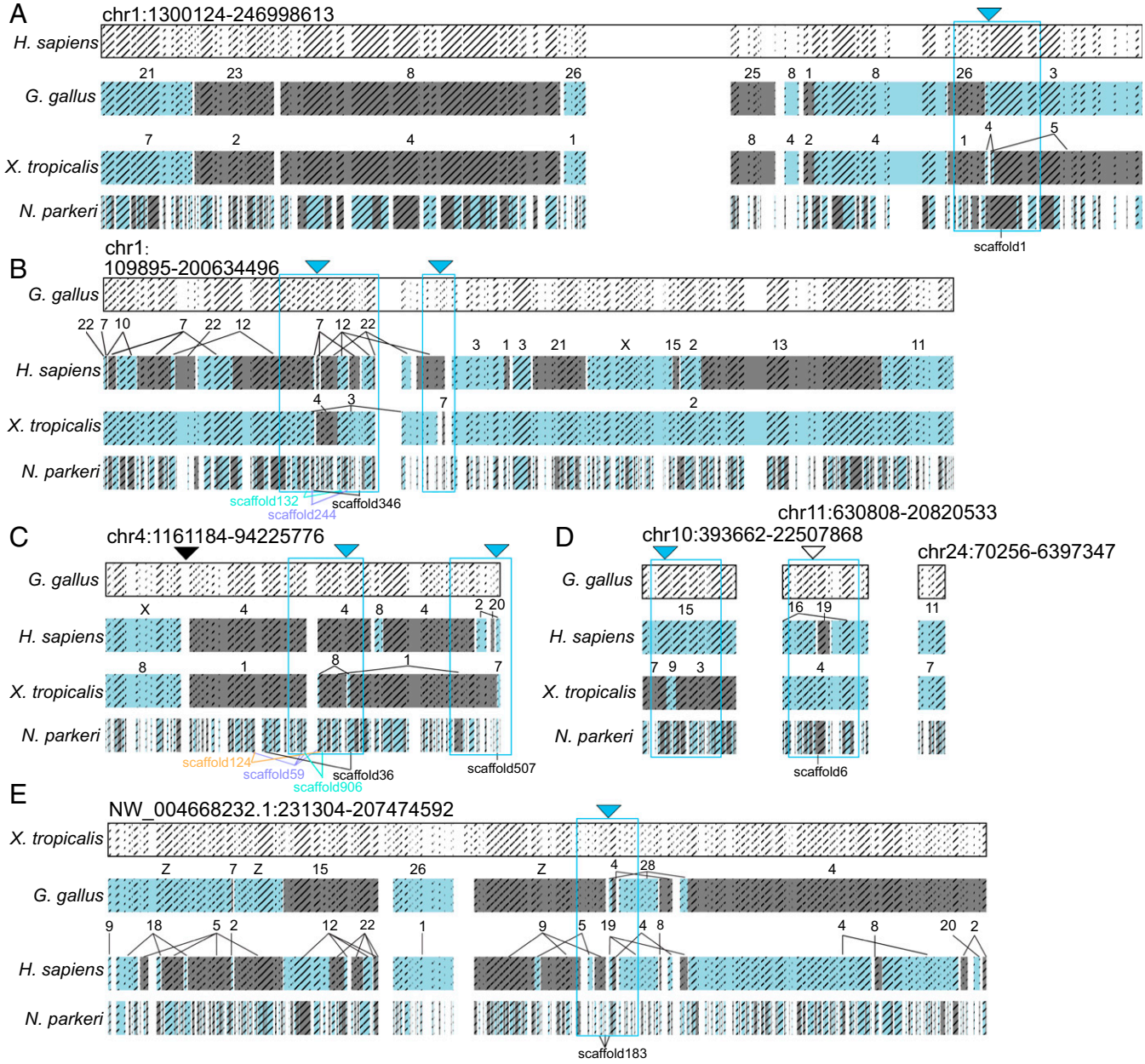

Fig. 3. Chromosome synteny blocks among human, chicken, *N. parkeri*, and *X. tropicalis*. (*A*) Homologous synteny blocks between human chr1 and other genomes. The homologous synteny blocks (HSBs; shaded areas) occur over the whole human chr1; gray and blue shadings are used to indicate different chromosomes/scaffolds. 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.

Supporting Information

Materials and Methods

Genome/Transcriptome Sequencing and Assembly

A female adult *Nanorana parkeri* was collected on the Qinghai-Tibetan Plateau at an elevation of about 4900m. Genomic DNA was extracted from muscle tissue. Paired-end DNA libraries with different insert size lengths (170 bp to 20 kb) were constructed, with the short insert size libraries yielding read-lengths of 100 and 150bp and long insert size libraries 49bp ends. All the sequences were generated via the Illumina Hiseq 2000 platform. In total, 323 Gb of raw reads were obtained from these libraries. For subsequent assembly, we performed a series of strict filtering steps to remove artificial duplications, adapter contamination, and low-quality reads. Briefly, reads from short-insert libraries (<2,000 bp) were first assembled into contigs on the basis of k-mer overlap information. Then, reads from long-insert libraries (≥2,000 bp) were aligned onto contigs to construct scaffolds. Finally, we used the paired-end information to retrieve read-pairs and then performed a local assembly of the collected reads to fill gaps between the scaffolds. Genome assembly quality was evaluated using GC content.

For transcriptome analysis, Poly (A) mRNAs were first isolated using oligonucleotide (dT) magnetic beads and disrupted into short segments. This was followed by cDNAs synthesis using random hexamer primers and reverse transcriptase. After end-repair, adapter-ligation and PCR amplification, each paired-end cDNA library was sequenced with a read length of 101 bp using the Illumina Hiseq 2000 sequencing platform. Reads with potential adaptor sequences and low-quality regions were trimmed by applying cutadapt (1) and Btrim (2), respectively.

Heterozygous SNP Detection and Estimation of Population History

To evaluate heterozygosity and its distribution, high-quality reads from short-insert (<2,000 bp) libraries were first realigned to the assembly with BWA (3). SOAPsnp (4) was then used to identify heterozygous SNPs. The probability at each possible site on the reference genome was calculated based on a Bayesian statistical model. The genotype with the highest probability at each position was inferred. We used a set of filters to obtain candidate SNPs, including 1) Phred score ≥20, 2) overall sequence depth ≤90, 3) at least 3 unique read mapped for each allele, 4) a minimum of 5 bp between SNPs, and 5) the approximate copy number of flanking sequences was < 2.

We further adopted the pairwise sequential Markovian coalescent (PSMC) model to investigate in detail the time of the most recent common ancestor of *N. parkeri*. Illumina short reads were re-aligned to the *N. parkeri* genome using BWA. Consensus sequences were called

using samtools (5) and then converted into the fastq format using bcftools and vcfutils. The bases were further filtered with Phred scores being not less than 20. Sequences were split into short segments of 500kb to perform bootstrapping of the PMSC estimates. Split sequences were then used to reconstruct the demographic history with the PSMC model with the following parameters: -N25 -t15 -r5 -b -p "4+25*2+4+6". One hundred bootstrap replicates assessed variance of the simulated results. Finally, we scaled the PSMC profiles by using generation time (g) = 3 yr, and neutral mutation rate per yr (μ) = 0.776e−09 (estimated herein).

Transposable Elements and Segmental Duplication

Transposable elements (TEs) were identified in the genome by a combination of homology-based and *de novo* approaches. The homology method identified known TEs using RepeatMasker and RepeatProteinMask (http://www.repeatmasker.org/) against the Repbase TE library (Repbase 16.10) at the DNA and protein level, respectively. A *de novo* repeat library was constructed using RepeatModeler and Piler (6). We used TRF (7) to predict tandem repeats. LTR_FINDER (8) was used to search the whole genome for the characteristic structure of full-length long terminal repeat retrotransposons (LTR). For comparison, we performed the same analyses on the genome of *X. tropicalis* to identify its repeat sequences. Genomic sequences of *X. tropicalis* were downloaded from the Ensembl Database (9).

We further used LTR_STRUC (10) to identify full-length LTRs for both *N. parkeri* and *X. tropicalis*. Both 5' and 3' LTR regions for each full-length LTR were extracted. All LTR pairs were aligned using MUSCLE (11). Insertion time was calculated based on the sequence similarity between two LTRs using the formula, $T = K/(2)(r)$, where K was the nucleotide distance between the LTRs calculated by the distmat program implemented in the EMBOSS package (12) and r was the rate of nucleotide substitution.

To identify segmental duplication in the genome of *N. parkeri*, we generated genome alignments using Lastz (http://www.bx.psu.edu/miller_lab/) with parameters T=2, C=2, H=2000, Y=3400, L=6000, K=2200. The self-versus-self approach was performed on the repeat-masked genome. Alignment blocks were then joined using CHAINNET (13). We selected block-chains with lengths >500 bp and similarity >85%. Next, we extracted the sequences in the chains from the unmasked genome and aligned them again using Lastz. The resulting alignments with lengths extended to 1kb and having the identity greater than 90% were considered to be segmental duplications. We used the same method to explore segmental duplications in *X. tropicalis*.

To infer the transposable element correlations between *X. tropicalis* and *N. parkeri* geomes, we used 2 Mb non-overlapping sliding windows to calculate the numbers of different TE classes among the windows. Then we calculated the correlation coefficient for each two classes using R package, which was subsequently used (1- correlation coefficient) as the distance to cluster these TEs.

Annotation

We utilized three methods to predict protein-coding genes in the genome: homology-based, *ab initio*, and RNA-seq based predictions. We combined these results to establish a final gene-set. For homology-based gene prediction, we mapped protein sequences of *X. tropicalis*, *H. sapiens*, and *A. carolinensis* to the draft assembly using TblastN (14). We filtered the aligned sequences and query proteins, and passed them to GeneWise (15) to obtain accurate spliced alignments. For the RNA-based prediction, we first aligned the reads of *N. parkeri* to the genome using TopHat (16, 17) to identify candidate exon regions. We used Cufflinks (18) to construct transcripts. Predicted open reading frames (ORF) identified reliable transcripts via HMM-based training parameters. In addition, AUGUSTUS (19) was used to predict protein-coding genes based parameters trained from 1000 high-quality proteins from both the homolog-based predictions and RNA-seq predictions. An in-house pipeline (Augap) efficiently integrated the Cufflinks and AUGUSTUS gene-sets based on gene-structure. Finally, all gene evidence was merged to form a comprehensive and non-redundant gene-set. Gene functions were assigned according to the best match of their alignments using Blastp to the SwissProt and TrEMBL databases (20). Motifs and domains of genes were determined by InterProScan (21) against protein databases including ProDom, PRINTS, Pfam, SMART, PANTHER and PROSITE. Gene Ontology (GO; (22) IDs for each gene were obtained from the corresponding SwissProt and TrEMBL entries. All genes were aligned against the KEGG (23) proteins to detect possible gene pathways.

Genome Rearrangements Comparison

We compared the genome of *N. parkeri* to those of *Xenopus*, zebra finch, ostrich, human, platypus, soft-shell turtle, and green anole lizard. We used the reference gene-set of *X. tropicalis*. Protein-coding sequences for the other species were downloaded from Ensembl Database. First, we performed pairwise whole-genome alignments for each taxon using Lastz. Next, we joined the alignments using CHAINNET and then defined synteny blocks. An all-vs-all BLASTP alignment (<1e-5) was performed on the predicted polypeptides of every protein-coding gene. Single-copy orthologs were determined if (1) a reciprocal best hit existed, and (2) the identity score between the gene and its best hit was 20% larger than its second-best hit. These criteria allowed us to identify orthologous pairs among the amphibians, reptiles, including birds, and mammals. Additionally, the synteny blocks were used to determine the order of the single-copy orthologs. The total numbers of indels, translocations, and reversals of the genomic blocks were explored using a dynamic programming script. Finally we estimated the rate of genomic rearrangements between *X. tropicalis* and *N. parkeri* to be 0.043 per 100Mb per 1M yr, a rate similar to that of reptiles (0.039), but much smaller than that of either birds (0.128) or mammals (0.101).

Reconstruction of Ancestor Homologous Synteny Blocks (aHSB)

We performed whole genome alignment between Human, Chicken, *N. parkeri*, *X. tropicalis* using Lastz (http://www.bx.psu.edu/miller_lab/). Human was aligned in sequence to Chicken, *N. parkeri* and *X. tropicalis*. Alignment nets that representing the putative orthologous regions was created using chainNet (13). By merging co-linear alignments with inferCARs algorithm (24), the aHSB of tetrapod were then reconstructed. The breakpoints that refer to adjacent segments differed in the ancestor and the target species were identified in these aHSB. During this anlaysis, the genomes of human and chicken were downloaded from UCSC. The *X. tropicalis* genome was downloaded from NCBI (Xtropicalis_v7).

Detection of Highly Conserved Elements (HCEs)

To detect HCEs in these tetrapod lineages, we generated pair-wise alignments (Lastz) and multiple alignments (MULTIZ) of human, chicken, *N. parkeri*, and *X. tropicalis* with human as the reference. We used PhaseCons (25) to estimate the genome conservation index and further identified the HCEs. Briefly, we first used phyloFit to estimate a initial neutral phylogenetic model (also considered as the nonconserved model in PhastCons). Then we ran PhastCons twice, with first for estimation of conserved and nonconserved models and second for prediction of conserved elements, respectively. Finally, we identified 194,567 tetrapod HCEs, covering 12 Mb in the human genome. We found 2466 human Ref genes have more than 50% overlapping coding sequences with the HCEs. We performed GO term enrichment analysis of these genes. Significant GO terms (P <0.05) were identified using Fisher's exact test (Table S15). To investigate evolutionary constraint in amphibian lineage, the method used was the same as above. Finally, we identified 278,843 amphibian HCEs, covering 22Mb in the *N. parkeri* genome. We also sought to find some genes that are highly conserved in amphibians but divergent in Human and Chicken in terms of the coverage in the coding regions. We identified 217 genes whose coding regions are ≥70% covered by amphibian HCEs and ≤10% covered by tetrapod HCEs. The enriched GO categories for these genes are shown in Table S16.

Gene Family Expansion and Contraction

A gene family was considered to constitute a group of similar genes descended from a single gene in the last common ancestor of the targeted species. To evaluate the dynamic evolutionary processes in gene families, the protein-coding genes we downloaded the genomes of *A. carolinensis*, *G. gallus*, *H. sapiens*, *X. tropicalis* and *D. rerio* from the Ensembl Database. These data and the genes of *N. parkeri* were analyzed using TreeFam (26) to define gene families. For each species, the longest translation transcript was selected to represent each gene. We filtered

out genes with length < 50% of the median length within each family.

Expansions of gene families were obtained using CAFE (27), which uses a random birth and death model to study gene-gain and -loss in each family over a given phylogeny. A conditional *p*-value was calculated for each gene family and families with $p < 0.05$ were considered to have a significantly accelerated rate of expansion and/or contraction. GO enrichment for both the expansion and contraction gene families were analyzed.

Lineage-Specific Genes Analysis

To further characterize lineage-specific genes of the Tibetan frog, we first searched the above gene families that contained genes of *N. parkeri* only, and then used their protein sequences to search the other five species using BLAST; if the best hit had an identical portion that spanned over 40% of the gene's protein sequence length, then the gene was filtered out. A similar method was used to search for amphibian-specific genes (genes that are conserved between *N. parkeri* and *X. tropicalis* with an identity of greater than 60%, but showed less that 40% identity when aligned against the other four species). Similarly, aquatic animal-specific genes and the terrestrial animal-specific genes were also identified.

Divergence Time Estimation

Single-copy gene families (those with only one copy in each species) that had strong orthologous relationships were used to reconstruct a phylogeny and estimate the divergence times of the six species. Following alignment of protein sequences using Muscle43, the genes were concatenated into a 'supergene' for each species. Amino acid substitution model JTT + Gamma + I was used in PhyML (28) to reconstruct the phylogeny; the gamma parameter and proportion of invariable sites were derived from the maximum likelihood estimates. The root was determined by minimizing height of the whole tree via Treebest (http://treesoft.sourceforge.net/treebest.shtml). MCMCtree (29) in the PAML package was implemented to estimate the time of divergence for each ancestral node. Calibration-times were obtained from the TimeTree database (http://www.timetree.org/) (Supplementary Table S12).

Mutation Rate Estimation

All-vs-all blastp searches was used to identify one-to-one orthologs between *N. parkeri*, *X. tropicalis* and human, during which the human was separately aligned to *N. parkeri* and *X. tropicalis*. In total, 9964 orthologous genes were identified. Multiple sequence alignments for each

ortholog were reconstructed using MAFFT (30). 1,208,217 four-fold degenerate sites were extracted and concatenated to form a three-way alignment. A divergence time of 266 million years for the two frogs was used to obtain the mutation rate under a global clock model using baseml in PAML.

Supplementary Tables

Table S1. Statistics for the DNA libraries. Short insert size paired-end libraries contained 170bp, 250bp, 500bp, and 800bp; 100bp or 150bp sequence-lengths were generated for both ends. The remaining libraries belonged to the long insert size paired-end libraries and 49bp of sequences obtained for both ends were used to build scaffolds.

Pair-end libraries	**Insert size (bp)**	**Read lengths(bp)**	**Total data(Mb)**	**Sequence coverage (X)**	**High-quality size (G)**	**High-quality coverage (x)**
Solexa Reads	170	100	37,498	16.304	33,987	14.777
	250	150	45,321	19.705	36,580	15.904
	500	100/150	71,048	30.891	48,147	20.934
	800	100	31,320	13.617	25,892	11.257
	2k	49	40,336	17.538	26,525	11.533
	5k	49	35,323	15.358	10,647	4.629
	10k	49	29,151	12.674	5,087	2.212
	20k	49	33,022	14.358	3,970	1.726
Total			**323,020**	**140.443**	**190,834**	**82.971**

Table S2. Seventeen-mer statistics for genome size estimation of *Nanorana parkeri*.

Statistic	Value
kmer	17
Kmer_num	55,450,398,715
pkdepth	24
genome_size	2,310,433,279
used_base	68,931,334,555
Read_num	842,558,490
X	9.529.83

Table S3. Statistics for the assembly of the genome of *Nanorana parkeri*.

