scientific data

OPEN

DATA DESCRIPTOR

Long-read *de novo* assembly of the red-legged partridge (*Alectoris rufa*) genome

Rayner González-Prendes[1], Ramona Natacha Pena[2,3], Cristóbal Richart[4], Jesús Nadal[2] ✉ & Roger Ros-Freixedes[2,3] ✉

The red-legged partridge (*Alectoris rufa*) is a popular game bird species that is in decline in several regions of southwestern Europe. The introduction of farm-reared individuals of a distinct genetic make-up in hunting reserves can result in genetic swamping of wild populations. Here we present a *de novo* genome assembly for the red-legged partridge based on long-read sequencing technology. The assembled genome size is 1.14 Gb, with scaffold N50 of 37.6 Mb and contig N50 of 29.5 Mb. Our genome is highly contiguous and contains 97.06% of complete avian core genes. Overall, the quality of this genome assembly is equivalent to those available for other close relatives such as the Japanese quail or the chicken. This genome assembly will contribute to the understanding of genetic dynamics of wild populations of red-legged partridges with releases of farm-reared reinforcements and to appropriate management decisions of such populations.

Background & Summary

The red-legged partridge (*Alectoris rufa*) is a popular game bird species from southwestern Europe (France, Spain, Portugal, and northwestern Italy). It can also be found in flat areas of England and Wales, where it was introduced as a game species. This partridge species can survive in a wide range of habitats, favouring dry low grass areas. Wild populations in the Iberian Peninsula are currently in decline[1,2], mainly due to habitat degradation and hunting pressure. To meet the demand of birds for hunting purposes, farm-reared partridges are often released in game reserves.

Red-legged partridges belong to the Phasianidae family of the order Galliformes. Red-legged partridges and chickens (*Gallus gallus*) diverged about 65 million years ago[3]. Despite the relatively long divergence time, the genome synteny and karyotype of both species are highly conserved. The red-legged partridge has $2n = 78$ chromosomes[3] and the chicken has $2n = 80$ in its latest genome assembly (version GRCg7b; GenBank accession: GCA_016699485.1[4]). Most of the chromosomes are small microchromosomes, while only a few macrochromosomes are present in the karyotype[3,5,6]. This conservation in genome structure and organization between the red-legged partridge and the chicken, particularly on the macrochromosomes, is expected to facilitate the identification and characterization of orthologous genes and genomic regions.

The first red-legged partridge genome assembly[7], published in 2021, was based on short-read sequencing data, and by current standards it could be considered of draft level with an N50 equal to 11.5 Mb[7]. The assembly included linkage data based on a low-density genetic map and the placement of scaffolds to chromosomes relied considerably on conserved synteny assumptions with the better assembled reference genome of the chicken. However, avian microchromosomes have proved to be difficult to assemble even with current technology.

An improved genome assembly is crucial for better understanding genomic variation in wildlife populations and commercial stocks. Improved annotation of coding and non-coding genes benefits the functional interpretation of genome-wide scans for selection sweeps and genome-wide association studies, as well as aiding in the identification of genetic markers for assisting in selection of commercial stocks or gene editing targets. In wild populations, an improved genome assembly can also help in capturing population-specific variation that could be used in the design of conservation programs. So far, in the Galliformes, the genome assemblies of the

[1]Animal Breeding and Genomics, Wageningen University & Research, 6708PB, Wageningen, The Netherlands. [2]Departament de Ciència Animal, Universitat de Lleida, Lleida, Spain. [3]Agrotecnio-CERCA Center, Lleida, Spain. [4]Departament de Medicina i Cirurgia, Universitat Rovira i Virgili, Tarragona, Spain. ✉e-mail: jesus.nadal@udl.cat; roger.ros@udl.cat

chicken[8,9], Japanese quail (*Coturnix japonica*)[10,11], Gunnison sage-grouse[12], helmeted guineafowl[13], and turkey[14] are high-quality genome assemblies based on long-read sequencing technology. These high-quality assemblies allow for comparative studies within the Galliformes and raise the opportunities for further research.

In this study, we aimed to develop a new *de novo* assembly for the red-legged partridge species based on a combination of long- and short-read sequencing technologies. This new assembly will provide an improved reference genome that will be a useful resource for future research and practical applications, including conservation efforts.

Methods

Ethics statement. Sample collection was conducted in full compliance with Spanish laws and regulations, including the licence of "Las Ensanchas" for sampling shot partridges. The protocol was approved by the Committee on the Ethics of Animal Experiments of the University of Lleida (Ref. 1998-2012/05).

