

# Genome-Wide Analyses of the Relict Gull (*Larus Relictus*) Insights Into Evolutionary Implications

**Chao Yang**

Shaanxi Institute of Zoology

**Xuejuan Li**

Shaanxi Normal University

**Qingxiong Wang**

Shaanxi Institute of Zoology

**Hao Yuan**

Shaanxi Normal University

**Yuan Huang** (✉ [yuanh@snnu.edu.cn](mailto:yuanh@snnu.edu.cn))

Shaanxi Normal University <https://orcid.org/0000-0001-7683-9193>

**Hong Xiao**

Shaanxi Institute of Zoology

---

## Research article

**Keywords:** Whole-genome, PacBio sequencing, *Larus relictus*, Habitat loss, Population fragmentation

**Posted Date:** November 10th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-103622/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at BMC Genomics on April 29th, 2021. See the published version at <https://doi.org/10.1186/s12864-021-07616-z>.

# Abstract

## Background

The relict gull (*Larus relictus*), one of the least known Aves, was classified as vulnerable on the IUCN Red List and is a first-class national protected bird in China. Genomic resources for *L. relictus* are lacking, which limits the study of its evolution and its conservation.

## Results

In this study, based on the Illumina and PacBio sequencing platforms, we successfully assembled the genome of *L. relictus*, the first reference genome of the genus *Larus*. The size of the final assembled genome was 1.21 Gb, with a contig N50 of 8.11 Mb. A total of 18,454 protein-coding genes were predicted from the assembly results, with 16,967 (91.94%) of these genes annotated. The genome contained 92.52 Mb of repeat sequence, accounting for 7.63% of the assembly. The phylogenetic tree was constructed using 7,339 single-copy orthologous genes, which showed Charadriiformes located at the basal position and *Philomachus pugnax* as the closest relative of *L. relictus*. The divergence time between *L. relictus* and *P. pugnax* was ~68.44 Mya. The population dynamics of the Ordos breeding subpopulation in Hongjian Nur is a good confirmation that these birds are suffering from habitat loss and fragmentation.

## Conclusions

This assembled genome will be a valuable genomic resource for a range of genomic and conservation studies of *L. relictus* and helps to establish a foundation for further studies investigating whether the other three breeding subpopulations have combined with the Ordos breeding subpopulation. As the species is threatened by habitat loss and fragmentation, actions to protect *L. relictus* are suggested to improve the fragmentation of breeding populations.

## Background

The relict gull (*Larus relictus*) (Charadriiformes, Laridae), a middle-sized gull with a black-coloured head, had been known for nearly 50 years before it was regarded as a unique species [1]. *L. relictus* is one of the least known Aves [2,3]. It was classified as vulnerable (VU) on the IUCN Red List and is a first-class national protected bird in China. Its population size has been estimated at 10,000-19,999 (BirdLife International, 2020), and the vast majority of *L. relictus* (90%) reside in Hongjian Nur [4]. Their main wintering place is situated on the west coast of the Bohai Sea [5]. A small number of winter migratory individuals have been sighted in Hong Kong [6]. Therefore, the main threats to *L. relictus* are overwhelming lake shrinkage on breeding grounds and at stopover sites, as well as the loss of intertidal flats on wintering grounds [3]. A novel data-driven habitat suitability ranking approach for *L. relictus* species using remote sensing and GIS indicated that three threat factors, road networks, developed buildings and vegetation, were the most significant factors in the highly suitable region for this species

and supported the development of strategies or recommendations for sustainably managing the ecosystem to enhance the protection of *L. relictus* [7]. Because of the concentration of this population in Hongjian Nur, the birds having to disperse to breed and the phenomenon of abnormal mass mortality of *L. relictus* fledglings in 2011 due to food shortage, habitat loss, and habitat fragmentation are inferred to be the major causes of population fragmentation [4].

On the whole-genome level, genome sequencing technology is usually used to characterize genetic variations and acquire comprehensive molecular characterizations [8]. At present, only limited genetic information, regarding mitochondrial markers and population structures, is available for *L. relictus* [2,4,9,10], as well as dispersal and migration information [3]. However, no genome information for *L. relictus* has been published, which limits our understanding of the evolutionary and molecular mechanisms of some significant processes.

High-throughput sequencing technology has notably reduced sequencing costs [11] and marked the start of a new era of genomic studies [12]. Among them, long-read sequencing technologies such as Pacific Biosciences (PacBio) [13] can produce average lengths of over 10,000 bp [12]. PacBio technology has been used to obtain high-quality genome assemblies for several avian species, such as *Gallus gallus* (Galliformes) [14] and *Malurus cyaneus* (Passeriformes) [15].