	Soap Denovo		**SSPACE**
	Contig (bp)	**Scaffold (bp)**	**Scaffold (bp)**
N50	8,142	762,232	1,055,271
Longest	90,564	5,650,237	8,606,022
Total Number(>100bp)	568,375	139,595	135,808
Total Size	1,991,133,680	2,054,418,562	2,071,743,715

Table S4. Combination of the homology and *de novo* gene-sets. Genes in the genome of *Nanorana parkeri* were predicted using *de novo* and homology-based methods based on the protein-coding sequences of *Xenopus tropicalis* and *Homo sapiens*.

	Gene set	Gene	Average transcript length (bp)	Average CDS length (bp)	Average exons per gene	Average exon length (bp)	Average intron length (bp)
Homolog	*Xenopus tropicalis*	26,805	18020	1323	6	222	3359
	Homo sapiens	21,042	24900	1288	6.9	185	3969
	Anolis carolinensis	23,293	17016	1243	5.7	216	3332
De novo	Augap	23,811	23859	1433	6.8	212	3885
Final Gene set		23,408	24408	1382	6.9	199	3883

Note: Augap is an in-house pipeline that integrates the Augustus and the Cufflinks gene-sets.

Table S5. Proteomic prediction and functional annotation of the genome of *Nanorana parkeri*. Predictions were based on the functional protein databases SwissProt and TrEMBL from UniProt (http://www.uniprot.org), InterPro (http://www.ebi.ac.uk/interpro) and their associated GO annotations, and KEGG pathways for metabolic and cellular processes. About 96% of the gene-set was well annotated.

	Number	Percent(%)
Total	23408	
Annotated	22459	95.95
SwissProt	21523	91.95
TrEMBL	22315	95.33
InterPro	18362	78.44
KEGG	16154	69.01
GO	13375	57.14
Unannotated	949	4.05

Table S6. Comparison of the TE content of the *Nanorana parkeri* and *Xenopus tropicalis* genomes.

Type	*Nanorana parkeri*		*Xenopus tropicalis*	
	Repeat Size(bp)	% of Genome	Repeat Size(bp)	% of Genome
TRF	113,080,908	5.46	103,016,463	6.81
RepeatMasker	132,102,110	6.38	442,506,632	29.27
RepeatProteinMask	140,697,073	6.79	86,379,759	5.71
De novo	838,804,555	40.49	528,064,710	34.93
Total	970,122,551	46.83	652,417,394	43.16

Table S7. Subclassifation of the TE families in the genomes of *Nanorana parkeri* and *Xenopus tropicalis*.

	N. parkeri			*X. tropicalis*		
TE Classification	**Total length(Mb)**	**TE coverage (%)**	**Total genome coverage (%)**	**Total length(Mb)**	**TE coverage (%)**	**Total genome coverage (%)**
DNA Transposon	**243.0**	**25.1**	**11.7**	**365.9**	**56.1**	**24.2**
TcMar-Tc1	50.3	5.2	2.4	20.6	3.2	1.4
En-Spm	47.4	4.9	2.3	3.6	0.6	0.2
MuDR	22.5	2.3	1.1	9.3	1.4	0.6
hAT-Ac	16.5	1.7	0.8	36.9	5.7	2.4
hAT-Charlie	16.4	1.7	0.8	28.9	4.4	1.9
Harbinger	6.5	0.7	0.3	51.9	8.0	3.4
T2	2.0	0.2	0.1	35.8	5.5	2.4
Kolobok-T2	0.1	0.0	0.0	96.2	14.7	6.4
PIF-Harbinger	0.7	0.1	0.0	61.5	9.4	4.1
LTR	**378.7**	**39.0**	**18.3**	**162.2**	**24.9**	**10.7**
Gypsy	170.4	17.6	8.2	45.9	7.0	3.0
ERV1	51.5	5.3	2.5	19.6	3.0	1.3
Copia	48.9	5.0	2.4	12.6	1.9	0.8
ERVK	26.8	2.8	1.3	2.6	0.4	0.2
Pao	16.2	1.7	0.8	3.5	0.5	0.2
DIRS1	10.6	1.1	0.5	3.3	0.5	0.2
DIRS	9.6	1.0	0.5	6.8	1.0	0.4

	Ngaro	6.5	0.7	0.3	1.5	0.2	0.1
LINE		**218.6**	**22.5**	**10.6**	**111.8**	**17.1**	**7.4**
	L1	84.6	8.7	4.1	25.1	3.8	1.7
	CR1	35.1	3.6	1.7	37.9	5.8	2.5
	L2	30.3	3.1	1.5	24.1	3.7	1.6
	Penelope	26.9	2.8	1.3	18.6	2.8	1.2
	R2	15.1	1.6	0.7	1.0	0.2	0.1
SINE		**17.0**	**1.8**	**0.8**	**4.7**	**0.7**	**0.3**
Tandem repeats		**113.1**	**11.7**	**5.5**	**103.0**	**15.8**	**6.8**
Total		**970.1**	**100.0**	**46.8**	**652.4**	**100.0**	**43.2**

Table S8.1. GO enrichment for expanded gene families in *N. parkeri*. For each GO subcategory, a 2 X 2 contingency table was constructed by recording the numbers of families included or not included in a category of 'genome background' families and contracted families. Two-tailed X^2 tests were used to calculate statistical significance.

GO Term	Description	FDR q-value	Enrichment
GO:0004984	olfactory receptor activity	1.20E-80	19
GO:0004930	G-protein coupled receptor activity	3.01E-57	6.23
GO:0004888	transmembrane signaling receptor activity	7.41E-47	4.41
GO:0038023	signaling receptor activity	1.91E-45	4.12
GO:0004872	receptor activity	1.83E-43	3.69
GO:0004871	signal transducer activity	9.77E-36	3.26
GO:0060089	molecular transducer activity	8.37E-36	3.26
GO:0008527	taste receptor activity	5.55E-06	14.29
GO:0005198	structural molecule activity	1.01E-05	2.39

GO:0005201	extracellular matrix structural constituent	4.18E-05	5.61
GO:0008417	fucosyltransferase activity	1.22E-04	12.5
GO:0046920	alpha-(1->3)-fucosyltransferase activity	2.26E-03	14.51
GO:0017060	3-galactosyl-N-acetylglucosaminide 4-alpha-L-fucosyltransferase activity	2.27E-02	23.22
GO:0005003	ephrin receptor activity	2.38E-02	9.68
GO:0003950	NAD+ ADP-ribosyltransferase activity	2.58E-02	7.33
GO:0004982	N-formyl peptide receptor activity	2.49E-02	13.27
GO:0032190	acrosin binding	6.72E-02	17.42
GO:0005004	GPI-linked ephrin receptor activity	6.35E-02	17.42
GO:0008066	glutamate receptor activity	8.41E-02	5.81
GO:0001640	adenylate cyclase inhibiting G-protein coupled glutamate receptor activity	1.38E-01	13.93
GO:0048407	platelet-derived growth factor binding	1.56E-01	8.44

Table S8.2. GO enrichment for expanded gene families in X. tropicalis. For each GO subcategory, a 2 X 2 contingency table was constructed by recording the numbers of families included or not included in a category of 'genome background' families and contracted families. Two-tailed X2 tests were used to calculate statistical significance.

GO Term	Description	P-value	FDR q-value	Enrichment
GO:0004252	serine-type endopeptidase activity	3.15E-15	4.02E-12	7.96
GO:0008236	serine-type peptidase activity	1.32E-14	8.45E-12	7.61
GO:0017171	serine hydrolase activity	1.32E-14	5.63E-12	7.61
GO:0004175	endopeptidase activity	2.21E-13	7.05E-11	5.63
GO:0004984	olfactory receptor activity	2.44E-13	6.24E-11	7.78

GO:0004930	G-protein coupled receptor activity	4.52E-10	9.63E-08	3.17
GO:0070011	peptidase activity, acting on L-amino acid peptides	1.19E-09	2.18E-07	4.04
GO:0008233	peptidase activity	1.84E-09	2.94E-07	3.96
GO:0004888	transmembrane signaling receptor activity	2.05E-07	2.91E-05	2.4
GO:0038023	signaling receptor activity	1.79E-06	2.29E-04	2.2
GO:0004871	signal transducer activity	1.54E-05	1.78E-03	1.99
GO:0060089	molecular transducer activity	1.54E-05	1.64E-03	1.99
GO:0004872	receptor activity	2.32E-05	2.28E-03	1.96
GO:0003677	DNA binding	5.65E-05	5.16E-03	1.96
GO:0008527	taste receptor activity	5.94E-05	5.06E-03	8.11
GO:0004806	triglyceride lipase activity	4.95E-04	3.95E-02	7.78

Table S9.1. Forty-four genes of *Xenopus tropicalis* that do not occur in the Tibetan frog genome.

ID	Symbol	Description
ENSXETP00000053297	c16orf61	chromosome 16 open reading frame 61
ENSXETP00000058799	il28ra	interleukin 28 receptor, alpha (interferon, lambda receptor) precursor
ENSXETP00000053944	dazl	deleted in azoospermia-like
ENSXETP00000060265		uncharacterized protein
ENSXETP00000015592		uncharacterized protein
ENSXETP00000059869		uncharacterized protein
ENSXETP00000061814		uncharacterized protein
ENSXETP00000014993	fads6	fatty acid desaturase domain family, member 6
ENSXETP00000022010	spata20	spermatogenesis associated 20

ENSXETP00000060658	HIRIP3	HIRA interacting protein 3
ENSXETP00000058977	LYRM7	LYR motif-containing protein 7
ENSXETP00000063620	IL-21	interleukin-21 precursor
ENSXETP00000059113		Putative interleukin-2; Uncharacterized protein
ENSXETP00000063023	LAD1	ladinin 1
ENSXETP00000062478	ag1	anterior gradient 1
ENSXETP00000010586	ag1	anterior gradient 1
ENSXETP00000058693	PTCD2	pentatricopeptide repeat domain 2
ENSXETP00000043347	c1orf35	chromosome 1 open reading frame 35
ENSXETP00000013557	nt5dc1	5'-nucleotidase domain containing 1
ENSXETP00000059686	PINLYP	phospholipase A2 inhibitor and LY6/PLAUR domain containing
ENSXETP00000058695	PINLYP	phospholipase A2 inhibitor and LY6/PLAUR domain containing
ENSXETP00000061517	IL5RA	interleukin 5 receptor, alpha precursor
ENSXETP00000059176	MGC147402	uncharacterized protein LOC780132
ENSXETP00000018644	snrnp27	small nuclear ribonucleoprotein 27kDa (U4/U6.U5)
ENSXETP00000017916	ccdc73	coiled-coil domain containing 73
ENSXETP00000060006	TMEM25	transmembrane protein 25
ENSXETP00000060169	TMEM25	transmembrane protein 25
ENSXETP00000059433		uncharacterized protein LOC100170428
ENSXETP00000037150	tmem136	transmembrane protein 136
ENSXETP00000063961		uncharacterized protein
ENSXETP00000059420		secretory calcium-binding phosphoprotein acidic 1 precursor
ENSXETP00000063512	CCDC106	coiled-coil domain containing 106
ENSXETP00000061827	rab11fip4l	rab11 family-interacting protein 4-like
ENSXETP00000055973	nmu	neuromedin U
ENSXETP00000063278	nmu	neuromedin U

ENSXETP00000006606	hmgn1	high mobility group nucleosome binding domain 1
ENSXETP00000018440	tnip3	TNFAIP3 interacting protein 3
ENSXETP00000043487		uncharacterized protein
ENSXETP00000058912	SFXN4	sideroflexin 4
ENSXETP00000058851	SFXN4	sideroflexin 4
ENSXETP00000000741		Uncharacterized protein
ENSXETP00000063968	IL-15	interleukin 15
ENSXETP00000053339	zmat5	zinc finger, matrin-type 5
ENSXETP00000045379		uncharacterized protein

Table S9.2. GO terms enriched by the missing genes of *Nanorana*.

GO_ID	GO_Term	GO_Class	Pvalue	AdjustedPv
GO:0006940	regulation of smooth muscle contraction	BP	1.14E-06	6.68E-05
GO:0005125	cytokine activity	MF	0.000453562	0.011686251
GO:0005126	cytokine receptor binding	MF	0.000588309	0.011686251
GO:0045948	positive regulation of translational initiation	BP	0.001094032	0.011686251
GO:0060965	negative regulation of gene silencing by miRNA	BP	0.001094032	0.011686251
GO:0070935	3'-UTR-mediated mRNA stabilization	BP	0.001094032	0.011686251
GO:0032019	mitochondrial cloud	CC	0.001094032	0.011686251
GO:0008494	translation activator activity	MF	0.001094032	0.011686251
GO:0044427	chromosomal part	CC	0.001912983	0.019034365
GO:0043515	kinetochore binding	MF	0.002186927	0.019034365
GO:0003730	mRNA 3'-UTR binding	MF	0.003278686	0.02407785

GO:0016071	mRNA metabolic process	BP	0.0046795	0.0323436
GO:0031492	nucleosomal DNA binding	MF	0.0054588	0.036651944
GO:0005615	extracellular space	CC	0.00692194	0.041722796
GO:0006629	lipid metabolic process	BP	0.007259769	0.041722796
GO:0008354	germ cell migration	BP	0.007634384	0.041722796
GO:0045495	pole plasm	CC	0.007634384	0.041722796
GO:0006955	immune response	BP	0.008898154	0.04449077

Table S10.1. Pseudogenes in the genome of *Nanorana*.

ID	Symbol	Discription
ENSXETP00000030835	cbr1	carbonyl reductase 1
ENSXETP00000030837	cbr3	carbonyl reductase 3
ENSXETP00000020494	mttp.1	microsomal triglyceride transfer protein, gene 1
ENSXETP00000030764	rarres1	retinoic acid receptor responder (tazarotene induced) 1
ENSXETP00000015828	fam192a	uncharacterized protein LOC100145298
ENSXETP00000006559		uncharacterized protein
ENSXETP00000046394	trim33	tripartite motif containing 33
ENSXETP00000059804	dock9	dedicator of cytokinesis 9
ENSXETP00000003971	arhgap26	Rho GTPase activating protein 26
ENSXETP00000019772	galm	galactose mutarotase (aldose 1-epimerase)
ENSXETP00000030358		uncharacterized protein
ENSXETP00000060957	maea	macrophage erythroblast attacher
ENSXETP00000023313	tm9sf3	transmembrane 9 superfamily member 3
ENSXETP00000063698		uncharacterized protein
ENSXETP00000055332	timm23	translocase of inner mitochondrial membrane 23

		homolog
ENSXETP00000030368	atg16l1	ATG16 autophagy related 16-like 1
ENSXETP00000024292		uncharacterized protein
ENSXETP00000059913	slc7a15	solute carrier family 7 (cationic amino acid transporter, y+ system), member 15
ENSXETP00000060333	esam	endothelial cell adhesion molecule precursor
ENSXETP00000005128	mogs	mannosyl-oligosaccharide glucosidase
ENSXETP00000001418	chchd5	coiled-coil-helix-coiled-coil-helix domain containing 5
ENSXETP00000024313		uncharacterized protein
ENSXETP00000020935	hp1bp3	heterochromatin protein 1, binding protein 3
ENSXETP00000004854	XB-GENE-5863396	MGC84255 protein
ENSXETP00000056440		uncharacterized protein
ENSXETP00000063542		uncharacterized protein
ENSXETP00000059835		uncharacterized protein
ENSXETP00000063796	ss18	synovial sarcoma translocation, chromosome 18
ENSXETP00000061239		uncharacterized protein
ENSXETP00000029796	rps8	ribosomal protein S8
ENSXETP00000059692	fam135a	family with sequence similarity 135, member A
ENSXETP00000045168	iars	isoleucyl-tRNA synthetase
ENSXETP00000034321	avpr2.2	arginine vasopressin receptor 2 (nephrogenic diabetes insipidus), gene 2
ENSXETP00000012900	XB-GENE-5830282	hypothetical protein LOC100145511
ENSXETP00000062705	SH3TC2	SH3 domain and tetratricopeptide repeats 2
ENSXETP00000007344	dbf4b	DBF4 homolog B (S. cerevisiae)
ENSXETP00000002418		uncharacterized protein
ENSXETP00000035178	CBLB	Cbl proto-oncogene, E3 ubiquitin protein ligase B

ENSXETP00000007378	ccdc103	coiled-coil domain containing 103
ENSXETP00000032362	ccdc122	coiled-coil domain containing 122
ENSXETP00000001685	egfl7	EGF-like-domain, multiple 7
ENSXETP00000059095	WNT10A	wingless-type MMTV integration site family, member 10A
ENSXETP00000036668	btf3	basic transcription factor 3
ENSXETP00000013580	ccdc13	coiled-coil domain containing 13
ENSXETP00000000947	erbb2ip	erbb2 interacting protein
ENSXETP00000052227	ipo7	importin 7
ENSXETP00000044513	gipc1	GIPC PDZ domain containing family, member 1
ENSXETP00000063281	nfrkb	nuclear factor related to kappaB binding protein
ENSXETP00000060564	ISCA2	iron-sulfur cluster assembly 2 homolog (S. cerevisiae)
ENSXETP00000026265	dirc2	disrupted in renal carcinoma 2
ENSXETP00000029054	dlst	dihydrolipoamide S-succinyltransferase (E2 component of 2-oxo-glutarate complex)
ENSXETP00000024057	mkl2	MKL/myocardin-like 2
ENSXETP00000062038	polr3e	RNA polymerase III polypeptide E
ENSXETP00000059233		uncharacterized protein
ENSXETP00000032125	esrp1	epithelial splicing regulatory protein 1
ENSXETP00000004933	rims1	regulating synaptic membrane exocytosis 1
ENSXETP00000052749	cdc6	cell division cycle 6 homolog
ENSXETP00000032909	hhipl2	HHIP-like 2
ENSXETP00000061466	ALG13	asparagine-linked glycosylation 13 homolog (S. cerevisiae)
ENSXETP00000018286	igsf9b	immunoglobulin superfamily, member 9B
ENSXETP00000031981		uncharacterized protein