Sample collection and construction of sequencing libraries. Muscle samples for whole-genome sequencing were collected from 60 red-legged partridges, divided in two subpopulations. Half of the samples (n = 30) were from a population of wild partridges from a private game hunting estate in central Spain ("Las Ensanchas", Ciudad Real, Spain) and the other half (n = 30) were from a partridge rearing farm. These 60 samples were sequenced in an Illumina (San Diego, CA, USA) short-read platform, as described below. More detailed information about these subpopulations and an analyses of the genomic diversity and signals of selection processes based on these samples was provided by Ros-Freixedes *et al.*[15]. In addition, blood samples from two additional red-legged partridges, one from each subpopulation, were used for long-read sequencing in an Oxford Nanopore Technologies (Oxford, UK) platform. The individual from the wild subpopulation was a female captured approximately 60 km away from the location of the other 30 wild individuals. The individual from the farm-reared population was a male from a different farm to the other 30 farm-reared individuals.

Total genomic DNA was isolated using standard protocols consisting of proteinase K lysis and phenol-chloroform purification[16]. The DNA for the long-read sequencing experiment was isolated using wide-bore filter tips and the preparations were never frozen before sequencing to preserve the integrity of long chains of DNA. The purified DNA was quantified using a Qubit fluorometer (Invitrogen, Waltham, MA USA) and quality checked using a NanoDrop Spectrophotometer (ThermoScientific, Wilmington, DE, USA) and agarose gel electrophoresis.

The short-insert paired-end libraries for whole-genome sequencing were prepared with a PCR-free protocol using KAPA HyperPrep kit as detailed before[15]. The 60 libraries were sequenced on NovaSeq 6000 (Illumina) in paired-end mode with a read length of $2 \times 151 + 17 + 8$ bp following the manufacturer's protocol for dual indexing. Image analysis, base calling and quality scoring of the run were processed using the manufacturer's software Real Time Analysis (RTA 3.4.4, Illumina) and followed by generation of FASTQ sequence files. For each sample, a minimum of 20 Gb of sequencing data was generated, which represents a sequencing depth of ~20x of the partridge genome.

For the long-read sequencing experiment, one sequencing library was prepared from a pool of equal amounts of the two DNA samples. The library was built using the Ligation sequencing kit SQK-LSK109 from Oxford Nanopore Technologies. Briefly, 4.0 μg of the pool DNA was repaired and end-repaired using NEBNext FFPE DNA Repair Mix (New England BioLabs, Ipswich, MA, USA) and the NEBNext UltraII End Repair/dA-Tailing Module (New England BioLabs) and followed by the sequencing adaptors ligation, purified by 0.4X AMPure XP Beads and eluted in Elution Buffer (SQK-LSK109). The sequencing run was performed on GridION Mk1 (Oxford Nanopore Technologies) using a Flowcell R9.4.1 FLO-MIN106D (Oxford Nanopore Technologies) and the sequencing data was collected for 90 h. The quality parameters of the sequencing run were monitored by the MinKNOW platform version 3.6.5 in real time and basecalled with Guppy version 3.2.10.

Genome assembly. For the genome *de novo* assembly we used the long-read sequencing data of the DNA pool as well as short-read sequencing data of one individual. The latter was a female from the farm-reared subpopulation and was selected randomly among the 60 individuals with available short-read sequencing data. We used Flye[17] for initial assembly and NextPolish[18] for the polishing step. This approach allowed us to combine the benefits of both long-read and short-read sequencing technologies to produce a high-quality assembly of the red-legged partridge genome. LRScaf[19] software was then utilized to identify and close gaps in the assembly. The completeness of the final assembly in terms of gene space was evaluated using BUSCO[20] v4.1.2, which was run in genome mode (-m genome) with the vertebrae (vertebrata_odb10) and avian (aves_odb10) datasets. The final draft assembly was aligned against the quail and chicken reference genomes using the default parameters in D-genie (https://dgenies.toulouse.inra.fr) using the Minimap2[21,22] v2.26 aligner.

Repeats annotation. We generated a *de novo* repeat library using the BuildDatabase tool from RepeatModeler[23] v1.0.11. RepeatMasker[24] v4.0.7 and a custom-built repeat library from RepeatModeler were used to comprehensively identify and annotate the repetitive elements in the genome. The custom RepeatModeler library and custom R scripts were used to investigate the differences in repeat content between the scaffolds. Each scaffold was split into bins (each bin corresponding to 2% of the scaffold length), allowing us to compare between the scaffolds by relative length.

