Genomics Data 11 (2017) 132–134

Contents lists available at ScienceDirect

Genomics Data

journal homepage: www.elsevier.com/locate/gdata

Data in Brief

Red-legged partridge (*Alectoris rufa*) *de-novo* transcriptome assembly and identification of gene-related markers

Natalia Sevane *, Javier Cañon, Paulina G. Eusebi, Ignacio Gil, Susana Dunner

Departamento de Producción Animal, Facultad de Veterinaria, Universidad Complutense, Madrid 28040, Spain

ARTICLE INFO

Article history:
Received 22 January 2017
Accepted 3 February 2017
Available online 5 February 2017

Keywords:
RNA-seq
Transcriptomic signatures
Immune response
SNPs
Microsatellites

ABSTRACT

The red-legged partridge (*Alectoris rufa*) has a great socio-economic importance as a game species and is reared by millions in farms in several European countries. The ability to respond to a wide spectrum of pathogens and environmental changes is key for farm-reared animals that, as such, face even higher pathogen exposure and specifically for those submitted to restocking programs. In this study, RNA-sequencing and *de-novo* assembly of genes expressed in different immune tissues were performed. The raw FASTQ files were submitted to the NCBI SRA database with accession number PRJNA289204. A total of 94.2 million reads were obtained and assembled into 51,403 contigs using OASES software. The final annotated partridge immune transcriptome comprises almost 7000 unigenes, available as FASTA in the supplementary material. A total of 12,828 microsatellites and 33,857 Single Nucleotide Polymorphisms (SNPs) were identified. The candidate gene sequences and the large number of potential genetic markers from the red-legged partridge transcriptome reliably identified through the use for the first time of a high coverage 100-bp paired-end RNA-seq protocol, provide new tools for future studies in this and related species, thus contributing to the ongoing development of genomic resources in avian species. Further investigation into candidate genes and gene-associated markers will help to uncover individual variability in the resistance to infections and other external aggressions in partridges.

Specifications	
Organism/cell line/tissue	*Alectoris rufa*
Sex	Males and females
Sequencer or array type	Genome Analyzer GAIIx (Illumina)
Data format	Raw data (FASTQ); analyzed data (FASTA)
Experimental factors	*De-novo* transcriptome assembly of *Alectoris rufa*
Experimental features	Non-infectious challenges to elicit both innate and acquired immune responses, sixteen RNA samples from spleen, bursa of Fabricius, thymus and skin sequenced, *de-novo* transcriptome assembly and annotation
Consent	N/A
Sample source location	Burgos, Spain

1. Direct link to deposited data

The raw data obtained in this study is deposited in the NCBI SRA database with accession number PRJNA289204 (https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA289204). The *de-novo* assembled and annotated *Alectoris rufa* transcriptome is available as supplementary material.

2. Introduction

The red-legged partridge (*Alectoris rufa*), whose natural range stretches southwestern Europe, including the Iberian Peninsula, France and Italy, belongs to the order *Galliformes* and the family *Phasianidae*. Recent socio-economic and environmental changes, mainly the increasing hunting pressure and habitat quality loss, have led to a dramatic decrease of its wild populations in 95% of its original range [10], and the release of farm-reared partridges to maintain natural populations. This situation has raised new concerns about both its human-mediated hybridization with the chukar partridge (*Alectoris chukar*) [1] and sanitary issues [9], thus calling for major the attention by governments, game managers and conservationists.

Infectious diseases have been cited as major factors driving population dynamics [13]. Recent studies have shown that captive-reared partridges are vectors of important avian and even human pathogens (e.g. [4]), and also that reintroductions seem to fail in the absence of medical treatments to decrease parasites [4]. Moreover, frequent enteric diseases increase management difficulties and limit their productivity in captivity, specially since the ban on the use of antimicrobials in poultry production was introduced by the European Union in 2006 [2]. Thus, the ability to respond to a wide spectrum of pathogens and

* Corresponding author.
E-mail address: nsevane@ucm.es (N. Sevane).

http://dx.doi.org/10.1016/j.gdata.2017.02.003

Table 1
De-novo assembly and annotation results summary of transcripts from red-legged partridge immune tissues.