ENSXETP00000045673	RORB	RAR-related orphan receptor B
ENSXETP00000051396	atp1a3	ATPase, Na+/K+ transporting, alpha 3 polypeptide
ENSXETP00000010293		uncharacterized protein
ENSXETP00000001581	prlr	prolactin receptor
ENSXETP00000025357	dlgap2	discs, large homolog-associated protein 2
ENSXETP00000048372	smad4.1	SMAD family member 4, gene 1
ENSXETP00000063057	dcaf8	DDB1 and CUL4 associated factor 8
ENSXETP00000050631	dhodh	dihydroorotate dehydrogenase
ENSXETP00000011400	pbx1	pre-B-cell leukemia homeobox 1
ENSXETP00000013341	pdap1	PDGFA associated protein 1
ENSXETP00000060804		uncharacterized protein
ENSXETP00000010280	selp	selectin P (granule membrane protein 140kDa, antigen CD62)
ENSXETP00000041127	TAPBPL	TAP binding protein-like
ENSXETP00000043891	gabbr1	gamma-aminobutyric acid (GABA) B receptor, 1
ENSXETP00000007193		uncharacterized protein
ENSXETP00000038836	USP48	ubiquitin specific peptidase 48
ENSXETP00000017332	cxcl14	chemokine (C-X-C motif) ligand 14
ENSXETP00000061366	MAP9	microtubule-associated protein 9
ENSXETP00000014317	rnf126	ring finger protein 126
ENSXETP00000059198		uncharacterized protein
ENSXETP00000028848	gcg	glucagon
ENSXETP00000022904	ccdc81	coiled-coil domain containing 81
ENSXETP00000058325	ZDHHC23	zinc finger, DHHC-type containing 23
ENSXETP00000047741	capn13	calpain 13
ENSXETP00000024139	map2k4	mitogen-activated protein kinase kinase 4

ENSXETP00000025700	grm5	glutamate receptor, metabotropic 5
ENSXETP00000024171	spag9	sperm associated antigen 9
ENSXETP00000059609		
ENSXETP00000060581	RPGR	retinitis pigmentosa GTPase regulator
ENSXETP00000000822		uncharacterized protein
ENSXETP00000009046	XB-GENE-1014282	Putative ortholog of protocadherin gamma B5, 5 of 5
ENSXETP00000042238	pik3r6	phosphoinositide-3-kinase, regulatory subunit 6
ENSXETP00000011096	yipf6	Yip1 domain family, member 6
ENSXETP00000029916	mknk2	MAP kinase interacting serine/threonine kinase 2
ENSXETP00000050392	gsc	goosecoid homeobox
ENSXETP00000054267	rpl7	ribosomal protein L7
ENSXETP00000055429	XB-GENE-5726497	hypothetical LOC496989
ENSXETP00000031774	rps6ka6	ribosomal protein S6 kinase, 90kDa, polypeptide 6
ENSXETP00000023482	tubd1	tubulin, delta 1
ENSXETP00000055412	XB-GENE-5889123	hypothetical LOC496784
ENSXETP00000058942	hn1	hematological and neurological expressed 1
ENSXETP00000055875	TRDC	T cell receptor delta constant
ENSXETP00000012700	llgl1	lethal giant larvae homolog 1
ENSXETP00000024430	rhebl1	Ras homolog enriched in brain like 1
ENSXETP00000053075	hbe1	hemoglobin, epsilon 1
ENSXETP00000006019	josd2	Josephin domain containing 2
ENSXETP00000002310	dmap1	DNA methyltransferase 1 associated protein 1
ENSXETP00000056886		uncharacterized protein
ENSXETP00000062023		uncharacterized protein
ENSXETP00000036971	ltbp4	latent transforming growth factor beta binding protein 4
ENSXETP00000029371	plekhm2	pleckstrin homology domain containing, family M (with

		RUN domain) member 2
ENSXETP00000053772	c1orf131	chromosome 1 open reading frame 131
ENSXETP00000034398	adamtsl3	ADAMTS-like 3
ENSXETP00000019067	asb14	ankyrin repeat and SOCS box containing 14
ENSXETP00000035675	stil	SCL/TAL1 interrupting locus
ENSXETP00000035667	cmpk1	cytidine monophosphate (UMP-CMP) kinase 1, cytosolic
ENSXETP00000001428		uncharacterized protein
ENSXETP00000019498	hook1	hook homolog 1
ENSXETP00000043293		Uncharacterized protein
ENSXETP00000061282	hist1h2ah	histone cluster 1, H2ah
ENSXETP00000058458	DYNC1H1	dynein, cytoplasmic 1, heavy chain 1
ENSXETP00000062391		uncharacterized protein
ENSXETP00000041159	camk2b	calcium/calmodulin-dependent protein kinase (CaM kinase) II beta
ENSXETP00000043240	ttc4	tetratricopeptide repeat domain 4
ENSXETP00000060650	HUWE1	HECT, UBA and WWE domain containing 1, E3 ubiquitin protein ligase
ENSXETP00000061578	RELB	v-rel reticuloendotheliosis viral oncogene homolog B
ENSXETP00000043096	dennd4b	DENN/MADD domain containing 4B
ENSXETP00000003954	taf3	TAF3 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 140kDa
ENSXETP00000039680	rufy1	RUN and FYVE domain containing 1
ENSXETP00000062733	C5orf25	chromosome 5 open reading frame 25
ENSXETP00000033036	ttc39c	tetratricopeptide repeat domain 39C
ENSXETP00000058566		uncharacterized protein

ENSXETP00000060029		uncharacterized protein
ENSXETP00000062433		uncharacterized protein
ENSXETP00000059931		uncharacterized protein
ENSXETP00000060900		uncharacterized protein
ENSXETP00000061052		uncharacterized protein
ENSXETP00000062261		uncharacterized protein
ENSXETP00000030723	ado	2-aminoethanethiol (cysteamine) dioxygenase
ENSXETP00000062925		uncharacterized protein
ENSXETP00000054435		uncharacterized protein
ENSXETP00000059165		uncharacterized protein
ENSXETP00000063452		uncharacterized protein
ENSXETP00000059581		histone H2A
ENSXETP00000056119		histone H2A
ENSXETP00000056584	HIST1H2BJ	histone cluster 1, H2bj
ENSXETP00000060607		uncharacterized protein

Table S10.2. GO terms enriched by the pseudogenes of *Nanorana*.

GO_ID	GO_Term	GO_Class	Pvalue	AdjustedPv
GO:0015074	DNA integration	BP	1.13E-06	0.000806596
GO:0006259	DNA metabolic process	BP	5.63E-05	0.02007535
GO:0071844	cellular component assembly at cellular level	BP	0.000103433	0.024582461
GO:0000785	chromatin	CC	0.000148587	0.026485693
GO:0034622	cellular macromolecular complex assembly	BP	0.000227415	0.032429347

GO:0000786	nucleosome	CC	0.000608864	0.036650922
GO:0006334	nucleosome assembly	BP	0.000775704	0.036650922
GO:0006139	nucleobase-containing compound metabolic process	BP	0.001418283	0.046494869

Table S11. Statistic summary of segmental duplication events in the two frog genomes.

	Nanorana parkeri		*Xenopus tropicalis*	
Copy number	**Clusters**	**Total length of DNA segments (bp)**	**Clusters**	**Total length of DNA segments (bp)**
>=100	0	0	19	388306
>=50	0	0	75	728616
>=20	15	173420	512	2533018
>=10	152	803839	1593	6182641
>=5	961	4060463	4974	13504693
>=2	22781	41087419	70771	125530808

Table S12. Comparison of genome rearrangement comparison among amphibian, reptile, bird and mammal.

	Split time (Myr)	**Orthologs**	**Orthologs in blocks**	**Conserved size (100Mbp)**	**# Transfer**	**# Indel**	**# Reverse**	**# Total mutations**	**Rearrangement rate**
Human vs Platypus	166	11372	10441	15.774	33	405	90	528	0.100821697
X. tropicalis* vs *N. parkeri	256	11499	11296	10.872	22	162	53	237	0.0425764
Zebra finch vs Ostrich	110	10756	10740	11.566	76	127	123	326	0.128118466
Soft-shell turtle vs Green anole lizard	257	11796	11192	14.31	33	175	82	290	0.039427137

Note: # Transfer = number of translocation of blocks; # Indel = number of insertion and deletion of blocks; # Reverse = number of reverse blocks.

Table S13.1. Conserved segments shared among human, chicken, *X. tropicalis* and *N. parkeri*, which were used in reconstructing aHSBs (ancestral homologous synteny blocks).

No. aligned *N. parkeri* scaffolds (total length)	1482 (1.47Gb, 71%)
No. aligned *X. tropicalis* chromosomes (total length)	10 (1.27Gb, 89%)
Total length of conserved segment in *X. tropicalis*	490 Mb (34%)
Total length of conserved segment in *N. parkeri*	665 Mb (32%)
Total length of conserved segment in Human	1301Mb (41%)
Total length of conserved segment in Chicken	491Mb (45%)
No. of conserved segments	2370

Table S13.2. Statistics of reconstructed ancestral homologous synteny blocks (aHSBs)

No. reconstructed aHSBs	114
No. aHSBs that correspond to only one chicken chromosome	105
No. aHSBs that correspond to only one human chromosome	74
Max. length of aHSBs	76 Mb
Min. length of aHSBs	26 Kb
Max. no. *N. parkeri* scaffolds in aHSBs	175
Min. no. *N. parkeri* scaffolds in aHSBs	1

Table S14.Intrachromosomal synteny comparison between human, chicken, *X. tropicalis* and *N. parkeri*.

aHSB ID	Size (bp)	total conserved segment	syntenic conserved segments in Human	# syntenic conserved segments in Chicken	# syntenic conserved segments in *X. tropicalis*	# syntenic conserved segments in *N. parkeri*
aHSB 39	76430651	175	137	129	118	66
aHSB 30	44152525	141	126	122	120	40
aHSB 62	35052695	90	74	68	58	26
aHSB 29	31207996	84	76	71	63	29
aHSB 47	30943496	92	71	69	64	24
aHSB 10	27228105	64	51	51	43	26
aHSB 57	26461476	97	93	91	75	41
aHSB 78	24674973	92	77	74	66	30
aHSB 32	22992509	95	79	69	61	28
aHSB 91	21897753	68	26	26	25	19
aHSB 89	18970721	48	28	26	26	14
aHSB 48	15818327	56	46	45	33	32
aHSB 22	14722092	58	45	42	42	31
aHSB 51	13896763	50	48	46	43	20
aHSB 14	13743741	21	17	16	10	7
aHSB 76	13245122	30	24	24	14	5
aHSB 94	12813742	69	46	44	44	27
aHSB 34	11745640	25	25	25	24	9
aHSB 58	10882868	58	24	21	21	20
aHSB 20	10428188	15	11	9	6	6

Notes: 1. Human genome used as reference to perform alignment and assign orientation, majorly because of its high quality. Thus, the intra-chromosomal synteny in human might overestimated. 2. As *N. parkeri* and *X. tropicalis* have no chromosomal assemblies. The intra-chromosomal synteny may not be complete (only in scaffold-level).

Table S15. Enriched GO terms in the genes near to / within the HCEs that are shared among human, chicken, *N. parkeri* and *X. tropicalis*.

GO term	**Description**	**GO Class**	P-value	FDR q-value	Enrichment (N, B, n, b)
GO:0007186	G-protein coupled receptor signaling pathway	BP	5.52E-14	4.22E-10	2.11 (2373,100,832,74)
GO:0015988	energy coupled proton transmembrane transport, against electrochemical gradient	BP	8.66E-13	3.31E-09	14.13 (2373,11,168,11)
GO:0015991	ATP hydrolysis coupled proton transport	BP	8.66E-13	2.20E-09	14.13 (2373,11,168,11)
GO:0006818	hydrogen transport	BP	8.82E-13	1.68E-09	13.04 (2373,13,168,12)
GO:0015992	proton transport	BP	8.82E-13	1.35E-09	13.04 (2373,13,168,12)
GO:0050690	regulation of defense response to virus by virus	BP	7.80E-12	9.93E-09	17.17 (2373,16,95,11)
GO:1902600	hydrogen ion transmembrane transport	BP	1.29E-11	1.40E-08	12.95 (2373,12,168,11)
GO:0098655	cation transmembrane transport	BP	3.12E-11	2.98E-08	3.71 (2373,93,234,34)
GO:0006334	nucleosome assembly	BP	6.64E-11	5.63E-08	2.28 (2373,47,906,41)
GO:0065004	protein-DNA complex assembly	BP	1.00E-10	7.64E-08	2.22 (2373,52,906,44)
GO:0050688	regulation of defense response to virus	BP	1.57E-10	1.09E-07	14.46 (2373,19,95,11)
GO:0002504	antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	BP	3.35E-10	2.13E-07	3.53 (2373,32,525,25)
GO:0002495	antigen processing and presentation of peptide antigen via MHC class II	BP	3.35E-10	1.97E-07	3.53 (2373,32,525,25)
GO:0019886	antigen processing and presentation of exogenous peptide antigen via MHC class II	BP	3.35E-10	1.83E-07	3.53 (2373,32,525,25)
GO:0002831	regulation of response to biotic stimulus	BP	8.01E-10	4.08E-07	13.08 (2373,21,95,11)
GO:0046916	cellular transition metal ion homeostasis	BP	9.73E-10	4.64E-07	11.77 (2373,12,168,10)

GO:0006810	transport	BP	4.33E-09	1.94E-06	1.78 (2373,670,173,87)
GO:0051234	establishment of localization	BP	4.69E-09	1.99E-06	1.77 (2373,683,173,88)
GO:0071824	protein-DNA complex subunit organization	BP	1.02E-08	4.11E-06	2.02 (2373,61,906,47)
GO:0034728	nucleosome organization	BP	1.17E-08	4.48E-06	2.06 (2373,56,906,44)
GO:0006812	cation transport	BP	1.54E-08	5.59E-06	2.73 (2373,137,254,40)
GO:0016192	vesicle-mediated transport	BP	2.17E-08	7.55E-06	3.36 (2373,186,114,30)
GO:0051179	localization	BP	2.97E-08	9.87E-06	1.68 (2373,741,173,91)
GO:0007015	actin filament organization	BP	3.31E-08	1.05E-05	7.65 (2373,31,130,13)
GO:0034220	ion transmembrane transport	BP	3.63E-08	1.11E-05	2.80 (2373,134,234,37)
GO:0034723	DNA replication-dependent nucleosome organization	BP	4.33E-08	1.27E-05	2.51 (2373,24,906,23)
GO:0006335	DNA replication-dependent nucleosome assembly	BP	4.33E-08	1.22E-05	2.51 (2373,24,906,23)
GO:0055085	transmembrane transport	BP	5.52E-08	1.50E-05	2.79 (2373,184,171,37)
GO:0007188	adenylate cyclase-modulating G-protein coupled receptor signaling pathway	BP	6.44E-08	1.70E-05	2.61 (2373,27,808,24)
GO:0006333	chromatin assembly or disassembly	BP	6.86E-08	1.74E-05	2.36 (2373,30,906,27)
GO:0034314	Arp2/3 complex-mediated actin nucleation	BP	6.90E-08	1.70E-05	18.54 (2373,6,128,6)
GO:0045010	actin nucleation	BP	6.90E-08	1.65E-05	18.54 (2373,6,128,6)
GO:1902578	single-organism localization	BP	9.42E-08	2.18E-05	1.83 (2373,553,171,73)
GO:0098662	inorganic cation transmembrane transport	BP	1.04E-07	2.34E-05	3.42 (2373,77,234,26)
GO:0007166	cell surface receptor signaling pathway	BP	1.12E-07	2.44E-05	1.34 (2373,451,834,212)
GO:0055076	transition metal ion homeostasis	BP	1.35E-07	2.86E-05	8.83 (2373,16,168,10)
GO:0044765	single-organism transport	BP	1.71E-07	3.53E-05	1.85 (2373,526,171,70)
GO:0007187	G-protein coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger	BP	1.88E-07	3.77E-05	2.16 (2373,36,947,31)
GO:0006754	ATP biosynthetic process	BP	2.24E-07	4.38E-05	12.90 (2373,8,161,7)
GO:0006892	post-Golgi vesicle-mediated transport	BP	2.51E-07	4.79E-05	7.30 (2373,30,130,12)
GO:1902476	chloride transmembrane transport	BP	3.12E-07	5.80E-05	3.04 (2373,18,738,17)

GO:0098661	inorganic anion transmembrane transport	BP	3.12E-07	5.67E-05	3.04 (2373,18,738,17)
GO:0002697	regulation of immune effector process	BP	7.57E-07	1.34E-04	7.85 (2373,35,95,11)
GO:0098660	inorganic ion transmembrane transport	BP	8.14E-07	1.41E-04	2.99 (2373,95,234,28)
GO:0072522	purine-containing compound biosynthetic process	BP	9.11E-07	1.55E-04	7.05 (2373,23,161,11)
GO:0002376	immune system process	BP	9.72E-07	1.61E-04	2.76 (2373,272,95,30)
GO:0007268	synaptic transmission	BP	1.31E-06	2.12E-04	1.62 (2373,141,832,80)
GO:0007215	glutamate receptor signaling pathway	BP	1.33E-06	2.12E-04	2.66 (2373,21,808,19)
GO:0006879	cellular iron ion homeostasis	BP	1.48E-06	2.31E-04	10.99 (2373,9,168,7)
GO:0043900	regulation of multi-organism process	BP	1.58E-06	2.41E-04	6.01 (2373,54,95,13)
GO:0006821	chloride transport	BP	2.01E-06	3.00E-04	2.88 (2373,19,738,17)
GO:0035235	ionotropic glutamate receptor signaling pathway	BP	2.20E-06	3.22E-04	2.96 (2373,14,802,14)
GO:0007214	gamma-aminobutyric acid signaling pathway	BP	2.54E-06	3.65E-04	3.16 (2373,13,750,13)
GO:0002478	antigen processing and presentation of exogenous peptide antigen	BP	2.61E-06	3.69E-04	5.80 (2373,56,95,13)
GO:0019884	antigen processing and presentation of exogenous antigen	BP	2.61E-06	3.63E-04	5.80 (2373,56,95,13)
GO:0009260	ribonucleotide biosynthetic process	BP	3.48E-06	4.75E-04	7.02 (2373,21,161,10)
GO:0009152	purine ribonucleotide biosynthetic process	BP	3.48E-06	4.67E-04	7.02 (2373,21,161,10)
GO:0007154	cell communication	BP	3.65E-06	4.81E-04	1.48 (2373,208,832,108)
GO:0090066	regulation of anatomical structure size	BP	3.81E-06	4.93E-04	5.02 (2373,46,144,14)
GO:0009201	ribonucleoside triphosphate biosynthetic process	BP	4.02E-06	5.11E-04	10.32 (2373,10,161,7)
GO:0009206	purine ribonucleoside triphosphate biosynthetic process	BP	4.02E-06	5.03E-04	10.32 (2373,10,161,7)
GO:0009145	purine nucleoside triphosphate biosynthetic process	BP	4.02E-06	4.95E-04	10.32 (2373,10,161,7)
GO:0048002	antigen processing and presentation of peptide antigen	BP	4.94E-06	5.99E-04	5.50 (2373,59,95,13)
GO:0090382	phagosome maturation	BP	5.32E-06	6.35E-04	9.89 (2373,10,168,7)
GO:0006811	ion transport	BP	5.32E-06	6.25E-04	2.11 (2373,204,254,46)
GO:0006164	purine nucleotide biosynthetic process	BP	6.04E-06	6.98E-04	6.70 (2373,22,161,10)