Gene prediction and annotation. The genome is currently being annotated using the ENSEMBL annotation pipeline, which is available as part of the Ensembl Rapid Release[25].

Alignment metrics. We assessed the quality of the alignment using the short reads from the 59 sequenced red-legged partridges (excluding the one used for genome *de novo* assembly) to our reference genome. To do

Features	Red-legged partridge		Quail	Chicken	
	Arufa2	Arufa1	Cjaponica v2.1	GRCg6a	GRCg7b
Total sequence length (bp)	1,142,509,937	1,027,480,606	927,656,957	1,065,348,650	1,053,332,251
Total ungapped length (bp)	1,141,765,707	996,170,856	917,263,224	1,055,564,190	1,049,948,333
Gaps between scaffolds	0	0	519	68	0
No. of scaffolds	426	10,598	2,531	524	214
Scaffold N50 (bp)	37,566,138	11,577,318	2,975,000	20,785,086	90,861,225
Scaffold L50	9	25	86	12	4
No. of contigs	546	29,132	9,642	1,402	677
Contig N50 (bp)	29,494,169	97,167	511,217	17,655,422	18,834,961
Contig L50	11	2,907	515	19	18
Total number of chromosomes	—	—	32	34	42
GC (%)	42	41	41	42	42
Genome coverage depth (x)	60	96	73	82	102

Table 1. Assembly statistics of red-legged partridge, quail and chicken genomes* using Nanopore long reads and Illumina short reads. *Arufa2: GCA_947331505.1[37]; Arufa1: GCA_019345075.1[27]; Cjaponica v2.1: GCA_001577835.2[28]; GRCg6a: GCA_000002315.5[29]; and GRCg7b: GCA_016699485.1[4].

so, DNA sequence short reads were pre-processed using Trimmomatic[26] to remove adapter sequences from the reads. Then, we mapped the reads to our reference genome for *A. rufa* (Arufa2) as well as other already available reference genomes. The other reference genomes tested were a previously published reference genome for *A. rufa* (Arufa1[7]; GenBank accession: GCA_019345075.1[27]) and those of two close poultry species: the Japanese quail (version 2.1; GenBank accession: GCA_001577835.2[28]) and the chicken (version GRCg6a; GenBank accession: GCA_000002315.5[29]). Additionally, we used data from 28 sequenced chicken mapped to the chicken reference genome as a control. To map the reads we used the BWA-MEM[30] algorithm. Duplicates were marked with Picard (http://broadinstitute.github.io/picard). Alignment metrics were extracted using Picard.

Genetic variation. We benchmarked our new Arufa2 reference genome by comparing to Arufa1 and the quail reference genome. To do so, single nucleotide polymorphisms (SNPs) and short insertions and deletions (indels) were identified with the variant caller GATK HaplotypeCaller[31,32] (GATK 3.8.0) using default settings with all three reference genomes. Variant discovery with GATK HaplotypeCaller was performed separately for each individual and then a variant set for the whole population was obtained by jointly genotyping all individuals at the variant sites with GATK GenotypeGVCFs. We retained all biallelic variants to analyse the transitions-to-transversions ratio (Ti/Tv) in order to avoid biases caused by the alteration of the allele frequency spectrum. This parameter was calculated with VCFtools[33]. To enable comparisons, only data from scaffolds that aligned to chromosomes 1 to 28 of the quail reference genome were considered for Arufa2 and Arufa1.

Data Records

The short-read sequencing data are available at the NCBI Sequence Read Archive (SRA) with BioProject accession PRJNA824288[34]. The long-read sequencing data are also available at the NCBI SRA with BioProject accession PRJNA1070845[35]. The short-read sequencing data that was used for the genome assembly together with the long-read sequencing data is available at the NCBI SRA with BioSample accession SAMN31710140[36]. The assembled draft genome of red-legged partridge is available at NCBI GenBank under the accession number of GCA_947331505.1[37]. The sequence variants called using our draft genome assembly as a reference are available at the EMBL-EBI European Variation Archive (EVA) with accession number ERP161362[38].

Technical Validation

Quality assessment of the genome assembly. The genome of the red-legged partridge was assembled using long-read data generated in a GridIon Nanopore platform and short-read data from the Illumina NovaSeq 6000 sequencing system. Approximately three million long reads were produced with an N50 read length of 40,470 bp and a mean read length of 20,672 bp. Long reads were assembled and polished with short reads with the Flye[17] assembler, resulting in a 1.14 Gb genome draft size with 546 contigs (Table 1). Of the 546 contigs, 223 had a length equal or superior to 25 Mb and 168 to 50 Mb. The longest contig had a length of 111.5 Mb. The draft was scaffolded using LRScaf[19] to produce a final assembly consisting of 426 scaffolds with a scaffold N50 of 37.6 Mb. The number of unassigned bases was 2,039.