In this study, the first contig-level genome of *L. relictus* was constructed using both Illumina HiSeq and PacBio sequencing platforms. Then, we assessed various genomic characteristics and performed comparative analyses. These genomic data will facilitate population studies of *L. relictus* and benefit the comprehensive protection of this vulnerable avian species.

## Results

### Genome sequencing and assembly

Approximately 106.29 Gb of raw sequencing data were obtained using the Illumina HiSeq platform, including three 250-bp insert libraries and two 350-bp insert libraries (Table S1). The sequencing depth was 87.85X. We then used the PacBio sequencing platform to obtain long reads for assembling the genome and retained approximately 30.50 Gb raw data. After filtering out low-quality and short-length reads, the read N50 and mean read length were 12,712 bp and 8,418 bp, respectively. Finally, a 1.21 Gb assembly with a contig N50 of ~8.11 Mb was obtained for *L. relictus* (Table S2), with a GC content of ~43.11%. The genome consisted of 1,313 contigs, with the longest contig being ~29.7 Mb long (Table S2).

Approximately 99.97% of the clean reads could be mapped to the contigs, with 93.33-93.77% properly mapped reads (Table S3). The CEGMA analysis identified 416 CEGs (core eukaryotic genes), accounting for 90.83% of all 458 CEGs, and 175 CEGs (70.56%) could be detected with homology to the 248 highly conserved CEGs (Table S4). In addition, 4,555 (92.7%) of the 4,915 highly conserved Aves orthologues

from BUSCO v3.0.2 were identified in the assembly (Table S5). These results show that the assembled *L. relictus* genome sequence was complete and had a low error ratio.

## Genome Annotation

The consensus gene set included a total of 18,454 PCGs. The average gene length, exon length, and intron length were 20,749.08 bp, 164.24 bp, and 1,996.77 bp, respectively (Table S6). The final prediction results revealed 17,452 (94.57%) PCGs supported by homology-based and RNA-seq-based methods (Fig. S1), which showed a good gene prediction effect. A total of 16,967 (91.94%) predicted PCGs in the *L. relictus* genome were annotated and functionally classified by the GO, KEGG, KOG, TrEMBL and NR databases (Table S7). Noncoding RNAs were also identified and annotated, including 208 miRNAs (microRNA genes), 73 rRNAs and 289 tRNAs. A total of 221 pseudogenes were identified in the *L. relictus* genome.

A total of 92.52 Mb of repeat sequence was annotated, composing 7.63% of the total genome length. We found that class I transposable elements (TEs) (RNA transposons or retrotransposons) occupied ~8.22% of the genome assembly. Among class I TEs, 5.85% were LINEs (long interspersed elements), 1.12% were LTR (long terminal repeat) elements, and 0.02% were SINEs (short interspersed elements) (Table S8). The LINE percentage of *L. relictus* was larger than that of *Charadrius vociferus* (4.53%), while SINEs were obviously less represented than in *C. vociferus* (0.13%) [16]. The *L. relictus* genome also contained class II TEs (DNA transposons), which occupied ~0.28% of the genome.

## Gene families

Comparison of the *L. relictus* genome assembly with the genomes of seven other avian species showed that a total of 18,454 genes of *L. relictus* could be clustered into 12,681 gene families, including 192 unique genes belonging to 56 gene families (Table S9). The proportion of species-specific genes within *L. relictus* genomes was obviously larger than that of other sampled genomes. In addition, 8,650 gene families were shared among all sampled species. The phylogenetic relationships based on 7,339 single-copy orthologous genes indicated that Charadriiformes was basal to three other orders (Fig. 1). *L. relictus* showed the closest relationship to another member of the order Charadriiformes, *P. pugnax*.

## Positive selection genes and functional enrichment

We found that 519 single-copy orthologous genes were under positive selection in the *L. relictus* genome. The Gene Ontology (GO) annotation classifies the positively selected genes (PSGs) in terms of three categories: cellular component, biological process, and molecular function (Fig. S2a). Cellular component annotations were primarily cell (117 genes; GO:0005623), cell part (117 genes; GO:0044464), and organelle (96 genes; GO:0043226). Molecular functions were mainly binding (93 genes; GO:0005488) and catalytic activity (47 genes; GO:0003824). Biological process annotations were mainly cellular process (98 genes; GO:0009987), biological regulation (81 genes; GO:0065007), and metabolic process (75 genes; GO:0008152). In addition, we also identified the biochemical pathways of the PSGs. The KEGG

annotation of the PSGs suggested the presence of 29 pathways related to cellular processes (33 genes), genetic information processing (43 genes), human diseases (10 genes), environmental information processing (29 genes), metabolism (55 genes), and organismal systems (24 genes) (Fig. S2b).