GO:0006417	regulation of translation	BP	7.11E-06	8.11E-04	2.11 (2373,76,562,38)
GO:0007267	cell-cell signaling	BP	8.08E-06	9.07E-04	1.49 (2373,187,832,98)
GO:0015698	inorganic anion transport	BP	1.03E-05	1.14E-03	2.73 (2373,20,738,17)
GO:0009142	nucleoside triphosphate biosynthetic process	BP	1.16E-05	1.26E-03	9.38 (2373,11,161,7)
GO:0044700	single organism signaling	BP	1.36E-05	1.46E-03	1.48 (2373,191,832,99)
GO:0023052	signaling	BP	1.36E-05	1.44E-03	1.48 (2373,191,832,99)
GO:1901659	glycosyl compound biosynthetic process	BP	1.38E-05	1.44E-03	6.98 (2373,19,161,9)
GO:0042455	ribonucleoside biosynthetic process	BP	1.38E-05	1.42E-03	6.98 (2373,19,161,9)
GO:0042451	purine nucleoside biosynthetic process	BP	1.38E-05	1.40E-03	6.98 (2373,19,161,9)
GO:0046129	purine ribonucleoside biosynthetic process	BP	1.38E-05	1.39E-03	6.98 (2373,19,161,9)
GO:0009163	nucleoside biosynthetic process	BP	1.38E-05	1.37E-03	6.98 (2373,19,161,9)
GO:0019882	antigen processing and presentation	BP	1.39E-05	1.36E-03	5.07 (2373,64,95,13)
GO:0035338	long-chain fatty-acyl-CoA biosynthetic process	BP	1.39E-05	1.34E-03	79.10 (2373,5,18,3)
GO:0046949	fatty-acyl-CoA biosynthetic process	BP	1.39E-05	1.33E-03	79.10 (2373,5,18,3)
GO:0006875	cellular metal ion homeostasis	BP	1.45E-05	1.37E-03	3.22 (2373,58,254,20)
GO:0015682	ferric iron transport	BP	1.48E-05	1.38E-03	8.99 (2373,11,168,7)
GO:0033572	transferrin transport	BP	1.48E-05	1.36E-03	8.99 (2373,11,168,7)
GO:0072512	trivalent inorganic cation transport	BP	1.48E-05	1.35E-03	8.99 (2373,11,168,7)
GO:0032535	regulation of cellular component size	BP	1.49E-05	1.34E-03	5.83 (2373,35,128,11)
GO:0031124	mRNA 3'-end processing	BP	1.52E-05	1.35E-03	2.75 (2373,31,556,20)
GO:0007165	signal transduction	BP	1.66E-05	1.46E-03	1.20 (2373,768,819,318)
GO:0046390	ribose phosphate biosynthetic process	BP	1.69E-05	1.47E-03	6.14 (2373,24,161,10)
GO:0006446	regulation of translational initiation	BP	1.72E-05	1.48E-03	3.18 (2373,20,560,15)
GO:2000192	negative regulation of fatty acid transport	BP	2.29E-05	1.94E-03	42.37 (2373,3,56,3)
GO:0006730	one-carbon metabolic process	BP	2.43E-05	2.04E-03	40.91 (2373,3,58,3)

GO:0030003	cellular cation homeostasis	BP	2.69E-05	2.23E-03	3.64 (2373,66,168,17)
GO:1900371	regulation of purine nucleotide biosynthetic process	BP	2.72E-05	2.23E-03	2.58 (2373,18,818,16)
GO:0030808	regulation of nucleotide biosynthetic process	BP	2.72E-05	2.21E-03	2.58 (2373,18,818,16)
GO:0030802	regulation of cyclic nucleotide biosynthetic process	BP	2.72E-05	2.18E-03	2.58 (2373,18,818,16)
GO:0009127	purine nucleoside monophosphate biosynthetic process	BP	2.82E-05	2.24E-03	8.60 (2373,12,161,7)
GO:0009168	purine ribonucleoside monophosphate biosynthetic process	BP	2.82E-05	2.22E-03	8.60 (2373,12,161,7)
GO:0031047	gene silencing by RNA	BP	2.89E-05	2.25E-03	4.72 (2373,14,359,10)
GO:0035574	histone H4-K20 demethylation	BP	2.97E-05	2.29E-03	2.62 (2373,13,906,13)
GO:0071616	acyl-CoA biosynthetic process	BP	3.20E-05	2.44E-03	65.92 (2373,6,18,3)
GO:0035337	fatty-acyl-CoA metabolic process	BP	3.20E-05	2.42E-03	65.92 (2373,6,18,3)
GO:0035336	long-chain fatty-acyl-CoA metabolic process	BP	3.20E-05	2.39E-03	65.92 (2373,6,18,3)
GO:0035384	thioester biosynthetic process	BP	3.20E-05	2.37E-03	65.92 (2373,6,18,3)
GO:0030833	regulation of actin filament polymerization	BP	3.47E-05	2.55E-03	6.67 (2373,25,128,9)
GO:0045839	negative regulation of mitosis	BP	3.64E-05	2.65E-03	6.01 (2373,11,287,8)
GO:0006873	cellular ion homeostasis	BP	3.74E-05	2.70E-03	2.60 (2373,68,336,25)
GO:0006641	triglyceride metabolic process	BP	3.85E-05	2.75E-03	32.96 (2373,16,18,4)
GO:0009416	response to light stimulus	BP	3.90E-05	2.75E-03	2.23 (2373,62,515,30)
GO:0000041	transition metal ion transport	BP	4.05E-05	2.84E-03	7.06 (2373,16,168,8)
GO:0007270	neuron-neuron synaptic transmission	BP	4.12E-05	2.86E-03	2.30 (2373,27,802,21)
GO:0097035	regulation of membrane lipid distribution	BP	4.29E-05	2.95E-03	11.56 (2373,6,171,5)
GO:0051983	regulation of chromosome segregation	BP	4.46E-05	3.04E-03	4.31 (2373,13,424,10)
GO:0032273	positive regulation of protein polymerization	BP	4.69E-05	3.17E-03	8.65 (2373,15,128,7)
GO:0030838	positive regulation of actin filament polymerization	BP	4.69E-05	3.14E-03	8.65 (2373,15,128,7)
GO:0044272	sulfur compound biosynthetic process	BP	4.94E-05	3.28E-03	2.93 (2373,20,607,15)
GO:0000289	nuclear-transcribed mRNA poly(A) tail shortening	BP	5.07E-05	3.33E-03	3.41 (2373,15,556,12)
GO:0006898	receptor-mediated endocytosis	BP	5.07E-05	3.31E-03	7.79 (2373,28,87,8)

GO:0006638	neutral lipid metabolic process	BP	5.11E-05	3.30E-03	31.02 (2373,17,18,4)
GO:0006639	acylglycerol metabolic process	BP	5.11E-05	3.28E-03	31.02 (2373,17,18,4)
GO:0015985	energy coupled proton transport, down electrochemical gradient	BP	5.21E-05	3.31E-03	14.74 (2373,4,161,4)
GO:0015986	ATP synthesis coupled proton transport	BP	5.21E-05	3.28E-03	14.74 (2373,4,161,4)
GO:0098656	anion transmembrane transport	BP	5.42E-05	3.39E-03	2.44 (2373,25,738,19)
GO:0030814	regulation of cAMP metabolic process	BP	5.51E-05	3.42E-03	2.59 (2373,17,808,15)
GO:0006942	regulation of striated muscle contraction	BP	5.60E-05	3.45E-03	5.56 (2373,16,240,9)
GO:0071702	organic substance transport	BP	5.92E-05	3.61E-03	1.81 (2373,395,173,52)
GO:0065008	regulation of biological quality	BP	5.95E-05	3.61E-03	1.57 (2373,469,254,79)
GO:0015672	monovalent inorganic cation transport	BP	6.15E-05	3.69E-03	3.78 (2373,56,168,15)
GO:0031123	RNA 3'-end processing	BP	6.24E-05	3.72E-03	2.59 (2373,33,556,20)
GO:0019432	triglyceride biosynthetic process	BP	6.35E-05	3.76E-03	56.50 (2373,7,18,3)
GO:0046460	neutral lipid biosynthetic process	BP	6.35E-05	3.73E-03	56.50 (2373,7,18,3)
GO:0046463	acylglycerol biosynthetic process	BP	6.35E-05	3.70E-03	56.50 (2373,7,18,3)
GO:0034204	lipid translocation	BP	6.97E-05	4.03E-03	13.88 (2373,4,171,4)
GO:0045332	phospholipid translocation	BP	6.97E-05	4.00E-03	13.88 (2373,4,171,4)
GO:0030799	regulation of cyclic nucleotide metabolic process	BP	7.02E-05	4.00E-03	2.37 (2373,22,818,18)
GO:0044711	single-organism biosynthetic process	BP	7.37E-05	4.17E-03	2.70 (2373,120,176,24)
GO:0030832	regulation of actin filament length	BP	7.55E-05	4.24E-03	6.18 (2373,27,128,9)
GO:0008064	regulation of actin polymerization or depolymerization	BP	7.55E-05	4.21E-03	6.18 (2373,27,128,9)
GO:0055072	iron ion homeostasis	BP	7.93E-05	4.39E-03	7.61 (2373,13,168,7)
GO:0006952	defense response	BP	8.03E-05	4.41E-03	3.81 (2373,170,55,15)
GO:0010862	positive regulation of pathway-restricted SMAD protein phosphorylation	BP	8.38E-05	4.57E-03	10.25 (2373,6,193,5)
GO:0071173	spindle assembly checkpoint	BP	8.96E-05	4.85E-03	7.09 (2373,7,287,6)
GO:0007094	mitotic spindle assembly checkpoint	BP	8.96E-05	4.82E-03	7.09 (2373,7,287,6)
GO:0002431	Fc receptor mediated stimulatory signaling pathway	BP	9.09E-05	4.85E-03	11.98 (2373,36,33,6)

GO:0002433	immune response-regulating cell surface receptor signaling pathway involved in phagocytosis	BP	9.09E-05	4.82E-03	11.98 (2373,36,33,6)
GO:0038096	Fc-gamma receptor signaling pathway involved in phagocytosis	BP	9.09E-05	4.79E-03	11.98 (2373,36,33,6)
GO:0038094	Fc-gamma receptor signaling pathway	BP	9.09E-05	4.75E-03	11.98 (2373,36,33,6)
GO:1901135	carbohydrate derivative metabolic process	BP	9.70E-05	5.04E-03	2.00 (2373,263,176,39)
GO:0001895	retina homeostasis	BP	9.81E-05	5.06E-03	98.87 (2373,2,24,2)
GO:0055065	metal ion homeostasis	BP	9.93E-05	5.09E-03	2.79 (2373,67,267,21)
GO:0072583	clathrin-mediated endocytosis	BP	1.01E-04	5.15E-03	18.18 (2373,6,87,4)
GO:0006869	lipid transport	BP	1.02E-04	5.16E-03	5.68 (2373,22,171,9)
GO:0032271	regulation of protein polymerization	BP	1.04E-04	5.23E-03	5.96 (2373,28,128,9)
GO:0051784	negative regulation of nuclear division	BP	1.05E-04	5.22E-03	5.51 (2373,12,287,8)
GO:0048193	Golgi vesicle transport	BP	1.08E-04	5.34E-03	4.16 (2373,57,130,13)
GO:0009124	nucleoside monophosphate biosynthetic process	BP	1.11E-04	5.47E-03	7.37 (2373,14,161,7)
GO:0009156	ribonucleoside monophosphate biosynthetic process	BP	1.11E-04	5.43E-03	7.37 (2373,14,161,7)
GO:0046034	ATP metabolic process	BP	1.11E-04	5.40E-03	3.18 (2373,80,168,18)
GO:0045653	negative regulation of megakaryocyte differentiation	BP	1.13E-04	5.47E-03	2.44 (2373,15,906,14)
GO:0030001	metal ion transport	BP	1.13E-04	5.44E-03	2.38 (2373,110,254,28)
GO:0032925	regulation of activin receptor signaling pathway	BP	1.17E-04	5.57E-03	38.48 (2373,5,37,3)
GO:0050801	ion homeostasis	BP	1.21E-04	5.72E-03	2.35 (2373,78,350,27)
GO:0051716	cellular response to stimulus	BP	1.25E-04	5.88E-03	1.16 (2373,907,834,369)
GO:0072521	purine-containing compound metabolic process	BP	1.33E-04	6.23E-03	2.16 (2373,222,168,34)
GO:0032891	negative regulation of organic acid transport	BP	1.34E-04	6.24E-03	31.78 (2373,4,56,3)
GO:2000191	regulation of fatty acid transport	BP	1.34E-04	6.20E-03	31.78 (2373,4,56,3)
GO:0055080	cation homeostasis	BP	1.38E-04	6.32E-03	2.44 (2373,75,324,25)
GO:0030817	regulation of cAMP biosynthetic process	BP	1.38E-04	6.31E-03	2.57 (2373,16,808,14)
GO:0006826	iron ion transport	BP	1.46E-04	6.65E-03	7.06 (2373,14,168,7)

GO:0006629	lipid metabolic process	BP	1.53E-04	6.89E-03	8.39 (2373,110,18,7)
GO:0010959	regulation of metal ion transport	BP	1.55E-04	6.95E-03	3.11 (2373,51,254,17)
GO:0032970	regulation of actin filament-based process	BP	1.57E-04	7.00E-03	3.18 (2373,64,198,17)
GO:0071174	mitotic spindle checkpoint	BP	1.58E-04	6.99E-03	5.09 (2373,8,408,7)
GO:0031577	spindle checkpoint	BP	1.58E-04	6.95E-03	5.09 (2373,8,408,7)
GO:0031279	regulation of cyclase activity	BP	1.58E-04	6.93E-03	2.54 (2373,16,817,14)
GO:0051339	regulation of lyase activity	BP	1.58E-04	6.89E-03	2.54 (2373,16,817,14)
GO:0043085	positive regulation of catalytic activity	BP	1.63E-04	7.08E-03	2.32 (2373,258,107,27)
GO:1903510	mucopolysaccharide metabolic process	BP	1.69E-04	7.29E-03	5.58 (2373,9,331,7)
GO:0045087	innate immune response	BP	1.76E-04	7.56E-03	4.06 (2373,138,55,13)
GO:0030240	skeletal muscle thin filament assembly	BP	1.80E-04	7.67E-03	103.17 (2373,2,23,2)
GO:0032312	regulation of ARF GTPase activity	BP	1.80E-04	7.64E-03	15.82 (2373,6,100,4)
GO:0055117	regulation of cardiac muscle contraction	BP	1.82E-04	7.69E-03	6.29 (2373,11,240,7)
GO:0051345	positive regulation of hydrolase activity	BP	1.88E-04	7.87E-03	2.99 (2373,141,107,19)
GO:0016344	meiotic chromosome movement towards spindle pole	BP	1.88E-04	7.83E-03	71.91 (2373,2,33,2)
GO:0033206	meiotic cytokinesis	BP	1.88E-04	7.78E-03	71.91 (2373,2,33,2)
GO:0051305	chromosome movement towards spindle pole	BP	1.88E-04	7.74E-03	71.91 (2373,2,33,2)
GO:0051653	spindle localization	BP	1.88E-04	7.70E-03	71.91 (2373,2,33,2)
GO:0016458	gene silencing	BP	1.98E-04	8.06E-03	3.83 (2373,19,359,11)
GO:0050994	regulation of lipid catabolic process	BP	1.99E-04	8.08E-03	18.83 (2373,9,56,4)
GO:0008610	lipid biosynthetic process	BP	2.00E-04	8.09E-03	14.65 (2373,45,18,5)
GO:0006631	fatty acid metabolic process	BP	2.01E-04	8.09E-03	22.93 (2373,23,18,4)
GO:0055082	cellular chemical homeostasis	BP	2.06E-04	8.25E-03	3.16 (2373,76,168,17)
GO:0007264	small GTPase mediated signal transduction	BP	2.21E-04	8.79E-03	2.96 (2373,127,120,19)
GO:0009126	purine nucleoside monophosphate metabolic process	BP	2.22E-04	8.78E-03	3.03 (2373,84,168,18)
GO:0009167	purine ribonucleoside monophosphate metabolic process	BP	2.22E-04	8.73E-03	3.03 (2373,84,168,18)