To evaluate the completeness and accuracy of the successive assembly versions, we used BUSCO[20] and a whole-genome alignments approach. BUSCO results showed that all four assembly versions had over 96% of the expected vertebrate gene sets (Fig. 1). The four assembly versions that were tested included assembly produced with alternative software tools wtdbg2[39] or Flye[17], followed or not by a polishing step[18]. Among these assemblies, the assembly produced with Flye software and a polishing step emerged as the most comprehensive, boasting the highest number of complete BUSCOs coupled with a relatively low missing count (3,231 complete, 41 fragmented, and 82 missing BUSCOs).

The proposed assembly, generated with Flye[17] and polished, had the best BUSCO scores and, therefore, it was evaluated using the avian gene set. The completeness assessment results revealed that, out of 8,338 total avian

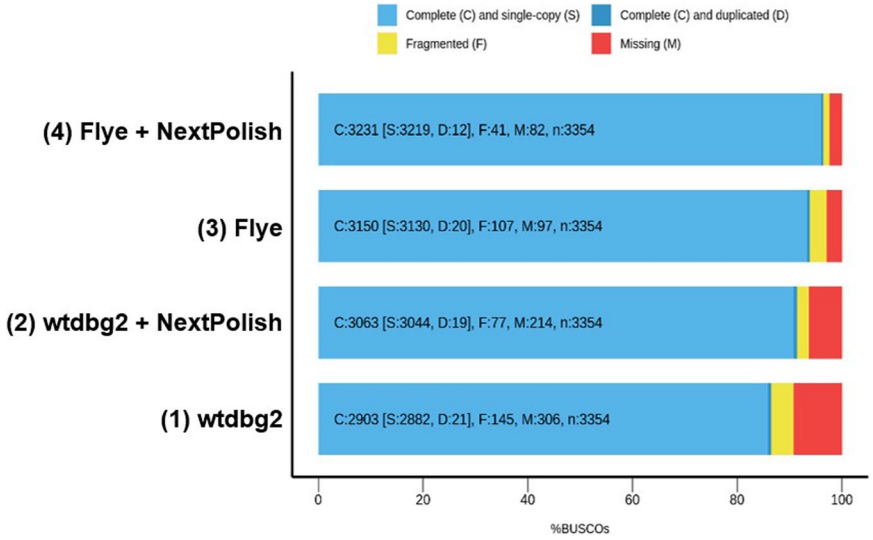

Fig. 1 Assembly completeness measured in BUSCO scores. Percentage of aligned genes (x-axis) for the vertebrae (n = 3,354) gene set in four red-legged partridge Arufa2 assembly versions. The percentages of complete (C) orthologs split between and single-copy (S) and duplicated (D) orthologs, fragmented (F) orthologs, and missing (M) orthologs are shown. Each bar (y-axis) represents a genome draft version for the Arufa2 genome: produced with alternative software tools wtdbg2, without (1) or with (2) a polishing step, and Flye, without (3) and with (4) a polishing step.