## Population structure

On a global scale, *L. relictus* was considered to exist in four separate breeding subpopulations: the Ordos, Far East, Gobi, and Central Asia populations [17]; however, since the end of last century, no individuals of the Far East, Gobi, and Central Asia breeding subpopulations have been observed and the largest colonies of the Ordos breeding subpopulation successfully migrated from Taolimiao-Alashan Nur on the Ordos Plateau in Inner Mongolia into Hongjian Nur, Shaanxi Province, China, in 2001. Ninety percent of the world's known *L. relictus* are gathered in Hongjian Nur, which could be used as a long-term monitoring site for this species [4].

## Discussion

### Genomic characteristics and evolution

The genome size of *L. relictus* was similar to those of some other Aves, such as *P. pugnax* (1.25 Gb) [18]. The GC content of the *L. relictus* genome was similar to that of other Aves, such as 41.43% in *Tetraothis szechenyii* [19], an average of 42.5% in *P. pugnax* and an average of 41.5% in *G. gallus* [18]. This proportion of repeat sequences is similar to that found in previous studies, in which almost all avian genomes contained lower levels of repeat elements than other animal genomes, with percentages of approximately 4-10% [16].

The timescale results indicated that the ancestral lineages of *L. relictus* and *P. pugnax* diverged approximately 68.44 Mya (million years ago) (Fig. 1), later than the divergence time between the complex clade containing the genus *Larus* and the clade including *Philomachus* [20].

### Population dynamics

*L. relictus* nests were surveyed, and 7,708 and 7,604 nests were recorded in Hongjian Nur in 2010 and 2011, respectively; the species it appears to have reached its breeding peak in 2010 and dropped sharply over the past 10 years to 5,140 nests in 2012, 4,990 nests in 2013, and 4,856 nests in 2014. Unfortunately, only 2000-3000 nests have been recorded during the last two years 2019-2020 (unpublished data). Five new breeding sites of Ordos individuals were found recently: Aotai Lake, a wetland located in the western suburb of Hohhot City in Inner Mongolia, China (40°40'N, 111°23'E) [21], 285 nests; Gouchi Wetland, Dingbian County, Shaanxi Province, China (37°44'N, 107°30'E) [22], 434 nests; Kangbao Noel Lake National Wetland Park, Kangbao County, Zhangjiakou, Hebei Province, China (41°49~50'N, 111°35~36'E) [23], 806 nests; Shuangmaotou Lake, Ningxia Hui Autonomous Region, China (37°43~44'N, 106°46~48'E) [24], 410 nests; Degeduwulan Lake, Taibus Banner, Inner Mongolia, China (41°48'N, 115°46~48'E) [24], 410 nests.

The population dynamics of the Ordos breeding subpopulation in Hongjian Nur provides a good confirmation that these birds are suffering from habitat loss and fragmentation, and interference by human activity is inferred to be the major cause of population fragmentation.

## Conclusions

The whole-genome sequence of *L. relictus* was successfully assembled employing the Illumina and PacBio sequencing platform. The size of the final assembled genome was 1.21 Gb, with a contig N50 of 8.11 Mb and 92.52 (7.63%) Mb of repeat sequence, and 18,454 protein-coding genes were predicted with 16,967 (91.94%) of these genes annotated.

The population of relict gull (*L. relictus*) was expressed very low genetic diversity while lacking a large geographical population. In this study, the genome information of *L. relictus* which is the first assembled reference genome of the genus *Larus*, will be effectively to investigate the evolutionary and molecular mechanisms of some significant processes in this species.

## Methods

### Sampling and sequencing

A naturally dead *L. relictus* fledgling from Hongjian Nur (39°04' N, 109°53' E), Yulin, Shaanxi Province, was collected and identified by H. Xiao, and the specimen (voucher number YG01) was deposited in the animal specimens museum of the Shaanxi Institute of Zoology, Xi'an, Shaanxi Province, China. Our team is a wildlife protection agency under the Xi'an branch of the Chinese Academy of Sciences, cooperating and working with the authority department on Hongjian Nur for nearly 20 years, mainly devoted to the protection of the relict gull. To protect *Larus relictus*, this project has been approved and received permission from the Nature Reserve Authority of Hongjian Nur.