GO:0042278	purine nucleoside metabolic process	BP	2.34E-04	9.16E-03	2.17 (2373,208,168,32)
GO:0046128	purine ribonucleoside metabolic process	BP	2.34E-04	9.11E-03	2.17 (2373,208,168,32)
GO:0009119	ribonucleoside metabolic process	BP	2.34E-04	9.06E-03	2.17 (2373,208,168,32)
GO:0006890	retrograde vesicle-mediated transport, Golgi to ER	BP	2.55E-04	9.85E-03	4.69 (2373,11,368,8)
GO:1901657	glycosyl compound metabolic process	BP	2.55E-04	9.80E-03	2.16 (2373,209,168,32)
GO:0009116	nucleoside metabolic process	BP	2.55E-04	9.75E-03	2.16 (2373,209,168,32)
GO:0045652	regulation of megakaryocyte differentiation	BP	2.56E-04	9.71E-03	2.31 (2373,17,906,15)
GO:0051782	negative regulation of cell division	BP	2.65E-04	1.00E-02	5.09 (2373,13,287,8)
GO:0019752	carboxylic acid metabolic process	BP	2.72E-04	1.02E-02	9.89 (2373,80,18,6)
GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	BP	2.76E-04	1.03E-02	2.14 (2373,148,240,32)
GO:0045932	negative regulation of muscle contraction	BP	2.85E-04	1.06E-02	12.41 (2373,5,153,4)
GO:0032924	activin receptor signaling pathway	BP	2.91E-04	1.08E-02	32.07 (2373,6,37,3)
GO:0038127	ERBB signaling pathway	BP	2.95E-04	1.09E-02	4.13 (2373,65,106,12)
GO:0007173	epidermal growth factor receptor signaling pathway	BP	2.95E-04	1.08E-02	4.13 (2373,65,106,12)
GO:0009123	nucleoside monophosphate metabolic process	BP	3.02E-04	1.10E-02	2.96 (2373,86,168,18)
GO:0009161	ribonucleoside monophosphate metabolic process	BP	3.02E-04	1.10E-02	2.96 (2373,86,168,18)
GO:0006171	cAMP biosynthetic process	BP	3.06E-04	1.11E-02	56.50 (2373,2,42,2)
GO:0090257	regulation of muscle system process	BP	3.10E-04	1.12E-02	3.74 (2373,30,254,12)
GO:0060359	response to ammonium ion	BP	3.28E-04	1.17E-02	2.37 (2373,18,833,15)
GO:1901293	nucleoside phosphate biosynthetic process	BP	3.40E-04	1.21E-02	4.61 (2373,32,161,10)
GO:0009165	nucleotide biosynthetic process	BP	3.40E-04	1.21E-02	4.61 (2373,32,161,10)
GO:0090503	RNA phosphodiester bond hydrolysis, exonucleolytic	BP	3.46E-04	1.22E-02	5.67 (2373,7,359,6)
GO:0006986	response to unfolded protein	BP	3.52E-04	1.24E-02	1.98 (2373,28,944,22)
GO:0045761	regulation of adenylate cyclase activity	BP	3.57E-04	1.25E-02	2.55 (2373,15,807,13)
GO:0002682	regulation of immune system process	BP	3.68E-04	1.28E-02	2.74 (2373,199,87,20)
GO:2000816	negative regulation of mitotic sister chromatid separation	BP	3.91E-04	1.36E-02	6.20 (2373,8,287,6)

GO:0051985	negative regulation of chromosome segregation	BP	3.91E-04	1.35E-02	6.20 (2373,8,287,6)
GO:1902100	negative regulation of metaphase/anaphase transition of cell cycle	BP	3.91E-04	1.35E-02	6.20 (2373,8,287,6)
GO:0045841	negative regulation of mitotic metaphase/anaphase transition	BP	3.91E-04	1.34E-02	6.20 (2373,8,287,6)
GO:0033048	negative regulation of mitotic sister chromatid segregation	BP	3.91E-04	1.33E-02	6.20 (2373,8,287,6)
GO:0033046	negative regulation of sister chromatid segregation	BP	3.91E-04	1.33E-02	6.20 (2373,8,287,6)
GO:0002252	immune effector process	BP	4.05E-04	1.37E-02	3.23 (2373,56,197,15)
GO:0010882	regulation of cardiac muscle contraction by calcium ion signaling	BP	4.08E-04	1.37E-02	6.59 (2373,9,240,6)
GO:0045017	glycerolipid biosynthetic process	BP	4.09E-04	1.37E-02	19.53 (2373,27,18,4)
GO:0048205	COPI coating of Golgi vesicle	BP	4.15E-04	1.38E-02	5.54 (2373,7,367,6)
GO:0048200	Golgi transport vesicle coating	BP	4.15E-04	1.38E-02	5.54 (2373,7,367,6)
GO:0032369	negative regulation of lipid transport	BP	4.20E-04	1.39E-02	25.42 (2373,5,56,3)
GO:0010608	posttranscriptional regulation of gene expression	BP	4.21E-04	1.39E-02	1.74 (2373,109,562,45)
GO:0032377	regulation of intracellular lipid transport	BP	4.21E-04	1.38E-02	2,373.00 (2373,1,1,1)
GO:0032380	regulation of intracellular sterol transport	BP	4.21E-04	1.37E-02	2,373.00 (2373,1,1,1)
GO:0032383	regulation of intracellular cholesterol transport	BP	4.21E-04	1.37E-02	2,373.00 (2373,1,1,1)
GO:0006637	acyl-CoA metabolic process	BP	4.26E-04	1.38E-02	35.95 (2373,11,18,3)
GO:0035383	thioester metabolic process	BP	4.26E-04	1.37E-02	35.95 (2373,11,18,3)
GO:0009108	coenzyme biosynthetic process	BP	4.26E-04	1.37E-02	35.95 (2373,11,18,3)
GO:0090501	RNA phosphodiester bond hydrolysis	BP	4.35E-04	1.39E-02	3.75 (2373,12,474,9)
GO:0034724	DNA replication-independent nucleosome organization	BP	4.44E-04	1.41E-02	2.21 (2373,19,906,16)
GO:0006336	DNA replication-independent nucleosome assembly	BP	4.44E-04	1.41E-02	2.21 (2373,19,906,16)
GO:0051726	regulation of cell cycle	BP	4.59E-04	1.45E-02	1.64 (2373,183,459,58)
GO:0051701	interaction with host	BP	4.74E-04	1.49E-02	6.18 (2373,16,168,7)
GO:0006200	ATP catabolic process	BP	4.74E-04	1.48E-02	32.51 (2373,73,3,3)
GO:0051412	response to corticosterone	BP	4.78E-04	1.49E-02	8.69 (2373,4,273,4)
GO:0009259	ribonucleotide metabolic process	BP	4.81E-04	1.49E-02	2.09 (2373,216,168,32)

GO:0009150	purine ribonucleotide metabolic process	BP	4.81E-04	1.49E-02	2.09 (2373,216,168,32)
GO:0071436	sodium ion export	BP	4.89E-04	1.50E-02	15.61 (2373,3,152,3)
GO:0036376	sodium ion export from cell	BP	4.89E-04	1.50E-02	15.61 (2373,3,152,3)
GO:1903416	response to glycoside	BP	4.89E-04	1.49E-02	15.61 (2373,3,152,3)
GO:0009125	nucleoside monophosphate catabolic process	BP	4.93E-04	1.50E-02	32.07 (2373,74,3,3)
GO:0009128	purine nucleoside monophosphate catabolic process	BP	4.93E-04	1.49E-02	32.07 (2373,74,3,3)
GO:0009169	purine ribonucleoside monophosphate catabolic process	BP	4.93E-04	1.49E-02	32.07 (2373,74,3,3)
GO:0009158	ribonucleoside monophosphate catabolic process	BP	4.93E-04	1.48E-02	32.07 (2373,74,3,3)
GO:0045988	negative regulation of striated muscle contraction	BP	4.94E-04	1.48E-02	15.51 (2373,3,153,3)
GO:0006023	aminoglycan biosynthetic process	BP	4.94E-04	1.47E-02	2.31 (2373,14,952,13)
GO:0006024	glycosaminoglycan biosynthetic process	BP	4.94E-04	1.47E-02	2.31 (2373,14,952,13)
GO:0035249	synaptic transmission, glutamatergic	BP	5.06E-04	1.50E-02	2.44 (2373,17,802,14)
GO:0015031	protein transport	BP	5.13E-04	1.51E-02	1.85 (2373,306,168,40)
GO:0046907	intracellular transport	BP	5.19E-04	1.52E-02	1.87 (2373,310,160,39)
GO:0006937	regulation of muscle contraction	BP	5.20E-04	1.52E-02	3.81 (2373,27,254,11)
GO:0030203	glycosaminoglycan metabolic process	BP	5.33E-04	1.55E-02	2.20 (2373,17,952,15)
GO:0006022	aminoglycan metabolic process	BP	5.33E-04	1.55E-02	2.20 (2373,17,952,15)
GO:0050896	response to stimulus	BP	5.56E-04	1.61E-02	1.13 (2373,1069,834,423)
GO:0009314	response to radiation	BP	5.69E-04	1.64E-02	1.65 (2373,77,822,44)
GO:0051928	positive regulation of calcium ion transport	BP	5.72E-04	1.64E-02	5.45 (2373,12,254,7)
GO:0070588	calcium ion transmembrane transport	BP	5.74E-04	1.64E-02	3.58 (2373,34,234,12)
GO:0042776	mitochondrial ATP synthesis coupled proton transport	BP	5.76E-04	1.64E-02	14.83 (2373,3,160,3)
GO:0044255	cellular lipid metabolic process	BP	5.78E-04	1.64E-02	8.69 (2373,91,18,6)
GO:1901533	negative regulation of hematopoietic progenitor cell differentiation	BP	5.83E-04	1.65E-02	2.29 (2373,16,906,14)
GO:0051188	cofactor biosynthetic process	BP	5.98E-04	1.68E-02	32.96 (2373,12,18,3)

GO:0006897	endocytosis	BP	6.00E-04	1.68E-02	4.56 (2373,56,93,10)
GO:0006996	organelle organization	BP	6.09E-04	1.70E-02	1.67 (2373,362,204,52)
GO:0007167	enzyme linked receptor protein signaling pathway	BP	6.20E-04	1.73E-02	1.92 (2373,196,240,38)
GO:0055086	nucleobase-containing small molecule metabolic process	BP	6.28E-04	1.74E-02	1.94 (2373,247,168,34)
GO:0006163	purine nucleotide metabolic process	BP	6.31E-04	1.74E-02	2.06 (2373,219,168,32)
GO:0019693	ribose phosphate metabolic process	BP	6.31E-04	1.74E-02	2.06 (2373,219,168,32)
GO:0002429	immune response-activating cell surface receptor signaling pathway	BP	6.42E-04	1.76E-02	8.63 (2373,50,33,6)
GO:2001258	negative regulation of cation channel activity	BP	6.44E-04	1.76E-02	3.11 (2373,8,763,8)
GO:0032956	regulation of actin cytoskeleton organization	BP	6.71E-04	1.83E-02	3.77 (2373,59,128,12)
GO:0010965	regulation of mitotic sister chromatid separation	BP	7.02E-04	1.91E-02	4.07 (2373,11,424,8)
GO:0033045	regulation of sister chromatid segregation	BP	7.02E-04	1.90E-02	4.07 (2373,11,424,8)
GO:0033047	regulation of mitotic sister chromatid segregation	BP	7.02E-04	1.89E-02	4.07 (2373,11,424,8)
GO:0015936	coenzyme A metabolic process	BP	7.10E-04	1.91E-02	74.16 (2373,4,16,2)
GO:0045744	negative regulation of G-protein coupled receptor protein signaling pathway	BP	7.25E-04	1.94E-02	14.92 (2373,12,53,4)
GO:1901564	organonitrogen compound metabolic process	BP	7.29E-04	1.95E-02	1.79 (2373,315,177,42)
GO:0045184	establishment of protein localization	BP	7.55E-04	2.01E-02	1.80 (2373,316,171,41)
GO:0050996	positive regulation of lipid catabolic process	BP	7.56E-04	2.00E-02	96.86 (2373,7,7,2)
GO:0006816	calcium ion transport	BP	7.61E-04	2.01E-02	2.98 (2373,47,254,15)
GO:0043436	oxoacid metabolic process	BP	7.68E-04	2.02E-02	8.24 (2373,96,18,6)
GO:0006101	citrate metabolic process	BP	8.04E-04	2.11E-02	69.79 (2373,4,17,2)
GO:0006955	immune response	BP	8.28E-04	2.17E-02	3.51 (2373,160,55,13)
GO:0009725	response to hormone	BP	8.28E-04	2.16E-02	2.00 (2373,163,240,33)
GO:0044699	single-organism process	BP	8.34E-04	2.16E-02	1.15 (2373,1757,199,170)
GO:0042760	very long-chain fatty acid catabolic process	BP	8.43E-04	2.18E-02	1,186.50 (2373,1,2,1)
GO:0006082	organic acid metabolic process	BP	8.62E-04	2.22E-02	8.07 (2373,98,18,6)

GO:0016050	vesicle organization	BP	8.82E-04	2.27E-02	2.63 (2373,34,424,16)
GO:0043271	negative regulation of ion transport	BP	8.84E-04	2.26E-02	2.24 (2373,21,807,16)
GO:0060393	regulation of pathway-restricted SMAD protein phosphorylation	BP	8.90E-04	2.27E-02	7.68 (2373,8,193,5)
GO:0045822	negative regulation of heart contraction	BP	9.69E-04	2.47E-02	10.27 (2373,6,154,4)
GO:1903523	negative regulation of blood circulation	BP	9.69E-04	2.46E-02	10.27 (2373,6,154,4)
GO:0032200	telomere organization	BP	9.80E-04	2.48E-02	2.18 (2373,18,906,15)
GO:0000723	telomere maintenance	BP	9.80E-04	2.47E-02	2.18 (2373,18,906,15)
GO:1901532	regulation of hematopoietic progenitor cell differentiation	BP	9.80E-04	2.46E-02	2.18 (2373,18,906,15)
GO:0031998	regulation of fatty acid beta-oxidation	BP	9.97E-04	2.50E-02	21.19 (2373,6,56,3)
GO:0016820	hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances	MF	1.91E-27	3.15E-24	13.88 (2373,24,171,24)
GO:0042626	ATPase activity, coupled to transmembrane movement of substances	MF	1.91E-27	1.58E-24	13.88 (2373,24,171,24)
GO:0043492	ATPase activity, coupled to movement of substances	MF	4.67E-26	2.57E-23	13.32 (2373,25,171,24)
GO:0015399	primary active transmembrane transporter activity	MF	4.67E-26	1.93E-23	13.32 (2373,25,171,24)
GO:0015405	P-P-bond-hydrolysis-driven transmembrane transporter activity	MF	4.67E-26	1.54E-23	13.32 (2373,25,171,24)
GO:0042625	ATPase activity, coupled to transmembrane movement of ions	MF	6.55E-24	1.81E-21	13.88 (2373,21,171,21)
GO:0019829	cation-transporting ATPase activity	MF	6.55E-24	1.55E-21	13.88 (2373,21,171,21)
GO:0022804	active transmembrane transporter activity	MF	8.40E-16	1.73E-13	7.75 (2373,43,171,24)
GO:0008135	translation factor activity, nucleic acid binding	MF	6.50E-13	1.19E-10	3.68 (2373,28,598,26)
GO:0005230	extracellular ligand-gated ion channel activity	MF	7.59E-13	1.26E-10	2.96 (2373,28,802,28)
GO:0015662	ATPase activity, coupled to transmembrane movement of ions, phosphorylative mechanism	MF	9.56E-12	1.44E-09	14.30 (2373,10,166,10)
GO:0042623	ATPase activity, coupled	MF	1.17E-11	1.61E-09	5.26 (2373,66,171,25)
GO:0003743	translation initiation factor activity	MF	4.64E-11	5.89E-09	4.03 (2373,21,561,20)
GO:0016887	ATPase activity	MF	9.37E-11	1.11E-08	4.37 (2373,89,171,28)
GO:0015078	hydrogen ion transmembrane transporter activity	MF	1.48E-10	1.64E-08	12.84 (2373,11,168,10)
GO:0005524	ATP binding	MF	3.00E-10	3.10E-08	4.19 (2373,308,46,25)

GO:0032559	adenyl ribonucleotide binding	MF	5.69E-10	5.53E-08	4.07 (2373,317,46,25)
GO:0030554	adenyl nucleotide binding	MF	6.70E-10	6.16E-08	4.04 (2373,319,46,25)
GO:0005215	transporter activity	MF	3.21E-09	2.79E-07	2.84 (2373,200,171,41)
GO:0004888	transmembrane signaling receptor activity	MF	3.45E-09	2.85E-07	1.92 (2373,92,832,62)
GO:0046982	protein heterodimerization activity	MF	5.24E-09	4.12E-07	1.65 (2373,150,912,95)
GO:0008324	cation transmembrane transporter activity	MF	5.70E-09	4.28E-07	3.22 (2373,104,234,33)
GO:0005254	chloride channel activity	MF	2.36E-08	1.70E-06	3.22 (2373,17,738,17)
GO:0036442	hydrogen-exporting ATPase activity	MF	2.81E-08	1.94E-06	14.12 (2373,7,168,7)
GO:0022892	substrate-specific transporter activity	MF	7.68E-08	5.08E-06	2.81 (2373,178,171,36)
GO:0001882	nucleoside binding	MF	7.79E-08	4.96E-06	1.96 (2373,452,171,64)
GO:0005231	excitatory extracellular ligand-gated ion channel activity	MF	9.40E-08	5.76E-06	2.96 (2373,17,802,17)
GO:0032550	purine ribonucleoside binding	MF	1.65E-07	9.75E-06	1.95 (2373,449,171,63)
GO:0001883	purine nucleoside binding	MF	1.65E-07	9.42E-06	1.95 (2373,449,171,63)
GO:0035639	purine ribonucleoside triphosphate binding	MF	1.65E-07	9.10E-06	1.95 (2373,449,171,63)
GO:0032549	ribonucleoside binding	MF	1.93E-07	1.03E-05	1.94 (2373,451,171,63)
GO:0015103	inorganic anion transmembrane transporter activity	MF	3.12E-07	1.61E-05	3.04 (2373,18,738,17)
GO:0015108	chloride transmembrane transporter activity	MF	3.12E-07	1.56E-05	3.04 (2373,18,738,17)
GO:0032555	purine ribonucleotide binding	MF	3.39E-07	1.65E-05	1.91 (2373,457,171,63)
GO:0032553	ribonucleotide binding	MF	4.01E-07	1.89E-05	1.90 (2373,459,171,63)
GO:0017076	purine nucleotide binding	MF	4.01E-07	1.84E-05	1.90 (2373,459,171,63)
GO:0044769	ATPase activity, coupled to transmembrane movement of ions, rotational mechanism	MF	4.03E-07	1.80E-05	14.12 (2373,6,168,6)
GO:0046961	proton-transporting ATPase activity, rotational mechanism	MF	4.03E-07	1.75E-05	14.12 (2373,6,168,6)
GO:0016817	hydrolase activity, acting on acid anhydrides	MF	4.11E-07	1.74E-05	2.41 (2373,230,171,40)
GO:0016818	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	MF	4.11E-07	1.70E-05	2.41 (2373,230,171,40)
GO:0016462	pyrophosphatase activity	MF	4.11E-07	1.66E-05	2.41 (2373,230,171,40)
GO:0017111	nucleoside-triphosphatase activity	MF	4.16E-07	1.64E-05	2.48 (2373,224,171,40)