Category	Quantity
All (≥100 bp)	51,403
≥500 bp	15,831
≥1000 bp	4256
N50 (bp)	542
N90 (bp)	239
Max length (bp)	5464
Min length (bp)	100
Average length (bp)	503
Total length (Mb)	25.8
Number of transcripts analysed	51,403
Number of transcripts annotated	11,699 (22.76%)
Number of transcripts not annotated	39,704 (77.24%)

environmental changes will be a major evolutionary force driving survival of partridge populations facing the challenges prompted by restocking programs.

Up to date, a solely differential expression analysis was performed in Sevane et al. [12] on *Gallus gallus* given that the RNA-seq protocol used (single-end 75 bp) and the number of reads analysed (~12 million raw reads per sample) did not allow for transcriptome *de-novo* assembly. In this study, we analysed for the first time different tissues implicated in both innate and acquired immune response (IR) using a high coverage 100-bp paired-end RNA-seq protocol for the reliable identification of partridge genes and markers.

3. Experimental design, materials and methods

We performed non-infectious challenges on a set of 600 seven-month-old red-legged partridges by using a sheep red blood cells (SRBC) hemagglutination assay to measure total antibody titers and IgG activity (i.e. acquired IR) as well as a skin test with phytohemagglutinin (PHA) to quantify T cell-mediated IR (i.e. innate IR) (see [12] for details). A total of sixteen RNA samples from spleen ($n = 4$), bursa of Fabricius ($n = 4$), thymus ($n = 4$) and the skin from PHA injection area ($n = 4$) were extracted as detailed in Sevane et al. [12] and pooled (300 ng each). These samples included two males and two females from each tissue, and the animals displaying the highest and the lowest IR. After full-length cDNA synthesis with Mint2 (Evrogen) and RNA normalization using Trimmer2 (Evrogen), following closely the manufacturer's instruction, the pooled samples were loaded onto a pair-end flowcell for cluster generation (Cluster Station, Illumina). Two consecutive runs were made in a Genome Analyzer GAIIx (Illumina, Unidad de Genómica, Parque Científico de Madrid) using a 100-bp pair-end protocol. FASTQ files were submitted to the NCBI Sequence Read Archive (SRA) (http://ncbi.nlm.nih.gov/sra) with the accession number SRA: PRJNA289204.

4. Results and discussion

A total of 47.1 million raw pair-end reads were obtained using Illumina deep sequencing to investigate the genetic architecture underlying IRs in the red-legged partridge. After removing adaptor sequences along with reads containing >1% Ns (ambiguous bases) and trimming low-quality sequences (quality score Q < 30), 38.3 million clean pair-end reads (81%) remained. The average GC content and sequence length were 36% and 76 bp, respectively. Assembly of the clean reads using OASES v0.2 [11] with the default settings resulted in 51,403 transcripts ranging from 100 bp to 5464 bp, with an average size of 503 bp and a total length size of 25.8 Mb (Table 1, Fig. 1). The N50 length was 542 bp and the N90 239 bp. Several values for the k-mer size of the assembler were tested and the assembly obtained with k = 79 was kept as the most accurate.

Transcripts were functionally annotated based on similarity to 49,102 UniProt reference proteins [8] using a BLASTX minimum similarity threshold of $e < 10^{-9}$. This set of UniProt proteins included all the Uniref90 clusters belonging to the taxonomic node Aves, i.e. each protein belongs to only one cluster and there is a minimum similarity of 90% with the representative protein for all the members of the cluster. A total of 11,699 transcripts were annotated from the red-legged partridge immune transcriptome (Table 1), ranging from 157 bp to 4092 bp with an average length of 764 bp.

The set of transcripts annotated with the same reference protein was defined as unigene, and the largest consensus transcript sequence representative of each unigene was selected (Data S1, Table S1). The 6623 proteins included in Table S1 are the proteins that clustered transcripts from *A. rufa* with a BLASTX similarity meeting a threshold of an $e < 10^{-9}$. Among the annotated unigenes, about 3312 (50%) were matched to chicken (*Gallus gallus*), 1470 (22%) to turkey (*Meleagris gallopavo*), 775 (12%) to human (*Homo sapiens*), 620 (9%) to zebra finch (*Taeniopygia guttata*), and a small amount to other species (Table S1). The annotated partridge sequences showed a mean coverage of the gene to which they shared homology of 43%, with 22% above 80% of coverage. A total of 632 sequences showed 100% gene coverage. Functional data for each representative gene/protein were obtained from UniProt (Table S2): 31,521 gene ontology (GO) terms were assigned to 5344