DNA was extracted from the muscle using the CTAB method, and the DNA concentration and quality were measured using a NanoDrop 2000 and a Qubit Fluorometer, respectively. Both Illumina HiSeq 4000 and PacBio RSII sequencing pipelines were used. For the Illumina pipeline, five short fragment paired-end libraries (three of 270 bp and two of 350 bp) were constructed using the standard Illumina pipeline. The details of library construction are as follows. The genomic DNA was broken randomly using the ultrasonic method, and target fragments were filtered. The small fragment sequencing library was constructed through the steps of end repair, addition of polyA and adaptor, selection of target-size fragments and PCR. The size and quality of the library were evaluated using an Agilent 2100 and qPCR. The Illumina HiSeq 4000 sequencer was used for sequencing, with PE=150. For the long fragment library in the PacBio pipeline, the details of library construction are as follows. The genomic DNA was sheared using g-TUBE, followed by DNA damage-repair and end-repair. The dumbbell-type adapters were ligated, and exonuclease digestion was performed. BluePippin was used to select segments to obtain the sequencing library. In addition, total RNA was extracted from the heart, liver, spleen, lung and kidney of *L.*

relictus using TRIzol, and RNA concentrations were measured using NanoDrop 2000, Qubit 2.0 and Agilent 2100. The Illumina HiSeq 4000 platform was used for sequencing RNA data.

## Genome assembly assessment

Whole-transcriptomic data from the liver and an equal mix of five tissue RNA samples were used to assist genome annotations. Raw reads were filtered to remove adapter sequences and low-quality data, with clean reads assembled using Trinity [26]. After filtering out low-quality and short-length PacBio reads, LoRDEC [27] software was used for error correction of PacBio data employing HiSeq data. The HiSeq data were preliminarily assembled by Platanus [28] software. Using dbg2olc [29] software, mixed assembly was carried out by using the data after error correction and the preliminary assembly results of HiSeq data. Pilon [30] software was used to correct the assembly results using HiSeq data. To assess the completeness of the *L. relictus* genome assembly, we used two methods, with the first remapping the Illumina paired-end reads to the assembled genome and the second employing CEGMA v.2.5 [31] and BUSCO v3.0.2 databases.

## Genome Annotation

Three different strategies were used to predict gene structures, namely, ab initio-based, homologue-based and RNA-seq-based methods. EVM v1.1.1 [32] software was used to integrate the predicted genes and generate a consensus gene set. Then, GENSCAN [33], Augustus v2.4 [34], GlimmerHMM v3.0.4 [35], GeneID v1.4 [36] and SNAP (version2006-07-28) [37] were first used to perform the ab initio prediction. For homologue prediction, GeMoMa v1.3.1 [38] was used, primarily employing five species as references, i.e., *G. gallus*, *Meleagris gallopavo*, *Taeniopygia guttata*, *Ficedula albicollis* and *Parus major*. Third, HISAT v2.0.4 and StringTie v1.2.3 [39] were used for assembly based on RNA-seq reference data, and TransDecoder v2.0 [40] and GeneMarkS-T v5.1 [41] were applied to predict genes. PASA v2.0.2 [42] was used to predict unigene sequences assembled based on the whole transcriptome data without references. Finally, EVMv1.1.1 [32] was used to integrate the prediction results obtained by the above three methods, and PASA v2.0.2 [42] was used to predict alternative splice variants.

Software including LTR-FINDERv1.05 [43], MITE-Hunter [44], RepeatScout v1.05 [45] and PILER-DF v2.4 [46] was used for prediction of repetitive sequences in the *L. relictus* genome. A combination of structure-based and de novo strategies was used to construct repeat databases and then merged with Repbase [47] to form a final database. RepeatMasker v4.0.6 was used to identify repeat sequences with this final repeat database [48].

Using the Rfam [49] and miRbase [50] databases as references, rRNA and microRNA were identified by Infernal 1.1 [51]. The tRNA was predicted using tRNAscan-SEv1.3.1 [52]. After screening the true loci of the genome, GenBlastA v1.0.4 [53] was used to blast to search homologous gene sequences. Pseudogenes were then identified via GeneWise v2.4.1 [54] with premature stop codons and frame shifts.