GO:0097367	carbohydrate derivative binding	MF	5.45E-07	2.09E-05	1.86 (2373,484,171,65)
GO:0008066	glutamate receptor activity	MF	8.47E-07	3.18E-05	2.94 (2373,15,808,15)
GO:0004871	signal transducer activity	MF	1.88E-06	6.90E-05	1.51 (2373,197,819,103)
GO:0060089	molecular transducer activity	MF	1.88E-06	6.75E-05	1.51 (2373,197,819,103)
GO:0004693	cyclin-dependent protein serine/threonine kinase activity	MF	2.12E-06	7.44E-05	7.88 (2373,7,301,7)
GO:0097472	cyclin-dependent protein kinase activity	MF	2.12E-06	7.28E-05	7.88 (2373,7,301,7)
GO:0022890	inorganic cation transmembrane transporter activity	MF	2.35E-06	7.93E-05	3.00 (2373,88,234,26)
GO:0031683	G-protein beta/gamma-subunit complex binding	MF	2.83E-06	9.36E-05	3.13 (2373,13,757,13)
GO:0022857	transmembrane transporter activity	MF	2.97E-06	9.63E-05	2.41 (2373,156,234,37)
GO:0005388	calcium-transporting ATPase activity	MF	3.22E-06	1.02E-04	15.11 (2373,5,157,5)
GO:0004930	G-protein coupled receptor activity	MF	3.46E-06	1.08E-04	2.00 (2373,40,947,32)
GO:0008026	ATP-dependent helicase activity	MF	5.35E-06	1.64E-04	3.37 (2373,19,556,15)
GO:0070035	purine NTP-dependent helicase activity	MF	5.35E-06	1.61E-04	3.37 (2373,19,556,15)
GO:0048185	activin binding	MF	6.64E-06	1.96E-04	64.14 (2373,3,37,3)
GO:0017002	activin-activated receptor activity	MF	6.64E-06	1.93E-04	64.14 (2373,3,37,3)
GO:0005253	anion channel activity	MF	1.01E-05	2.87E-04	2.60 (2373,21,783,18)
GO:0038023	signaling receptor activity	MF	1.02E-05	2.85E-04	1.66 (2373,111,809,63)
GO:0016917	GABA receptor activity	MF	1.22E-05	3.37E-04	3.34 (2373,11,710,11)
GO:0015075	ion transmembrane transporter activity	MF	1.51E-05	4.09E-04	2.36 (2373,146,234,34)
GO:0004386	helicase activity	MF	1.52E-05	4.06E-04	2.75 (2373,31,556,20)
GO:0043168	anion binding	MF	1.63E-05	4.28E-04	1.69 (2373,551,171,67)
GO:0005319	lipid transporter activity	MF	1.67E-05	4.33E-04	10.41 (2373,8,171,6)
GO:0022834	ligand-gated channel activity	MF	1.89E-05	4.80E-04	1.86 (2373,44,985,34)
GO:0015276	ligand-gated ion channel activity	MF	1.89E-05	4.73E-04	1.86 (2373,44,985,34)

GO:0032403	protein complex binding	MF	2.11E-05	5.21E-04	4.82 (2373,164,39,13)
GO:0098748	endocytic adaptor activity	MF	2.65E-05	6.45E-04	21.82 (2373,5,87,4)
GO:0035615	clathrin adaptor activity	MF	2.65E-05	6.35E-04	21.82 (2373,5,87,4)
GO:0035575	histone demethylase activity (H4-K20 specific)	MF	2.97E-05	7.02E-04	2.62 (2373,13,906,13)
GO:0000166	nucleotide binding	MF	3.31E-05	7.71E-04	2.38 (2373,540,48,26)
GO:1901265	nucleoside phosphate binding	MF	3.44E-05	7.91E-04	2.38 (2373,541,48,26)
GO:0004890	GABA-A receptor activity	MF	3.65E-05	8.27E-04	3.34 (2373,10,710,10)
GO:0036094	small molecule binding	MF	3.70E-05	8.27E-04	2.30 (2373,581,48,27)
GO:0016787	hydrolase activity	MF	3.81E-05	8.41E-04	1.84 (2373,384,171,51)
GO:0004675	transmembrane receptor protein serine/threonine kinase activity	MF	3.82E-05	8.32E-04	48.10 (2373,4,37,3)
GO:0005024	transforming growth factor beta-activated receptor activity	MF	3.82E-05	8.21E-04	48.10 (2373,4,37,3)
GO:0022891	substrate-specific transmembrane transporter activity	MF	4.09E-05	8.66E-04	2.27 (2373,152,234,34)
GO:0004970	ionotropic glutamate receptor activity	MF	4.64E-05	9.71E-04	2.96 (2373,11,802,11)
GO:0005234	extracellular-glutamate-gated ion channel activity	MF	4.64E-05	9.59E-04	2.96 (2373,11,802,11)
GO:0008060	ARF GTPase activator activity	MF	5.00E-05	1.02E-03	18.98 (2373,5,100,4)
GO:0043167	ion binding	MF	6.42E-05	1.30E-03	1.82 (2373,941,50,36)
GO:0004012	phospholipid-translocating ATPase activity	MF	6.97E-05	1.39E-03	13.88 (2373,4,171,4)
GO:0046983	protein dimerization activity	MF	8.02E-05	1.58E-03	1.32 (2373,253,957,135)
GO:0050998	nitric-oxide synthase binding	MF	8.28E-05	1.61E-03	4.77 (2373,7,498,7)
GO:0042562	hormone binding	MF	9.11E-05	1.75E-03	11.15 (2373,7,152,5)
GO:0008092	cytoskeletal protein binding	MF	1.06E-04	2.02E-03	4.94 (2373,160,33,11)
GO:0004540	ribonuclease activity	MF	1.22E-04	2.28E-03	3.36 (2373,11,643,10)
GO:0015077	monovalent inorganic cation transmembrane transporter activity	MF	1.51E-04	2.80E-03	3.73 (2373,53,168,14)
GO:0015464	acetylcholine receptor activity	MF	1.68E-04	3.09E-03	7.23 (2373,5,328,5)
GO:0005548	phospholipid transporter activity	MF	1.69E-04	3.06E-03	9.91 (2373,7,171,5)

GO:0008565	protein transporter activity	MF	2.03E-04	3.65E-03	7.73 (2373,25,86,7)
GO:0016361	activin receptor activity, type I	MF	2.24E-04	3.98E-03	65.92 (2373,2,36,2)
GO:0034711	inhibin binding	MF	2.37E-04	4.16E-03	64.14 (2373,2,37,2)
GO:0051082	unfolded protein binding	MF	2.41E-04	4.19E-03	2.03 (2373,26,944,21)
GO:0005178	integrin binding	MF	2.41E-04	4.15E-03	20.28 (2373,12,39,4)
GO:0008331	high voltage-gated calcium channel activity	MF	2.43E-04	4.14E-03	10.23 (2373,4,232,4)
GO:0019900	kinase binding	MF	2.67E-04	4.50E-03	2.16 (2373,123,277,31)
GO:0008179	adenylate cyclase binding	MF	2.85E-04	4.75E-03	9.89 (2373,4,240,4)
GO:0019199	transmembrane receptor protein kinase activity	MF	2.90E-04	4.80E-03	3.34 (2373,12,592,10)
GO:0004016	adenylate cyclase activity	MF	3.06E-04	5.01E-03	56.50 (2373,2,42,2)
GO:0051015	actin filament binding	MF	3.12E-04	5.05E-03	5.93 (2373,25,128,8)
GO:0005245	voltage-gated calcium channel activity	MF	3.33E-04	5.35E-03	5.92 (2373,12,234,7)
GO:0004532	exoribonuclease activity	MF	3.46E-04	5.49E-03	5.67 (2373,7,359,6)
GO:0004535	poly(A)-specific ribonuclease activity	MF	3.46E-04	5.44E-03	5.67 (2373,7,359,6)
GO:0016896	exoribonuclease activity, producing 5'-phosphomonoesters	MF	3.46E-04	5.39E-03	5.67 (2373,7,359,6)
GO:0000175	3'-5'-exoribonuclease activity	MF	3.46E-04	5.34E-03	5.67 (2373,7,359,6)
GO:0005200	structural constituent of cytoskeleton	MF	3.49E-04	5.33E-03	5.21 (2373,32,128,9)
GO:0047696	beta-adrenergic receptor kinase activity	MF	3.68E-04	5.58E-03	51.59 (2373,2,46,2)
GO:0015085	calcium ion transmembrane transporter activity	MF	4.13E-04	6.21E-03	3.69 (2373,33,234,12)
GO:0008556	potassium-transporting ATPase activity	MF	4.89E-04	7.28E-03	15.61 (2373,3,152,3)
GO:1990239	steroid hormone binding	MF	4.89E-04	7.21E-03	15.61 (2373,3,152,3)
GO:0005391	sodium:potassium-exchanging ATPase activity	MF	4.89E-04	7.15E-03	15.61 (2373,3,152,3)
GO:0005102	receptor binding	MF	4.93E-04	7.15E-03	1.39 (2373,209,810,99)
GO:0046933	proton-transporting ATP synthase activity, rotational mechanism	MF	6.17E-04	8.86E-03	14.56 (2373,3,163,3)
GO:0019901	protein kinase binding	MF	7.52E-04	1.07E-02	2.14 (2373,112,277,28)
GO:0008378	galactosyltransferase activity	MF	7.75E-04	1.09E-02	13.48 (2373,3,176,3)

GO:0017124	SH3 domain binding	MF	8.67E-04	1.22E-02	2.19 (2373,23,800,17)
GO:0030674	protein binding, bridging	MF	8.73E-04	1.21E-02	2.07 (2373,29,789,20)
GO:0004013	adenosylhomocysteinase activity	MF	9.42E-04	1.30E-02	45.63 (2373,2,52,2)
GO:0016802	trialkylsulfonium hydrolase activity	MF	9.42E-04	1.29E-02	45.63 (2373,2,52,2)
GO:0016801	hydrolase activity, acting on ether bonds	MF	9.42E-04	1.28E-02	45.63 (2373,2,52,2)
GO:0000786	nucleosome	CC	9.87E-27	9.06E-24	2.62 (2373,65,906,65)
GO:1990104	DNA bending complex	CC	9.87E-27	4.53E-24	2.62 (2373,65,906,65)
GO:0044815	DNA packaging complex	CC	9.87E-27	3.02E-24	2.62 (2373,65,906,65)
GO:0032993	protein-DNA complex	CC	7.17E-23	1.65E-20	2.46 (2373,70,923,67)
GO:0030119	AP-type membrane coat adaptor complex	CC	4.86E-17	8.93E-15	23.44 (2373,14,94,13)
GO:0030131	clathrin adaptor complex	CC	1.27E-14	1.94E-12	23.14 (2373,12,94,11)
GO:0005765	lysosomal membrane	CC	1.29E-08	1.70E-06	5.41 (2373,47,168,18)
GO:0016469	proton-transporting two-sector ATPase complex	CC	2.61E-08	3.00E-06	14.21 (2373,7,167,7)
GO:0016020	membrane	CC	2.99E-08	3.05E-06	1.19 (2373,1079,810,438)
GO:0005576	extracellular region	CC	3.02E-08	2.77E-06	1.73 (2373,97,978,69)
GO:0044437	vacuolar part	CC	4.86E-08	4.06E-06	4.79 (2373,56,168,19)
GO:0045211	postsynaptic membrane	CC	5.41E-08	4.14E-06	1.99 (2373,74,807,50)
GO:0005885	Arp2/3 protein complex	CC	6.90E-08	4.87E-06	18.54 (2373,6,128,6)
GO:0005856	cytoskeleton	CC	7.77E-08	5.10E-06	7.56 (2373,120,34,13)
GO:0044459	plasma membrane part	CC	1.19E-07	7.30E-06	1.40 (2373,333,832,164)
GO:0005774	vacuolar membrane	CC	1.81E-07	1.04E-05	4.71 (2373,54,168,18)
GO:0005834	heterotrimeric G-protein complex	CC	2.13E-07	1.15E-05	2.93 (2373,19,767,18)
GO:0034707	chloride channel complex	CC	2.17E-07	1.11E-05	3.22 (2373,15,738,15)
GO:0030665	clathrin-coated vesicle membrane	CC	2.29E-07	1.10E-05	11.56 (2373,21,88,9)

GO:0044425	membrane part	CC	3.00E-07	1.38E-05	1.25 (2373,720,810,306)
GO:0097060	synaptic membrane	CC	3.23E-07	1.41E-05	1.85 (2373,89,808,56)
GO:0000228	nuclear chromosome	CC	3.29E-07	1.37E-05	2.28 (2373,31,906,27)
GO:0072562	blood microparticle	CC	8.32E-07	3.32E-05	36.51 (2373,13,25,5)
GO:0030662	coated vesicle membrane	CC	1.50E-06	5.73E-05	9.71 (2373,25,88,9)
GO:0043234	protein complex	CC	1.61E-06	5.90E-05	1.17 (2373,888,991,433)
GO:0005886	plasma membrane	CC	1.64E-06	5.79E-05	1.29 (2373,526,810,231)
GO:0098588	bounding membrane of organelle	CC	1.75E-06	5.94E-05	2.05 (2373,329,176,50)
GO:0030286	dynein complex	CC	1.81E-06	5.95E-05	4.51 (2373,10,526,10)
GO:0030120	vesicle coat	CC	2.47E-06	7.83E-05	5.07 (2373,14,368,11)
GO:0044444	cytoplasmic part	CC	2.48E-06	7.60E-05	1.38 (2373,1199,177,123)
GO:0031090	organelle membrane	CC	3.00E-06	8.87E-05	1.88 (2373,413,177,58)
GO:0030014	CCR4-NOT complex	CC	3.63E-06	1.04E-04	5.26 (2373,12,376,10)
GO:0033177	proton-transporting two-sector ATPase complex, proton-transporting domain	CC	4.11E-06	1.14E-04	14.47 (2373,5,164,5)
GO:0005615	extracellular space	CC	6.47E-06	1.75E-04	8.77 (2373,87,28,9)
GO:0044427	chromosomal part	CC	6.64E-06	1.74E-04	1.44 (2373,183,961,107)
GO:0030126	COPI vesicle coat	CC	9.23E-06	2.35E-04	6.45 (2373,7,368,7)
GO:0034704	calcium channel complex	CC	9.97E-06	2.47E-04	5.82 (2373,17,240,10)
GO:0044431	Golgi apparatus part	CC	1.22E-05	2.96E-04	1.98 (2373,156,368,48)
GO:0030659	cytoplasmic vesicle membrane	CC	1.29E-05	3.04E-04	2.81 (2373,83,254,25)
GO:0012506	vesicle membrane	CC	1.61E-05	3.70E-04	2.78 (2373,84,254,25)

GO:0044433	cytoplasmic vesicle part	CC	1.63E-05	3.65E-04	2.17 (2373,113,368,38)
GO:0032588	trans-Golgi network membrane	CC	1.85E-05	4.05E-04	13.20 (2373,13,83,6)
GO:0005869	dynactin complex	CC	2.10E-05	4.48E-04	57.41 (2373,4,31,3)
GO:0005794	Golgi apparatus	CC	2.11E-05	4.41E-04	2.62 (2373,135,188,28)
GO:0030122	AP-2 adaptor complex	CC	2.65E-05	5.41E-04	21.82 (2373,5,87,4)
GO:0098533	ATPase dependent transmembrane transport complex	CC	2.79E-05	5.57E-04	12.84 (2373,6,154,5)
GO:0044456	synapse part	CC	4.56E-05	8.90E-04	1.55 (2373,138,808,73)
GO:0045259	proton-transporting ATP synthase complex	CC	5.21E-05	9.96E-04	14.74 (2373,4,161,4)
GO:0005753	mitochondrial proton-transporting ATP synthase complex	CC	5.21E-05	9.75E-04	14.74 (2373,4,161,4)
GO:0033178	proton-transporting two-sector ATPase complex, catalytic domain	CC	6.70E-05	1.23E-03	14.12 (2373,4,168,4)
GO:0030117	membrane coat	CC	6.95E-05	1.25E-03	3.87 (2373,20,368,12)
GO:0005832	chaperonin-containing T-complex	CC	8.21E-05	1.45E-03	7.19 (2373,7,283,6)
GO:0005680	anaphase-promoting complex	CC	8.96E-05	1.55E-03	7.09 (2373,7,287,6)
GO:0000152	nuclear ubiquitin ligase complex	CC	8.96E-05	1.52E-03	7.09 (2373,7,287,6)
GO:0000307	cyclin-dependent protein kinase holoenzyme complex	CC	1.03E-04	1.72E-03	7.99 (2373,5,297,5)
GO:0005891	voltage-gated calcium channel complex	CC	1.15E-04	1.88E-03	7.61 (2373,8,234,6)
GO:0005868	cytoplasmic dynein complex	CC	1.33E-04	2.13E-03	4.51 (2373,7,526,7)
GO:0043228	non-membrane-bounded organelle	CC	1.57E-04	2.49E-03	2.89 (2373,410,34,17)
GO:0043232	intracellular non-membrane-bounded organelle	CC	1.57E-04	2.45E-03	2.89 (2373,410,34,17)
GO:0015629	actin cytoskeleton	CC	1.58E-04	2.41E-03	2.69 (2373,40,419,19)
GO:0005852	eukaryotic translation initiation factor 3 complex	CC	1.79E-04	2.69E-03	4.29 (2373,7,553,7)
GO:0044430	cytoskeletal part	CC	2.57E-04	3.80E-03	3.31 (2373,304,33,14)
GO:0090533	cation-transporting ATPase complex	CC	2.95E-04	4.30E-03	12.33 (2373,5,154,4)
GO:0043190	ATP-binding cassette (ABC) transporter complex	CC	4.21E-04	6.04E-03	2,373.00 (2373,1,1,1)
GO:0031209	SCAR complex	CC	4.42E-04	6.25E-03	105.47 (2373,5,9,2)
GO:0005890	sodium:potassium-exchanging ATPase complex	CC	4.89E-04	6.80E-03	15.61 (2373,3,152,3)

GO:0005887	integral component of plasma membrane	CC	5.68E-04	7.78E-03	1.49 (2373,126,832,66)
GO:0033179	proton-transporting V-type ATPase, V0 domain	CC	6.22E-04	8.40E-03	14.47 (2373,3,164,3)
GO:0033180	proton-transporting V-type ATPase, V1 domain	CC	6.71E-04	8.93E-03	14.12 (2373,3,168,3)
GO:0005884	actin filament	CC	7.29E-04	9.56E-03	26.97 (2373,8,33,3)
GO:0005813	centrosome	CC	7.39E-04	9.55E-03	1.97 (2373,69,525,30)
GO:0031461	cullin-RING ubiquitin ligase complex	CC	8.03E-04	1.02E-02	2.61 (2373,32,454,16)
GO:0032432	actin filament bundle	CC	9.95E-04	1.25E-02	15.41 (2373,22,28,4)
GO:0001725	stress fiber	CC	9.95E-04	1.23E-02	15.41 (2373,22,28,4)

Table S16.1. Genes located within the amphibian-specific HCEs.