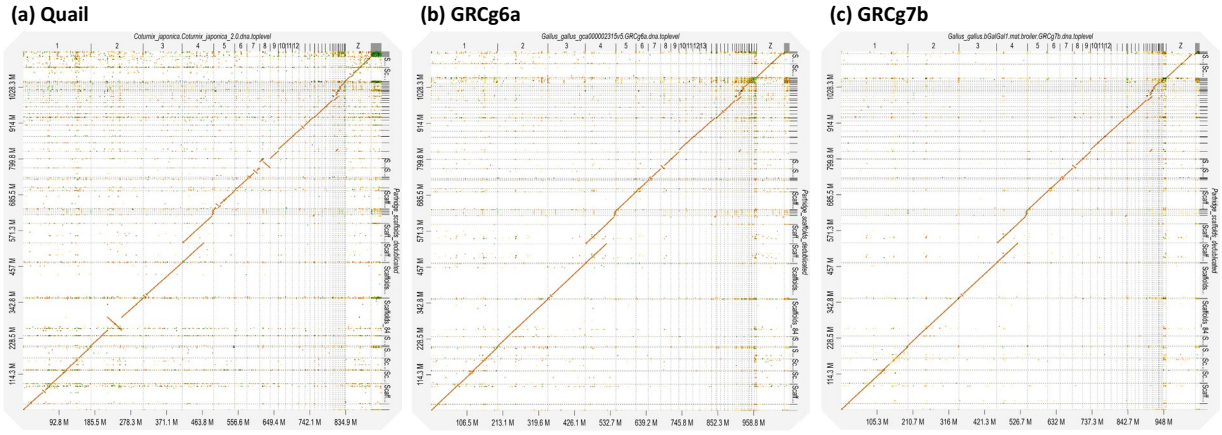

Fig. 2 Genome alignment plots. Quail (**a**) and chicken, GRCg6a (**b**) and GRCg7b (**c**), genomes aligned with the red-legged partridge genome build. Alignment shows high structural coherence between genomes. The y-axis represents segments of our genome assembly in megabases (Mb) and the x-axis represents respective aligned segments in the counterpart genome. Orange dots indicate regions of alignment, with the diagonal line suggesting conserved genomic regions. Chromosomes are represented in the top horizontal bar, labelled from 1 to Z. These labels correspond to individual chromosomes of the *C. japonica* or *G. gallus* genomes. The side vertical bar provides a scale indicating the percentage of aligned scaffolds for each chromosomal segment.

core genes queried, 8,093 (97.06%) were detected as complete, while 8,134 (97.55%) were identified as either complete or partial. A total of 204 core genes were missing, representing 2.45% of the entire set. The average number of orthologs per core gene was 1.00, and only 0.38% of the detected core genes had more than one ortholog. These findings suggest that our genome is highly contiguous and, in terms of its expected gene content, near completion. Indeed, this assembly improved the quality of the first red-legged partridge genome assembly[7], which identified 94.9% of single-copy avian orthologs (7,913 single-copy orthologs out of 8,338 proteins). The gene completeness assessment of our proposed assembly for *A. rufa* has higher values compared with prior findings including the chicken genome GRCg6a (95.4%[40]) and in the quail genome[11], yet slightly inferior to the latest chicken genome GRCg7b, which boasts a completeness of 99.1%.

Genome alignment comparison. The alignment comparing Arufa2 (Fig. 2) with the quail and chicken reference genomes reveals pronounced structural coherence. This is evidenced by a distinct correlation along the diagonal. Few regions exhibit misalignments, indicative of structural variations, inversions, or genomic

Parameter*	Red-legged partridge reads mapped to:				Chicken reads mapped to GRCg6a
	Arufa2	Arufa1	Cjaponica v2.1	GRCg6a	
Before marking duplicates					
Mean coverage depth (x)	22.3	24.8	21.4	20.2	22.8
SD coverage depth (x)	67.0	139.4	91.5	69.3	63.1
Het_Sensitivity	0.984	0.996	0.909	0.877	0.975
Het_Q	18.2	23.8	10.4	9.0	16.0
After marking duplicates					
Mean coverage depth (x)	17.7	21.7	18.3	17.6	19.4
SD coverage depth (x)	11.3	9.5	11.0	10.5	8.8
Exc_MapQ (%)	13.9	3.3	6.0	5.5	5.8
Exc_Dupe (%)	5.0	6.0	5.7	5.5	7.5
Exc_BaseQ (%)	1.3	1.4	1.3	1.3	1.4
Exc_Overlap (%)	14.2	15.9	15.0	15.0	13.8
Exc_Capped (%)	0.7	1.7	1.3	1.0	0.6
Exc_Total (%)	35.1	28.3	29.3	28.3	29.1
Het_Sensitivity	0.813	0.988	0.893	0.860	0.958
Het_Q	7.0	19.3	9.8	8.6	13.9

Table 2. Alignment metrics of 59 red-legged partridges sequenced with short reads against several reference genomes. *Mean coverage depth: mean coverage in bases of the genome territory, after all filters are applied. SD coverage depth: standard deviation of coverage of the genome after all filters are applied. Exc_XXX: fraction of aligned bases that were filtered out because they were in reads with low mapping quality (score < 20; Exc_MapQ), because they were in reads marked as duplicates (Exc_Dupe), because they were of low base quality (score < 20; Exc_BaseQ), because they were the second observation from an insert with overlapping reads (Exc_Overlap), because they would have raised coverage above the capped value (250x; Exc_Capped), or because all filters (Exc_Total). Het_Sensibility: theoretical sensitivity for all heterozygous variants. Het_Q: Phred-scaled Q score of the theoretical sensitivity for all heterozygous variants.

rearrangements. Some isolated regions show denser clustering of alignments, hinting at areas of conservation or duplication. Sporadic off-diagonal alignments may suggest translocations or genomic modifications. Overall, most chromosomes or segments present consistent alignment, which underscores the strong genomic synteny between the genomes. In essence, there is substantial genomic similarity and structural coherence.