To assign gene functions in the *L. relictus* genome, we aligned the genes to five functional databases using BLASTv2.2.3 [55] (E-value = 1e-5). The databases included GO (Gene Ontology) [56], KEGG (Kyoto Encyclopedia of Genes and Genomes) [57], KOG (Cluster of Orthologous Groups for eukaryotic complete genomes) [58], TrEMBL (Translated EMBL-Bank) [59] and NR (NCBI non-redundant amino acid sequences) [60].

## Phylogeny and positively selected genes

We used the whole-genome sequence of *L. relictus* and seven published whole-genome sequences of three Charadriiformes species (*L. relictus*, *P. pugnax* and *C. vociferus*), three Pelecaniformes (*Phaethon lepturus*, *Nipponia nippon* and *Egretta garzetta*), one Gruiformes (*Mesitornis unicolor*) and one Procellariiformes (*Fulmarus glacialis*). OrthoMCL was used to cluster gene families [61]. A total of 7,339 single-copy orthologues were identified, with protein sequences used for constructing phylogenetic trees. The protein sequences were aligned using MUSCLE [62] and then concatenated into a combined dataset. We then constructed the phylogenetic tree using the ML (maximum likelihood) algorithm with the JTT amino acid substitution model implemented in PhyML [63].

Based on the results of phylogenetic trees, divergence time was estimated using the MCMCTree program in PAML [64]. Divergence times and ages of fossil records were derived from TimeTree (<http://www.timetree.org/>) and applied as the time control. In addition, the CodeML program in PAML [65] included single-copy genes to detect positively selected genes in *the L. relictus* genome. Among them, the branch model (model=2, NSsites=0) was used to calculate  $\omega_0$  (dn/ds) of the foreground clade and obtained the average  $\omega_1$  of other clades. Then, model=0 and NSsites=0 were used to evaluate the  $\omega_2$  of the whole tree, and those genes with  $\omega_0 > \omega_2$  were selected.

## Abbreviations

CEG: Core Eukaryotic Gene; LINE: Long Interspersed Element; LTR: Long Terminal Repeat;

SINE: Short Interspersed Element; PSG: Positively Selected Gene

## Declarations

### Acknowledgements

We thank Tuokao Han for assisting us in collecting specimens and Liliang Lin for assisting us in plotting data.

### Authors' contributions

CY collected the sample, carried out all experiments, and wrote this paper. XJL analyzed sequencing data, and embellished the article. QXW collected the sample, and assisted in the programming. HY analyzed

sequencing data. HX conceived this idea and identified the sample. YH initiated this project and refined it, revised and approved the final manuscript .

## Funding

This work was supported by the National Natural Science Foundation of China (Grant No. 31601846 and 31801993), Natural Science Foundation of Shaanxi Province, China (Grant No. 2017JQ3014 and 2020JM-270), Projects for Department of Science and Technology of Shaanxi Province, China (Grant No. 2018ZDXM-NY-071 and 2019NY-089), Fundamental Research Funds for the Central Universities, China (Grant No. GK201803087 and GK202003052).

## Availability of data and materials

The authors declare that the data supporting the finding of this study are available in the article and its supplementary information files, and are available from the corresponding author.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests

## Author details

<sup>1</sup> College of Life Sciences, Shaanxi Normal University, Xi'an 710062, China

<sup>2</sup> Shaanxi Institute of Zoology, Xi'an 710032, China

## References

1. Auezov EM. Taxonomic evaluation and systematic status of *Larus relictus*. Moscow, J Acad Sci. 1971;50:235-42. (in Russian)
2. Yang C, Lian T, Wang Q, Huang Y, Xiao H. Structural characteristics of the Relict Gull (*Larus relictus*) mitochondrial DNA control region and its comparison to other Laridae. Mitochondrial DNA A DNA. 2016;27(4):2487-91.
3. Liu D, Zhang G, Jiang H, Chen L, Meng D, Lu J. Seasonal dispersal and longitudinal migration in the Relict Gull *Larus relictus* across the Inner-Mongolian Plateau. Peer J, 2017;5:e3380.
4. Yang C, Lian T, Wang Q, Huang Y, Xiao H. Preliminary study of genetic diversity and population structure of the Relict Gull *Larus relictus* (Charadriiformes Laridae) using mitochondrial and nuclear genes. Mitochondrial DNA, 2016,27(6):4246-49.