Gene ID	CDS conserved ratio	Gene name
Npar_R017495	0.812	JPH4
Npar_R007094	0.731	Olfm2
Npar_R015639	0.743	TSPAN17
Npar_R020116	0.75	NTRK3
Npar_R005660	0.735	LPPR4
Npar_R005659	0.86	lppr5
Npar_R018212	0.724	MYCBP
Npar_R018211	0.866	Rragc
Npar_R000561	0.707	RNF5
Npar_R013170	0.708	UBE2E2
Npar_R013171	1	UBE2E2
Npar_R004773	0.822	PCDHGC5
Npar_R020584	0.712	CRYBB1

Npar_R009436	0.786	VPS37D
Npar_R004570	0.703	-
Npar_R000933	0.76	ctbp2
Npar_R000932	0.814	Fam53c
Npar_R009203	0.871	RER1
Npar_R017093	0.731	KCNN3
Npar_R015220	0.735	ADCY6
Npar_R013877	0.733	Twist2
Npar_R002607	0.709	rest-a
Npar_R002612	0.736	ANAPC10
Npar_R002603	0.728	slc10a7-b
Npar_R006411	0.818	MTNR1A
Npar_R019924	0.781	NFIA
Npar_R019923	0.991	Nfia
Npar_R004823	0.72	Eif4e2
Npar_R007526	0.705	CSDE1
Npar_R011806	0.991	Rpl37
Npar_R011807	0.709	PRKAA1
Npar_R017574	1	TGas113e22.1
Npar_R009871	0.748	PRKAR1B
Npar_R007086	0.709	Nt5c1a
Npar_R007087	0.796	Hpcal4
Npar_R012961	0.833	Gje1
Npar_R012691	0.747	EFNA5
Npar_R000530	0.709	GYS1
Npar_R023165	0.78	S100Z

Npar_R009308	0.896	LHFPL2
Npar_R005108	0.767	Dpyd
Npar_R016358	0.719	MAEA
Npar_R004548	0.757	Hspa8
Npar_R001870	0.728	Fibp
Npar_R010685	0.709	Unc13c
Npar_R002761	0.952	C1orf70
Npar_R002766	0.766	TMEM88B
Npar_R022659	0.831	ube2s
Npar_R022647	0.847	U2af2
Npar_R003707	0.727	Lrrc3b
Npar_R007165	0.857	DR1
Npar_R007169	0.701	TECR
Npar_R010980	0.797	EIF2C3
Npar_R010978	0.784	eif2c4
Npar_R010985	0.755	TRAPPC3
Npar_R010979	0.823	EIF2C1
Npar_R005438	0.806	MYL2
Npar_R000078	0.909	TMEM57
Npar_R017453	0.752	DLGAP3
Npar_R017656	0.878	DLGAP3
Npar_R011198	0.795	rasl11b
Npar_R004836	0.827	Rps26
Npar_R004827	0.77	Gdf11
Npar_R018465	0.823	Ldb2
Npar_R011164	0.769	SYT6

Npar_R009456	0.77	ELOVL6
Npar_R015542	0.941	hand2
Npar_R007694	0.919	ODZ2
Npar_R013231	0.779	EPHA7
Npar_R004916	0.879	MSANTD1
Npar_R004912	0.746	GRK4
Npar_R005946	0.801	PITPNA
Npar_R007472	0.942	Stox2
Npar_R018047	0.958	CDK5R1
Npar_R009983	0.732	rgs7bp
Npar_R007852	0.708	TTC13
Npar_R018578	0.74	Grid2
Npar_R019159	0.739	wnt10a
Npar_R004375	0.712	GOLPH3
Npar_R001639	0.771	Cplx4
Npar_R001643	0.727	Polr2b
Npar_R018966	0.898	HIST2H2AB
Npar_R018976	0.918	histh2b
Npar_R018971	1	TGas006m08.1
Npar_R003335	0.764	EDIL3
Npar_R018258	0.749	SLC25A44
Npar_R018256	0.855	TOMM40L
Npar_R012656	0.783	ODZ2
Npar_R011522	0.84	Ptpn9
Npar_R007916	0.702	Agbl4
Npar_R022986	0.881	HPGD

Npar_R005286	0.851	Kcnn2
Npar_R014130	0.701	NPY5R
Npar_R005095	0.735	ARL15
Npar_R018041	0.774	mapk8
Npar_R012712	0.718	HTR1A
Npar_R019075	0.739	isoc1
Npar_R014326	0.815	Mxd4
Npar_R001592	0.769	JMJD1C
Npar_R004156	0.754	cs
Npar_R018816	0.873	VAMP8
Npar_R009230	0.898	Onecut2
Npar_R004085	0.72	TNK2
Npar_R019626	0.781	RIMS3
Npar_R000576	0.842	med10
Npar_R000577	0.792	UBE2QL1
Npar_R002723	0.718	tmem198
Npar_R019739	1	Clvs2
Npar_R020356	0.883	ISL1
Npar_R000993	0.772	sst1
Npar_R013721	0.809	arl8ba
Npar_R007418	0.862	DHX15
Npar_R008782	0.733	GUCA1A
Npar_R012926	0.716	Unc5c
Npar_R000876	0.801	WDR17
Npar_R011107	0.701	CACNA1I
Npar_R011108	0.761	Cacna1i

Npar_R000732	0.842	RAB3C
Npar_R000733	0.889	PDE4D
Npar_R000730	0.782	Rab3c
Npar_R017908	0.731	SHISA9
Npar_R003914	0.847	RPL36
Npar_R000367	0.748	Ssbp2
Npar_R005963	0.892	sox1
Npar_R008699	0.701	Syt2
Npar_R022231	0.835	dph1
Npar_R016363	0.917	Actc1
Npar_R013469	0.724	TNeu083d09
Npar_R013470	0.764	PRDM5
Npar_R022690	0.736	CAMTA1
Npar_R022699	0.736	PRDM16
Npar_R020684	0.747	-
Npar_R010718	0.822	ADAMTS6
Npar_R021790	0.843	PRDM6
Npar_R007877	0.836	Ptbp2
Npar_R015061	0.771	HIST1H1E
Npar_R015063	0.878	TGas006m08.1
Npar_R003122	0.769	lmo4.2
Npar_R003123	0.851	hs2st1
Npar_R002880	0.927	otx5
Npar_R002033	0.742	MCC
Npar_R015030	0.765	Tspan5
Npar_R005492	0.922	Arl4c

Npar_R012610	0.739	C3orf70
Npar_R020991	0.703	Rab3c
Npar_R020979	0.981	elavl4
Npar_R015717	0.912	Sstr4
Npar_R002691	0.845	MAML3
Npar_R014139	0.765	CLDN3
Npar_R002853	0.972	GNG12
Npar_R018275	0.734	Otof
Npar_R000158	0.833	Kcnh1
Npar_R000136	0.854	Prox1
Npar_R020530	0.869	CAPZB
Npar_R006969	0.709	WNT7B
Npar_R020947	0.841	PDE1C
Npar_R011101	0.844	NR2F2
Npar_R019304	0.818	Doc2b
Npar_R019805	0.763	Lrrc7
Npar_R007332	0.712	Chrm5
Npar_R001682	0.725	nmnat2
Npar_R000396	0.745	Mark1
Npar_R008240	0.869	h2afx
Npar_R018628	0.731	GSTENG00019215001
Npar_R005054	0.819	Polr2a
Npar_R006789	0.715	HNRNPR
Npar_R015894	0.775	thra
Npar_R013626	0.761	Serp1
Npar_R010818	0.877	Cplx1

Npar_R001283	0.937	Pcbp3
Npar_R010189	0.852	SPRYD3
Npar_R012812	0.753	TUBB3
Npar_R021075	0.701	SFRP2
Npar_R010379	0.903	Prkcg
Npar_R020904	0.741	dpysl3
Npar_R007677	0.931	HS6ST2
Npar_R019424	0.773	tmem151b
Npar_R016779	0.816	actc1
Npar_R017836	0.742	DLGAP3
Npar_R014068	0.724	ube2m
Npar_R004078	0.741	SCOP1
Npar_R008470	0.851	Cacng8
Npar_R008471	0.776	Cacng7
Npar_R003405	0.729	rasl11b
Npar_R003410	0.985	CHIC2
Npar_R018454	0.754	FAM49A
Npar_R018335	0.788	CRLF1
Npar_R011863	0.851	SPCS3
Npar_R006185	0.84	Tmem127
Npar_R021324	0.752	TRIM2
Npar_R011785	1	Snx27
Npar_R021335	0.802	nkain1
Npar_R019552	0.913	Ube2k
Npar_R010954	0.8	GABRB1
Npar_R019566	0.713	BEND4

Npar_R019556	0.737	Rbm47
Npar_R009518	0.92	Cnot3
Npar_R003515	0.893	csnk2b
Npar_R020445	0.924	ANP32B
Npar_R021517	0.771	wdr26
Npar_R004170	0.701	Mat1a
Npar_R018708	0.747	rpl18-b
Npar_R007031	0.746	camk2n2
Npar_R019721	0.838	SPSB1
Npar_R019708	0.859	Deaf1
Npar_R019710	0.76	Rere
Npar_R003871	0.728	PHF1
Npar_R000955	0.743	Fau
Npar_R013603	0.807	rbm4.1
Npar_R006950	0.857	HIST2H2AB
Npar_R006953	0.977	TGas006m08.1
Npar_R012439	0.853	CTNND2
Npar_R012438	0.812	CTNND2
Npar_R001901	0.865	tmem178
Npar_R007495	1	-
Npar_R019003	0.782	Trappc1
Npar_R003961	0.795	ARF6

Table S16.2. GO terms enriched by genes in amphibians-specific HCEs.

GO term	Description	GO_Class	P-value	FDR q-value	Enrichment (N, B, n, b)
GO:0045974	regulation of translation, ncRNA-mediated	BP	1.43E-04	1.00E+00	26.85 (12744,8,178,3)
GO:0040033	negative regulation of translation, ncRNA-mediated	BP	1.43E-04	8.75E-01	26.85 (12744,8,178,3)
GO:0035278	negative regulation of translation involved in gene silencing by miRNA	BP	1.43E-04	5.83E-01	26.85 (12744,8,178,3)
GO:0060849	regulation of transcription involved in lymphatic endothelial cell fate commitment	BP	1.94E-04	5.96E-01	71.60 (12744,2,178,2)
GO:0043586	tongue development	BP	2.12E-04	5.20E-01	23.87 (12744,9,178,3)
GO:0032879	regulation of localization	BP	2.99E-04	6.11E-01	1.77 (12744,1539,178,38)
GO:0050794	regulation of cellular process	BP	3.36E-04	5.89E-01	1.24 (12744,6820,178,118)
GO:0006402	mRNA catabolic process	BP	3.91E-04	6.00E-01	4.05 (12744,159,178,9)
GO:0007268	synaptic transmission	BP	5.59E-04	7.62E-01	2.77 (12744,362,178,14)
GO:0016071	mRNA metabolic process	BP	6.36E-04	7.81E-01	2.43 (12744,500,178,17)
GO:0055003	cardiac myofibril assembly	BP	6.91E-04	7.72E-01	16.52 (12744,13,178,3)
GO:0060341	regulation of cellular localization	BP	7.96E-04	8.15E-01	2.09 (12744,752,178,22)
GO:0008543	fibroblast growth factor receptor signaling pathway	BP	8.47E-04	8.00E-01	4.03 (12744,142,178,8)
GO:0006401	RNA catabolic process	BP	9.94E-04	8.72E-01	3.56 (12744,181,178,9)
GO:0016286	small conductance calcium-activated potassium channel activity	MF	1.94E-04	7.45E-01	71.60 (12744,2,178,2)
GO:0046875	ephrin receptor binding	MF	4.33E-04	8.31E-01	11.01 (12744,26,178,4)
GO:0044456	synapse part	CC	3.12E-06	4.46E-03	3.56 (12744,362,178,18)
GO:0035068	micro-ribonucleoprotein complex	CC	2.62E-05	1.88E-02	42.96 (12744,5,178,3)
GO:0000974	Prp19 complex	CC	5.20E-05	2.48E-02	35.80 (12744,6,178,3)

GO:0045202	synapse	CC	9.15E-05	3.27E-02	4.04 (12744,195,178,11)
GO:0008328	ionotropic glutamate receptor complex	CC	3.51E-04	1.00E-01	8.14 (12744,44,178,5)
GO:0005891	voltage-gated calcium channel complex	CC	6.65E-04	1.59E-01	9.88 (12744,29,178,4)
GO:0097060	synaptic membrane	CC	7.89E-04	1.61E-01	3.38 (12744,212,178,10)

Table S17. Calibration times used in the divergence time estimation.

Species 1	Species 2	Lower bound (Ma)	Upper bound (Ma)
A. carolinensis	*G. gallus*	259.7	299.8
A. carolinensis + G. gallus	*H. sapiens*	312.3	330.4
A. carolinensis + G. gallus + H. sapiens	*X. tropicalis + N. parkeri*	330.4	350.1
A. carolinensis + G. gallus + H. sapiens + X .tropicalis + N. parkeri	*D. rerio*	416	421.75

Supplementary Figures:

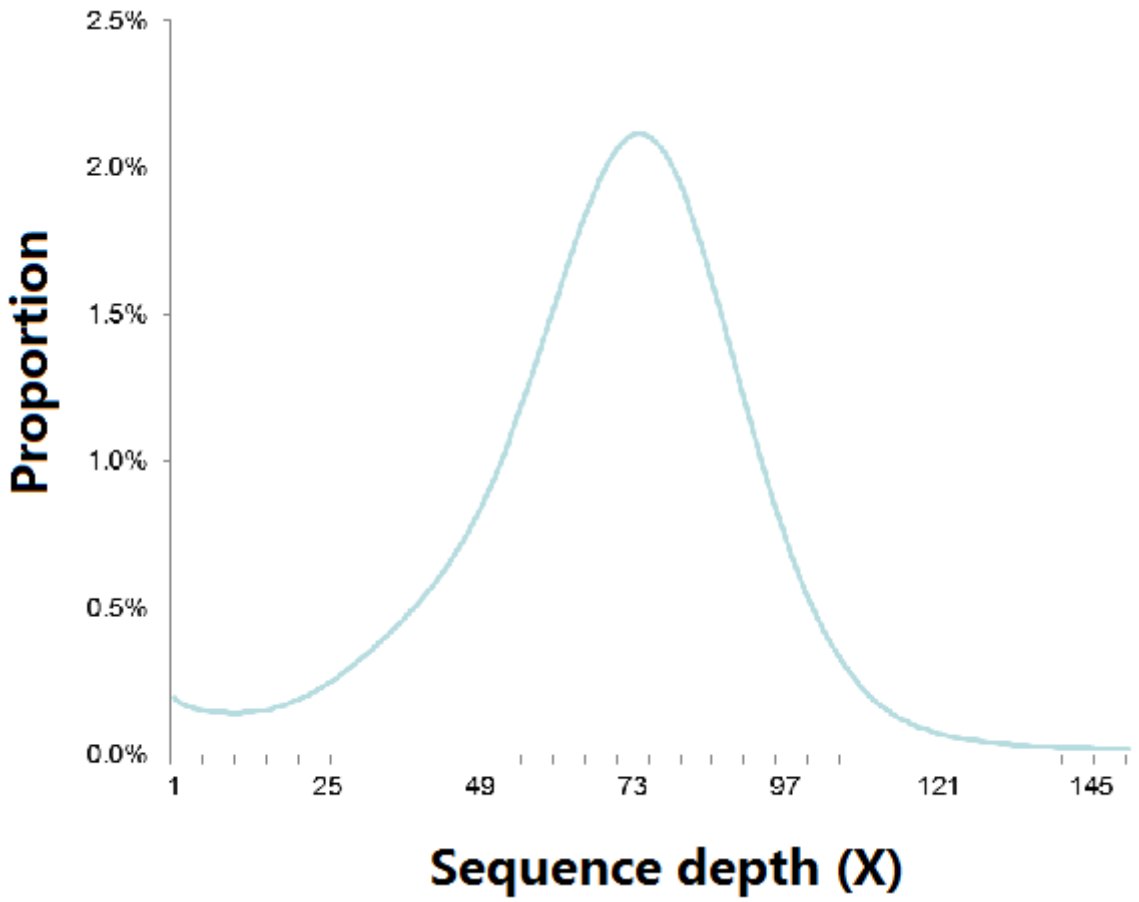

Fig. S1. Distribution of sequencing depth in the assembled genome of *Nanorana parkeri*.

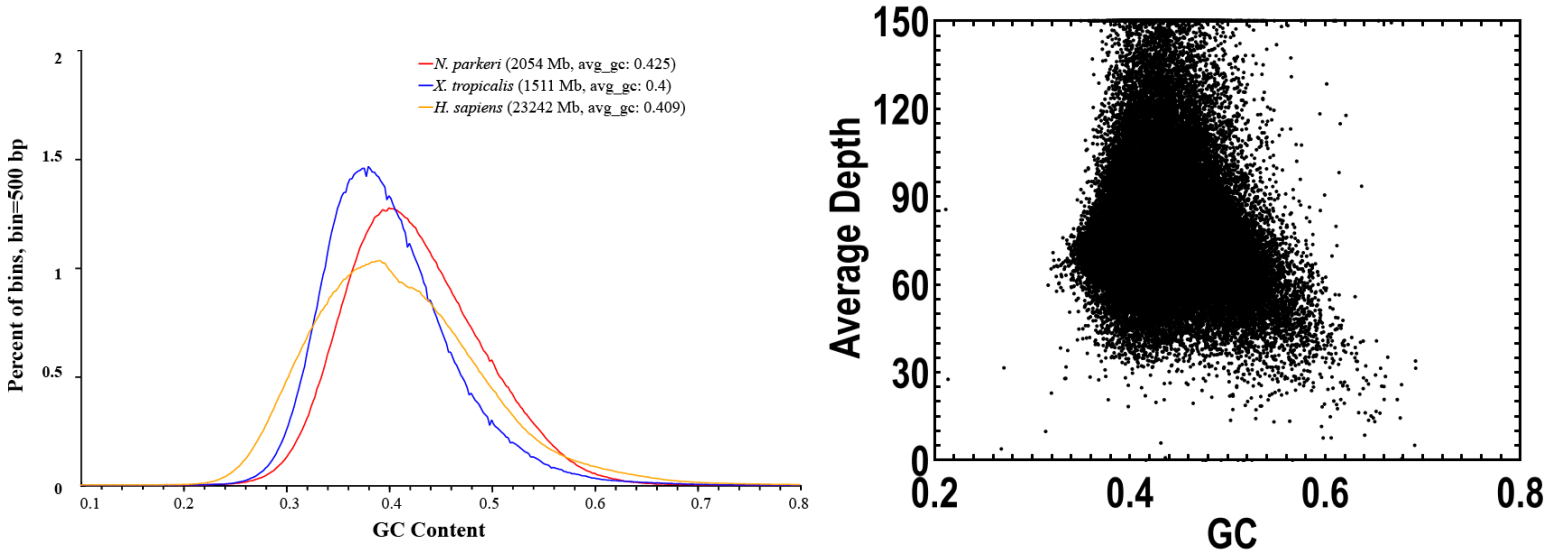

Fig. S2. GC content in the genome of *Nanorana parkeri*.