To annotate repeats in the genome assembly, we used RepeatModeler[23] to build a custom repeat library. Our analysis showed that repeats covered 10.45% of the genome, with the most common ones being LINE elements, which covered 6.35% of the genome. DNA transposons accounted for 0.76% of the bases, while LTRs and low complexity/simple repeats covered 0.53% and 1.58%, respectively. The remaining 1.23% of repeats were unclassified. These results provide insights into the repeat landscape of the red-legged partridge genome.

Alignment metrics. Alignment metrics for the tested reference genomes are provided in Table 2. The value of long-read sequencing technologies is evident in the improved coherence of genome assemblies. However, there were a larger proportion of reads with low mapping quality in the scaffolded version of Arufa2 (13.9% excluded reads due to mapping quality scores) than in the non-scaffolded version (5.7%), due to a duplication that spans part of two scaffolds and passed duplication filtering. This duplication was either not present or present in a collapsed form in Arufa1 and in the quail and chicken reference genomes. It cannot be ruled out that this duplication is an artefact due to the use of long-read sequence data from pooled DNA of two individuals[41]. Long-read sequencing presents a distinct advantage in structural variant detection. These reads span larger genomic regions, allowing for the accurate identification of insertions, deletions, inversions, and translocations that might be overlooked or inaccurately represented with short-read technologies[42,43].

Genetic variation. The Ti/Tv ratio was 2.50 with Arufa2, 2.54 with Arufa1, and 2.35 with Cjaponica. The Ti/Tv ratio was close to the generally expected values for whole-genome sequencing assuming spontaneous and neutral mutations[44] and, in particular, almost identical to the value of 2.53 reported in chicken[45]. The number of biallelic called variants was 16,210,910, 17,659,943, and 69,609,618 with Arufa2, Arufa1, and Cjaponica, respectively. The SNP/indel ratio was 7.8 for both Arufa2 and Arufa1 and 5.8 for Cjaponica. However, most variants called with Cjaponica were very rare. A total of 55% of the variants called with Cjaponica were singletons (i.e., the minor allele was observed only once) compared to 27% with Arufa2 and Arufa1. The number (and SNP/indel ratio) of variants with MAF $\geq$ 0.05 was only of 6.494.820 (8.5), 7.037.619 (8.4), and 19,631,161 (4.8) with Arufa2, Arufa1, and Cjaponica, respectively. Using a reference genome from a different species can result in an inflation of the number of called variants, as well as an enrichment of indels. These results are indicative of higher false positive rates when using a reference genome from a different species because indel calls are typically considered less reliable than SNP calls. In turn, this impacts the estimation of the other population genetic parameters.

Our findings underscore the importance of possessing a species-specific reference genome, such as this one for the red-legged partridge. By using a dedicated reference, we can enhance the precision with which we

estimate genetic population parameters. Such accurate estimations are instrumental in both characterizing and aiding conservation initiatives for wild populations of this species. Additionally, our reference genome presents a more refined scaffolding relative to previously available versions. Coupled with gene annotation, this opens the door to a range of other applications, notably the detection of selective sweeps.

Code availability

The software versions and configurations employed are described below, following the steps of the pipeline available at https://github.com/CarolinaPB/nanopore-assembly with custom modifications:

(1) Flye:–nano-raw raw.reads.fa.gz --out-dir draft_genome.fa --threads 16 and with wtdbg2 -x ont -g 4.6 m -i raw.reads.fa.gz -t 16 -fo dbg
(2) NextPolish: -genome draft_genome.fa -polish_type sr
(3) LRScaf: -c draft_genome.fa -l long_reads.fq -o scaffolded_genome.fa
(4) To retain a single copy of scaffolds and merge duplicated names, a Python script was utilized: https://github.com/CarolinaPB/Bioinfo_scripts/blob/main/remove_duplicates_fasta.py
(5) BUSCO, version v4.1.2 was first run with the vertebrate database using the command: -m genome -l vertebrae (vertebrata_odb10) -i assembly.fa -o busco_output. It was then run again with the Avian orthologs gene set (aves_odb10).
(6) RepeatModeler, version v1.0.11: BuildDatabase -name custom_repeat_db genome.fa
(7) RepeatMasker, version v4.0.7: -lib custom_repeat_db genome.fa
(8) ENSEMBL annotation pipeline: Uploaded the genome to ENSEMBL's servers
(9) The final draft assembly was employed as input for aligning with genomes from Fig. 2 using the default parameters in D-genie (https://dgenies.toulouse.inra.fr) using the Minimap2 v2.26 aligner.