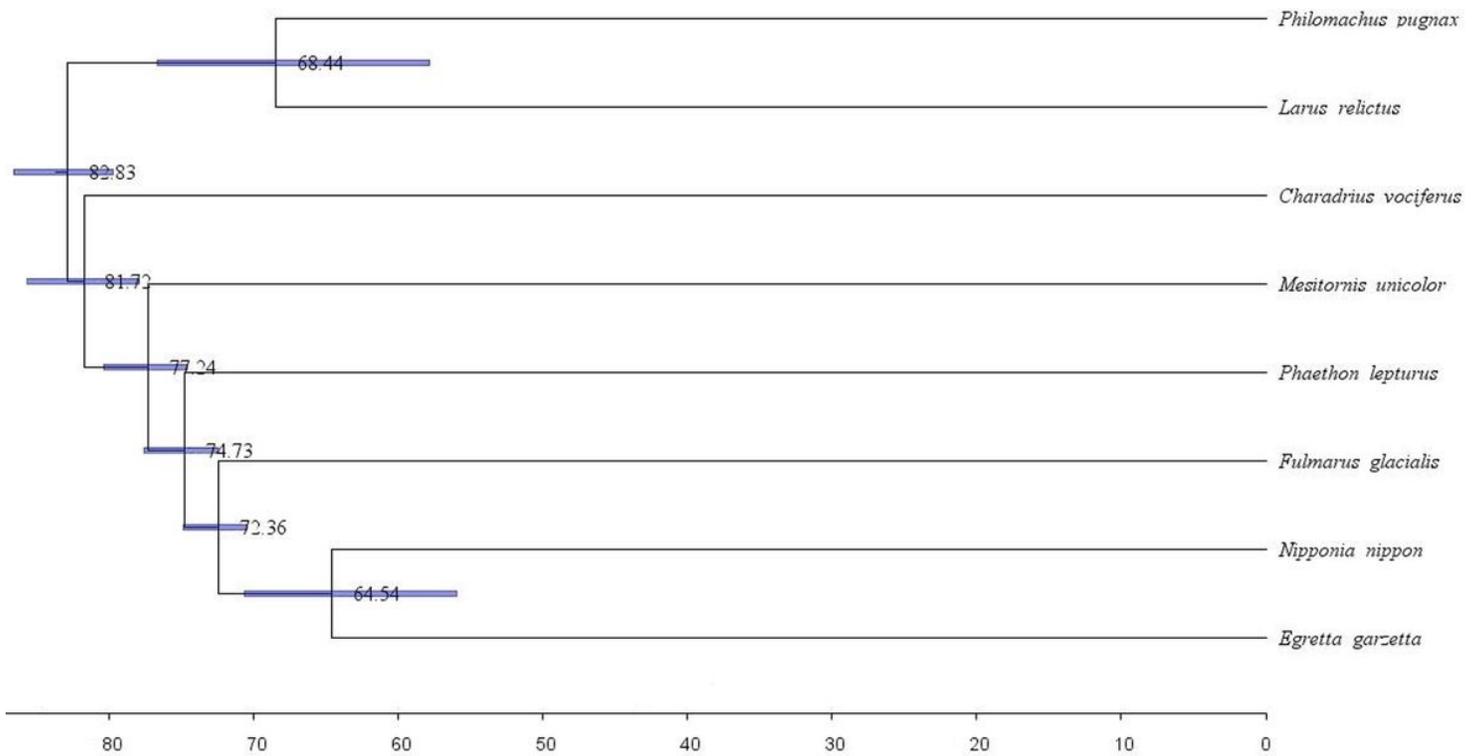
5. Liu Y, Lei JY, Zhang Y, Zhang ZW. The population, distribution and structure of Relict Gull community in Bohai Bay, in: Proceedings of the Eighth National Congress of China Ornithological Society and the Sixth Ornithological Symposium of the Mainland and Taiwan in China. 2005.
6. Yin L, Fei JL, Liu CY. Birds of Hong Kong and South China, eighth ed., Hong Kong Printing Department, Hong Kong. 1994.
7. Ikhumhen HO, Li TX, Lu SL. Assessment of a novel data driven habitat suitability ranking approach for *Larus relictus* specie using remote sensing and GIS. *Ecol Model*, 2020;432:109-221.
8. Zhang L, Li S, Luo J, Du P, Wu L, Li Y, Zhu X, Wang L, Zhang S, Cui J. Chromosome-level genome assembly of the predator *Propylea japonica* to understand its tolerance to insecticides and high temperatures. *Mol Ecol Resour*, 2020;20(1):292-307.
9. Yang C, Wang Q, Huang Y, Xiao H. Complete mitochondrial genome of Relict Gull, *Larus relictus* (Charadriiformes: Laridae). *Mitochondrial DNA*, 2016;27(1):411-2.
10. Kwon YS, Kim JH, Choe JC, Park YC. Low resolution of mitochondrial COI barcodes for identifying species of the genus *Larus* (Charadriiformes: Laridae). *Mitochondrial DNA*, 2012;23(2):157-66.
11. Bian L, Li F, Ge J, Wang P, Chang Q, Zhang S, Li J, Liu C, Liu K, Liu X, et al. Chromosome-level genome assembly of the greenfin horse-faced filefish (*Thamnaconus septentrionalis*) using Oxford Nanopore PromethION sequencing and Hi-C technology. *Mol Ecol Resour*, 2020;20(4):1069-79.
12. Giordano F, Aigrain L, Quail MA, Coupland P, Bonfield J, Davies R, Tischler G, Jackson D, Keane T, Li J, et al. De novo yeast genome assemblies from MinION, PacBio and MiSeq platforms. *Sci Rep*, 2017;7(1):3935.
13. Rhoads A, Au KF. PacBio sequencing and its applications. *Genom Proteom Bioinf*, 2015;13(5):278-89.
14. Warren W, Hillier L, Tomlinson C, Minx P, Kremitzki M, Graves T, Markovic C, Bouk N, Pruitt K, Thibaud-Nissen F, et al. A new chicken genome assembly provides insight into avian genome structure. *G3 (Bethesda)*, 2017;7(1):109-17.
15. Peñalba JV, Deng Y, Fang Q, Joseph L, Moritz C, Cockburn A. Genome of an iconic Australian bird: High-quality assembly and linkage map of the superb fairy-wren (*Malurus cyaneus*). *Mol Ecol Resour*, 2020;20(2):560-78.
16. Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, Storz JF, Antunes A, Greenwold MJ, Meredith RW, et al. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*, 2014;346(6215):1311-20.
17. He FQ, David M, Xing XJ, Ren YQ. A review on studies of the Relict Gull *Larus relictus*. *Chin. J. Zool*, 2002;37:65-70.
18. Lamichhaney S, Fan G, Widemo F, Gunnarsson U, Thalmann DS, Hoepfner MP, Kerje S, Gustafson U, Shi C, Zhang H, et al. Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nat Genet*, 2016;48(1):84-8.
19. Zhou C, James JG, Xu Y, Tu H, He X, Wen Q, Price M, Yang N, Wu Y, Ran J, et al. Genome-wide analysis sheds light on the high-altitude adaptation of the buff-throated partridge (*Tetraophasis*