(a) Distribution of GC content in the *N. parkeri* (red), *X. tropicalis* (blue), *H. sapiens* (yellow) genomes. Proportion of 500bp non-overlapping sliding windows with a given GC content is shown. (b) GC content versus sequencing depth. During this analysis, we used 10kb non-overlapping sliding windows across the genome to calculate the GC content and average sequencing depth.

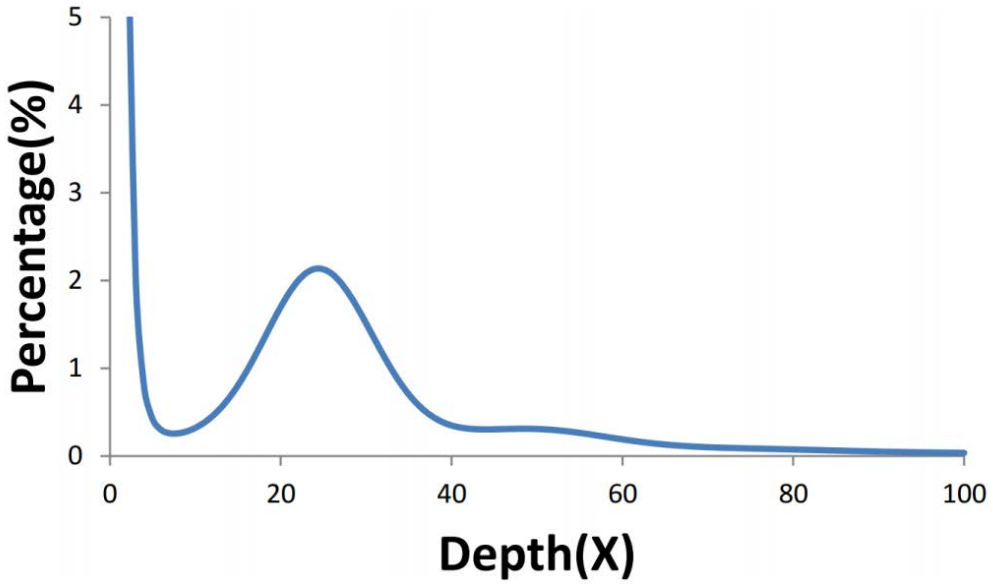

Fig. S3. 17-mer depth distribution for estimation of genome size.

Peak depth is at 24X. The total number of k-mers is 55,450,398,715. The genome size can be calculated from the formula G=K_num/K_depth.

The *Nanorana parkeri* genome size was therefore estimated to be 2.3Gb.

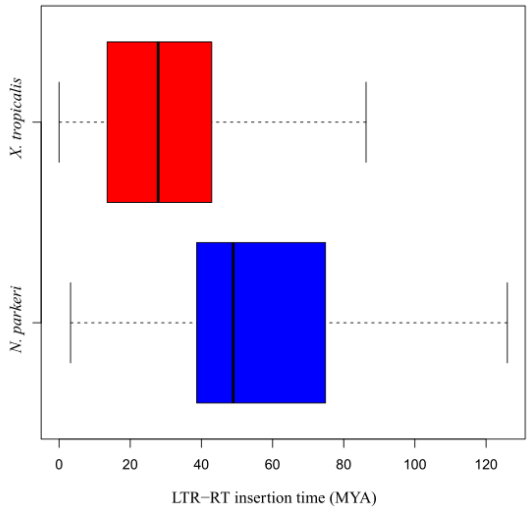

Fig. S4. Distributions of insertion times calculated for LTR-RTs in *N. parkeri* and *X. tropicalis*, using a mutation rate of 0.776E-9 per site per year.

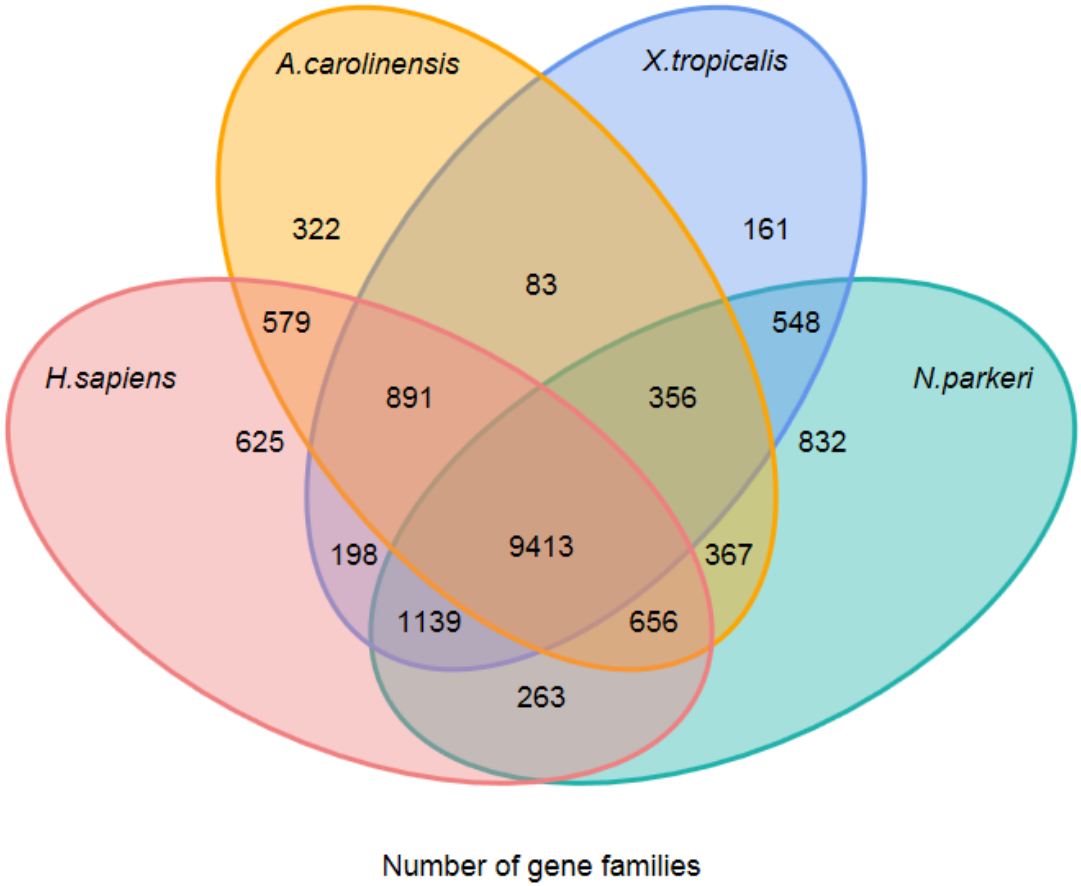

Fig. S5. Venn diagram showing unique and shared gene families between the *A. carolinensis, X. tropicalis, N. parkeri* and *H. sapiens* genomes. The number of gene families is listed in each of the diagram components.

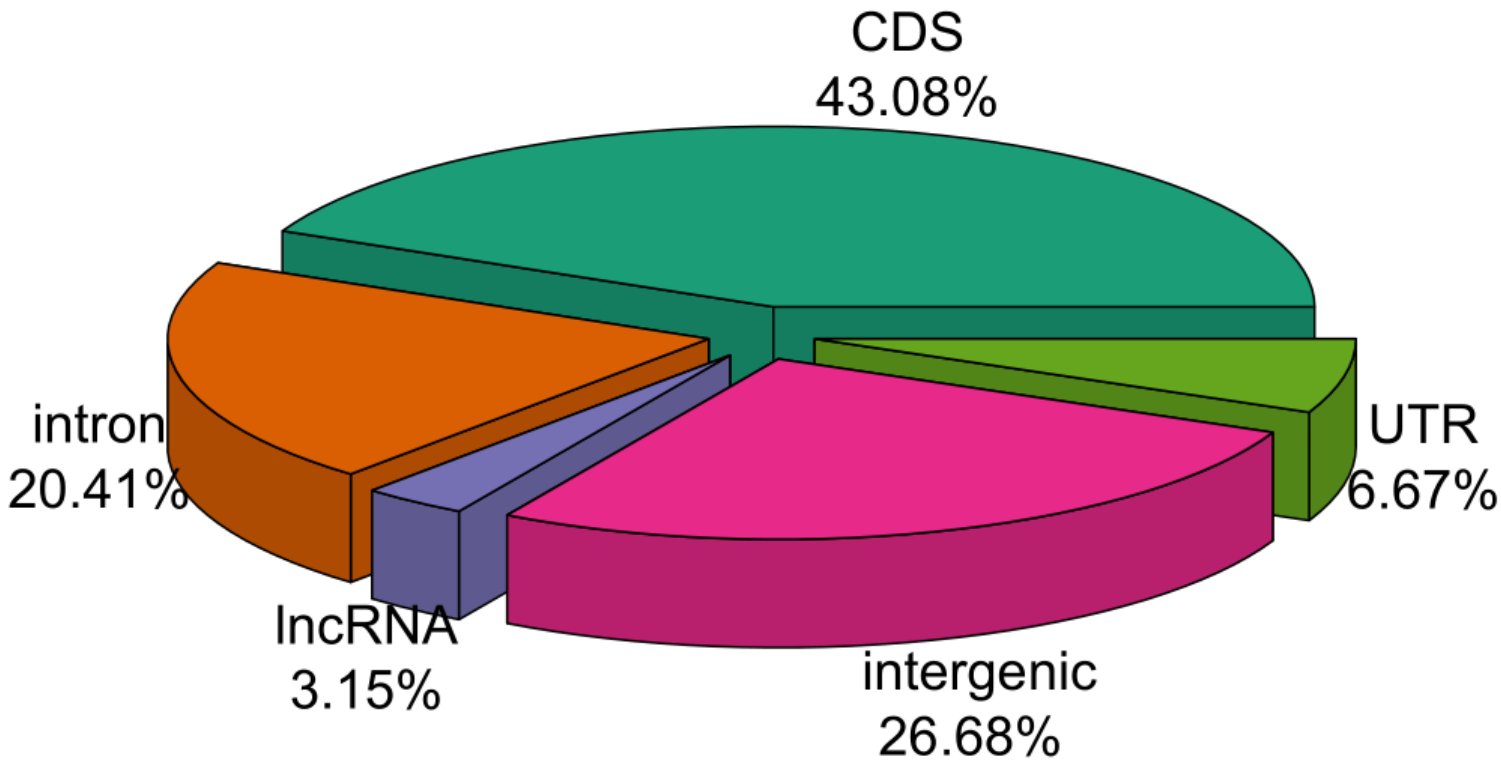

Fig. S6. Distribution pattern of tetrapod Highly Conserved Elements (HCEs).

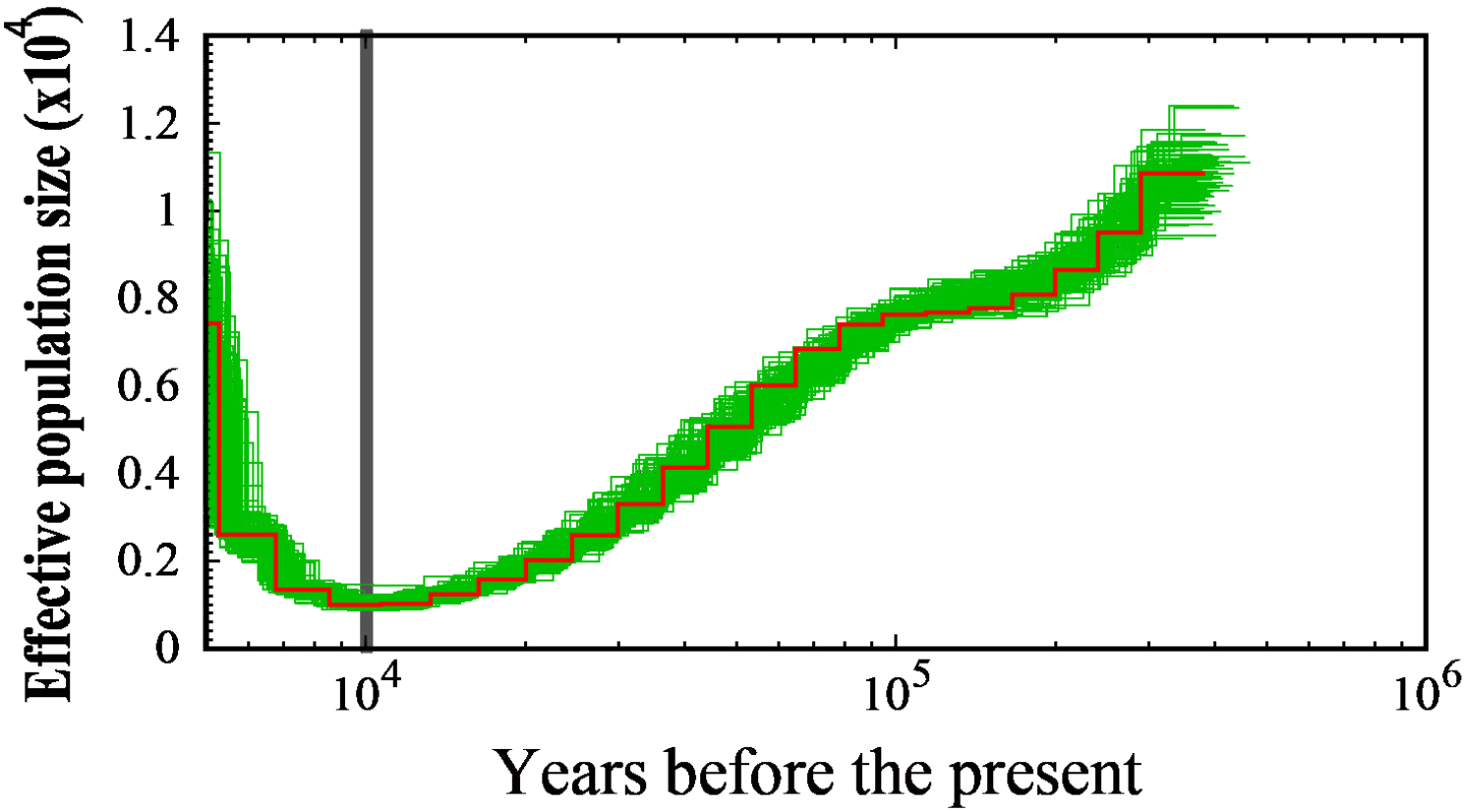

Fig. S7. Population history deduced for the Tibetan frog, *Nanorana parkeri*. Historical changes of effective population size (*Ne*); *Ne* decreased until 10,000 ybp, when subsequently, a rapid expansion occurred resulting to the current *Ne* at about 15,000. Gray vertical line refers to the approximate ending time of the last glacial maximum in this region (Tibet; 31).

References

1. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* 17(1):pp. 10-12.
2. Kong Y (2011) Btrim: a fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. *Genomics* 98(2):152-153.
3. Li H & Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754-1760.
4. Li R, *et al.* (2009) SNP detection for massively parallel whole-genome resequencing. *Genome Res* 19(6):1124-1132.
5. Li H, *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078-2079.
6. Edgar RC & Myers EW (2005) PILER: identification and classification of genomic repeats. *Bioinformatics* 21 Suppl 1:i152-158.
7. Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 27(2):573-580.
8. Xu Z & Wang H (2007) LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res* 35(Web Server issue):W265-268.
9. Flicek P, *et al.* (2012) Ensembl 2012. *Nucleic Acids Res* 40(Database issue):D84-90.
10. McCarthy EM & McDonald JF (2003) LTR_STRUC: a novel search and identification program for LTR retrotransposons. *Bioinformatics* 19(3):362-367.
11. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792-1797.
12. Rice P, Longden I, & Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet : TIG* 16(6):276-277.
13. Kent WJ, Baertsch R, Hinrichs A, Miller W, & Haussler D (2003) Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proc Natl Acad Sci USA* 100(20):11484-11489.
14. Kent WJ (2002) BLAT--the BLAST-like alignment tool. *Genome Res* 12(4):656-664.
15. Birney E, Clamp M, & Durbin R (2004) GeneWise and Genomewise. *Genome Res* 14(5):988-995.
16. Trapnell C, Pachter L, & Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25(9):1105-1111.
17. Trapnell C, *et al.* (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 7(3):562-578.
18. Trapnell C, *et al.* (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28(5):511-515.
19. Stanke M, Schoffmann O, Morgenstern B, & Waack S (2006) Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC bioinformatics* 7:62.
20. Bairoch A & Apweiler R (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res* 28(1):45-48.
21. Zdobnov EM & Apweiler R (2001) InterProScan--an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17(9):847-848.
22. Ashburner M, *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25(1):25-29.

23. Kanehisa M & Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28(1):27-30.
24. Ma J, *et al.* (2006) Reconstructing contiguous regions of an ancestral genome. *Genome Res* 16(12):1557-1565.
25. Siepel A, *et al.* (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 15(8):1034-1050.
26. Li H, *et al.* (2006) TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Res* 34(Database issue):D572-580.
27. 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.
28. Guindon S, *et al.* (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59(3):307-321.
29. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24(8):1586-1591.
30. Katoh K & Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30(4):772-780.
31. Owen LA & Benn DI (2005) Equilibrium-line altitudes of the Last Glacial Maximum for the Himalaya and Tibet: an assessment and evaluation of results. *Quaternary International* 138:55-78.