Received: 30 January 2024; Accepted: 18 July 2024;
Published online: 22 August 2024

References

1. Guzmán, J. L., Viñuela, J., Carranza, J., Porras, J. T. & Arroyo, B. Red-legged partridge Alectoris rufa productivity in relation to weather, land use, and releases of farm-reared birds. *Eur J Wildl Res* **66**, 87 (2020).
2. Cabodevilla, X., Estrada, A., Mougeot, F., Jimenez, J. & Arroyo, B. Farmland composition and farming practices explain spatio-temporal variations in red-legged partridge density in central Spain. *Science of The Total Environment* **799**, 149406 (2021).
3. Griffin, D. K., Robertson, L. B. W., Tempest, H. G. & Skinner, B. M. The evolution of the avian genome as revealed by comparative molecular cytogenetics. *Cytogenet Genome Res* **117**, 64–77 (2007).
4. *NCBI GenBank*. https://identifiers.org/ncbi/insdc.gca:GCA_016699485.1 (2021).
5. Kasai, F., Garcia, C., Arruga, M. V. & Ferguson-Smith, M. A. Chromosome homology between chicken *(Gallus gallus domesticus)* and the red-legged partridge *(Alectoris rufa)*; evidence of the occurrence of a neocentromere during evolution. *Cytogenet Genome Res* **102**, 326–330 (2003).
6. Ouchia-Benissad, S. & Ladjali-Mohammedi, K. Banding cytogenetics of the Barbary partridge Alectoris barbara and the Chukar partridge Alectoris chukar (Phasianidae): a large conservation with Domestic fowl Gallus domesticus revealed by high resolution chromosomes. *CCG* **12**, 171–199 (2018).
7. Chattopadhyay, B. *et al*. Novel genome reveals susceptibility of popular gamebird, the red-legged partridge (Alectoris rufa, Phasianidae), to climate change. *Genomics* **113**, 3430–3438 (2021).
8. International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695–716 (2004).
9. Huang, Z. *et al*. Evolutionary analysis of a complete chicken genome. *Proc. Natl. Acad. Sci. USA* **120**, e2216641120 (2023).
10. Kawahara-Miki, R. *et al*. Next-generation sequencing reveals genomic features in the Japanese quail. *Genomics* **101**, 345–353 (2013).
11. Morris, K. M. *et al*. The quail genome: insights into social behaviour, seasonal biology and infectious disease response. *BMC Biol* **18**, 14 (2020).
12. Oh, K. P., Aldridge, C. L., Forbey, J. S., Dadabay, C. Y. & Oyler-McCance, S. J. Conservation Genomics in the Sagebrush Sea: Population Divergence, Demographic History, and Local Adaptation in Sage-Grouse (Centrocercus spp.). *Genome Biology and Evolution* **11**, 2023–2034 (2019).
13. Shen, Q.-K. *et al*. Genomic Analyses Unveil Helmeted Guinea Fowl (*Numida meleagris*) Domestication in West Africa. *Genome Biology and Evolution* **13**, evab090 (2021).
14. Barros, C. P. *et al*. A new haplotype-resolved turkey genome to enable turkey genetics and genomics research. *GigaScience* **12**, giad051 (2022).
15. Ros-Freixedes, R., Pena, R. N., Richart, C. & Nadal, J. Genomic diversity and signals of selection processes in wild and farm-reared red-legged partridges (Alectoris rufa). *Genomics* **115**, 110591 (2023).
16. Green, M. R. & Sambrook, J. Isolation of High-Molecular-Weight DNA Using Organic Solvents. *Cold Spring Harb Protoc* **2017**, pdb.prot093450 (2017).
17. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* **37**, 540–546 (2019).
18. Hu, J., Fan, J., Sun, Z. & Liu, S. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* **36**, 2253–2255 (2020).
19. Qin, M. *et al*. LRScaf: improving draft genomes using long noisy reads. *BMC Genomics* **20**, 955 (2019).
20. Seppey, M., Manni, M. & Zdobnov, E. M. BUSCO: Assessing Genome Assembly and Annotation Completeness. in *Gene Prediction* (ed. Kollmar, M.) **vol. 1962** 227–245 (Springer New York, New York, NY, 2019).
21. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
22. Li, H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* **37**, 4572–4574 (2021).
23. Flynn, J. M. *et al*. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. USA* **117**, 9451–9457 (2020).
24. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. *CP in Bioinformatics* **25**, (2009).
25. Yates, A. D. *et al*. Ensembl 2020. *Nucleic Acids Research* gkz966 https://doi.org/10.1093/nar/gkz966 (2019).
26. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
27. *NCBI GenBank*. https://identifiers.org/ncbi/insdc.gca:GCA_019345075.1 (2021).
28. *NCBI GenBank*. https://identifiers.org/ncbi/insdc.gca:GCA_001577835.2 (2020).