- szechenyii*). Mol Genet Genomics, 2020;295(1):31-46.
20. Baker AJ, Pereira SL, Paton TA. Phylogenetic relationships and divergence times of Charadriiformes genera: multigene evidence for the Cretaceous origin of at least 14 clades of shorebirds. Biol Lett, 2007;3(2):205-9.
  21. Gao LJ, Hou JH, Dong JX, Chen C, Du MT. New distribution of the Relict Gulls in breeding season, at Aotai lake of Inner Mongolia. Chin J Zool, 2013;48(1):141-2.
  22. Xiao H, Wang KF, Feng N, Xu ZQ. New breeding distribution of Relict Gulls in Gouchi wetland, Dingbian, Shaanxi Province. Chin J Zool, 2013; 48(5):776-7.
  23. Wu Y, Zhang H, Zhang YQ, Deng L, Hong JM. New breeding distribution of the Relict Gull at Kangba Noel National Wetland Park in Zhangjiakou, Hebei Province. Chin J Zool, 2016; 51(4):668-70.
  24. Zhai H, Yu HX, Ma GD, Wei H, Li YL, Feng QD, Shi HN, Duan ZG, Tian R. A new record of Relict Gull in Ningxia and its breeding population and breeding sites. Chin J Wildlife, 2019; 40(1):196-9.
  25. Wu Y, Lu ZY, Liu YD, Cao ZZ, Zhang YQ, Hong JM. New distribution of breeding population of the Relict Gull at Degeduwulan Lake in Taibus Banner. Inner Mongolia. Wetl Sci Manag, 2019; 15:(4)65-6.
  26. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol, 2011;29(7):644-52.
  27. Salmela L, Rivals E. LoRDEC: accurate and efficient long read error correction. Bioinformatics, 2014;30(24):3506-14.
  28. Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, Yabana M, Harada M, Nagayasu E, Maruyama H, et al. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome Res, 2014;24(8):1384-95.
  29. Ye C, Hill C, Wu S, Ruan J, Ma Z. DBG2OLC: Efficient assembly of large genomes using long erroneous reads of the third generation sequencing technologies. Sci Rep, 2016;6:31900.
  30. Chin CS, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A, et al. Phased diploid genome assembly with single-molecule real-time sequencing. Nat Methods, 2016;13(12):1050-54.
  31. Parra G, Bradnam K, Korf I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics, 2007;23(9):1061-67.
  32. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. Automated eukaryotic gene structure annotation using evidence modeler and the program to assemble spliced alignments. Genome Biol, 2008;9(1):R7.
  33. Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. J Mol Biol, 1997;268(1):78-94.
  34. Stanke M, Waack S. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics, 2003;19 Suppl 2:215-25.

35. Majoros WH, Pertea M, Salzberg SL. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics*, 2004;20(16):2878-79.
36. Blanco E, Parra G, Guigó R. Using geneid to identify genes. *Curr Protoc Bioinformatics*, 2007; 4(4.3):1-28.
37. Korf I. Gene finding in novel genomes. *BMC Bioinformatics*, 2004;5:59.
38. Keilwagen J, Wenk M, Erickson JL, Schattat MH, Jan G, Frank H. Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res*, 2016;44(9):e89.
39. Pertea M, Kim D, Pertea GM. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc*, 2016;11(9):1650-67.
40. Haas BJ, Papanicolaou A. TransDecoder (Find Coding Regions Within Transcripts) [EB/OL]. <http://transdecoder.github.io>.
41. Tang S, Lomsadze A, Borodovsky M. Identification of protein coding regions in RNA transcripts. *Nucleic Acids Res*, 2015;43(12):e78.
42. Campbell MA, Haas BJ, Hamilton JP, Mount SM, Buell CR. Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis. *BMC Genomics*, 2006;7:327.
43. Xu Z, Wang H. LTR\_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res*, 2007;35(suppl. 2):W265-8.
44. Han Y, Wessler SR. MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res*, 2010;38(22):e199.
45. Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. *Bioinformatics*, 2005;21(Suppl 1):i351-8.
46. Edgar RC, Myers EW. PILER: Identification and classification of genomic repeats. *Bioinformatics*, 2005;21(Suppl 1):i152-8.
47. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res*, 2005;110(1-4):462-7.
48. Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics*, 2009; Supplement 5 (4.10):1-14.
49. Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, Bateman A. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res*, 2005;33(Database issue):D121-4.
50. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res*, 2006;34(Database issue):D140-4.
51. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, 2013;29(22):2933-5.
52. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*, 1997;25(5):955-64.
53. She R, Chu JSC, Wang K, Pei J, Chen NS. GenBlastA: enabling BLAST to identify homologous gene sequences. *Genome Res*, 2009;19(1):143-9.

54. Birney E, Clamp M, Durbin R. GeneWise and genomewise. *Genome Res.*, 2004;14:988-95.
55. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment searchtool. *J Mol Biol*, 1990;215:403-410.
56. Dimmer EC, Huntley RP, Alam-Faruque Y, Sawford T, O'Donovan C, Martin MJ, Bely B, Browne P, Chan WM, Eberhardt R. The UniProt-GO Annotation database in 2011. *Nucleic Acids Res*, 2012;40(Database issue):D565-70.
57. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, 2000;28(1):27-30.
58. Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova ND, Koonin EV. The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res*, 2001;29(1):22-8.
59. Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res*, 2003;31(1):365-70.
60. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res*, 2011;39(Database issue):D225-9.
61. Li L, Stoeckert CJ, Roos DS. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res*, 2003;13(9):2178-89.
62. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*, 2004;32(5):1792-7.
63. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*, 2010;59(3):307-21.
64. Cummings MP. PAML (Phylogenetic Analysis by Maximum Likelihood) John Wiley & Sons, Inc. 2004.
65. Schabauer H, Valle M, Pacher C, Stockinger H, Stamatakis A, Robinson-Rechavi M, Yang Z, Salamin N. SlimCodeML: An optimized version of CodeML for the branch-site model. In 26th IPDPS Workshops and PhD Forum, 2012;706-14.

## Figures



**Figure 1**

Timing of inferred divergence of *Larus relictus* and seven other Aves. Note: numbers on the nodes represent divergence times, and blue boxes indicate 95% HPD.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaltables.docx](#)
- [figureS1.bmp](#)
- [figureS2.bmp](#)