29. *NCBI GenBank*. https://identifiers.org/ncbi/insdc.gca:GCA_000002315.5 (2018).
30. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at http://arxiv.org/abs/1303.3997 (2013).
31. DePristo, M. A. *et al*. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**, 491–498 (2011).
32. Poplin, R. *et al*. *Scaling Accurate Genetic Variant Discovery to Tens of Thousands of Samples*. http://biorxiv.org/lookup/doi/10.1101/201178 10.1101/201178 (2017).
33. Danecek, P. *et al*. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
34. *NCBI Sequence Read Archive*. https://identifiers.org/ncbi/insdc.sra:SRP408849 (2023).
35. *NCBI Sequence Read Archive*. https://identifiers.org/ncbi/insdc.sra:SRP486622 (2024).
36. *NCBI Sequence Read Archive*. https://identifiers.org/ncbi/insdc.sra:SRS15813874 (2023).
37. *NCBI GenBank*. https://identifiers.org/ncbi/insdc.gca:GCA_947331505.1 (2023).
38. *EMBL-EBI European Variation Archive*. https://identifiers.org/ena.embl:ERP161362 (2024).
39. Ruan, J. & Li, H. Fast and accurate long-read assembly with wtdbg2. *Nat Methods* **17**, 155–158 (2020).
40. Li, M. *et al*. De Novo Assembly of 20 Chicken Genomes Reveals the Undetectable Phenomenon for Thousands of Core Genes on Microchromosomes and Subtelomeric Regions. *Molecular Biology and Evolution* **39**, msac066 (2022).
41. Goldberg, J. K., Allan, C. W., Copetti, D., Matzkin, L. M. & Bronstein, J. A pooled-sample draft genome assembly provides insights into host plant-specific transcriptional responses of a Solanaceae-specializing pest, *Tupiocoris notatus* (Hemiptera: Miridae). *Ecology and Evolution* **14**, e10979 (2024).
42. Chaisson, M. J. P. *et al*. Multi-platform discovery of haplotype-resolved structural variation in human genomes. *Nat Commun* **10**, 1784 (2019).
43. Chen, Y. *et al*. Deciphering the exact breakpoints of structural variations using long sequencing reads with DeBreak. *Nat Commun* **14**, 283 (2023).
44. Wang, J., Raskin, L., Samuels, D. C., Shyr, Y. & Guo, Y. Genome measures used for quality control are dependent on gene function and ancestry. *Bioinformatics* **31**, 318–323 (2015).
45. Derks, M. F. L. *et al*. A survey of functional genomic variation in domesticated chickens. *Genet Sel Evol* **50**, 17 (2018).

Acknowledgements

We are grateful for the contributions made by the Melgarejo family, Patricia, Luis and Ivan Maldonado and Tom Gullick. Thanks also to the "Las Ensanchas" staff, especially the game keepers, the Barranquero family and collaborators, the members of the Tom Gullick hunting team in Campo de Montiel and around the world, Federación de Caza de Castilla y León, Delegación Burgalesa, MUTUASPORT, and Real Federación Española de Caza (RFEC). Carolina Ponz helped in sampling. Fundació Universitat Rovira i Virgili funded the sequencing (grant no. 2060-398-454-455; Proyecto IT20041-S; C. R.).

Author contributions

R. González-Prendes: Methodology, Software, Validation, Investigation, Resources, Data curation, Writing – original draft, Writing – review and editing. R.N. Pena: Conceptualization, Methodology, Investigation, Resources, Writing – original draft, Writing – review and editing. C. Richart: Funding acquisition. J. Nadal: Conceptualization, Resources, Writing – review and editing, Funding acquisition. R. Ros-Freixedes: Conceptualization, Methodology, Software, Validation, Investigation, Resources, Data curation, Writing – original draft, Writing – review and editing.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.N. or R.R.-F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.