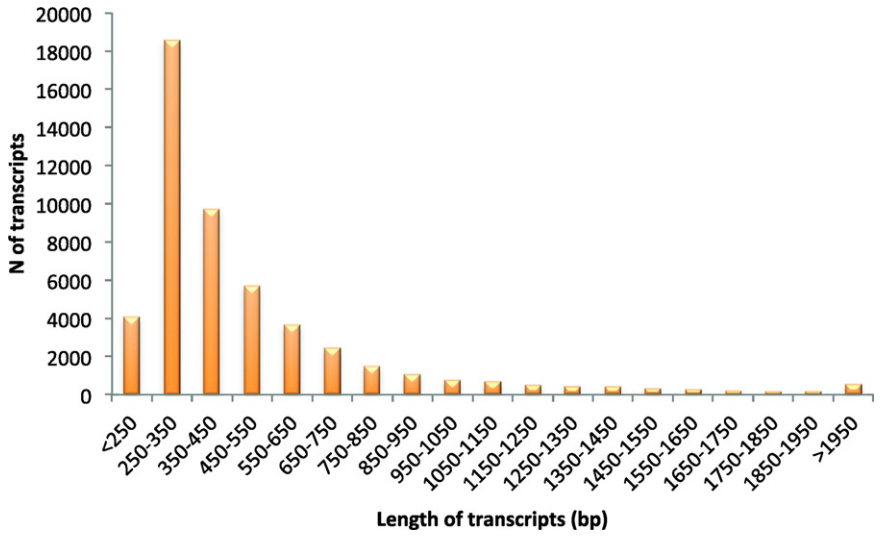

Fig. 1. *De-novo* assembly distribution of transcripts from red-legged partridge immune tissues.

of the annotated unigenes; 1461 unigenes were assigned domains; 428 were assigned enzyme annotations (EC number); and 6128 InterPro motifs.

A total of 12,828 microsatellites were located in the red-legged partridge transcriptome (Table S3, Table S4) using the Genome-wide Microsatellite Analyzing Tool (GMATo v1.2) [14] adding a quality filtering to detect those consisting of >10 repetitions. The most common microsatellite was di-nucleotide (73%). The filtered reads were mapped to the transcripts using Bowtie 2 (v2.2.1) [5] and the single nucleotide polymorphisms (SNPs) called with SAMTools (v0.1.18), BCFtools [6] and VCFtools [3]. Variants were filtered according to the following criteria: two minimum number of alternate bases; SNP within 3 bp around a gap were filtered out; window size of 10 for filtering adjacent gaps; minimum Genotype Quality (GQ) value 30. Finally, a total of 33,857 SNPs situated in 18,371 transcripts were retained (Table S5, Table S6). Among these, 8299 (25%) were in 3025 annotated sequences. The ratio of transitions and transversions (Ts/Tv) at the SNP sites was 3.17. This high Ts/Tv ratio is generally a good measure of a low frequency of false positives in SNP development [7], supporting a high validation rate for the selected SNPs in this study. The GC ratio at the SNP sites was 47.2%.

In conclusion, using RNA-seq we provide the first comprehensive catalogue of red-legged partridge genes expressed in different immune tissues, including thymus, bursa of Fabricius, spleen and skin, after non-infectious challenges aimed at elicit both innate and acquired IRs. The candidate gene sequences and gene-related markers identified here may help understanding key and unique metabolic and immune processes in this species, providing a base for functional genomic and ecological research as well as a resource to understand the immune response in birds.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2017.02.003.

Conflict of interest

The authors have not declared any potential conflicts.

Acknowledgements

This project was funded by the Centre for Industrial Technological Development (CDTI) (IDI-2011-0754) and by ClimGen project (Climate Genomics for Farm Animal Adaptation - FACCE-ERA47 2015-2017). N.S. was recipient of an INNCORPORA-Torres Quevedo contract (PTQ-11-04753) funded by the European Social Fund.

References

[1] F. Barbanera, O.R.W. Pergams, M. Guerrini, G. Forcina, P. Panayides, F. Dini, Genetic consequences of intensive management in game birds. Biol. Conserv. 143 (2010) 1259–1268.
[2] C. Cogliani, H. Goossens, C. Greko, Restricting antimicrobial use in food animals: lessons from Europe. Microbe 6 (2011) 274–279.
[3] P. Danecek, A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, et al., The variant call format and VCFtools. Bioinformatics 27 (2011) 2156–2158.
[4] S. Díaz-Fernández, B. Arroyo, F. Casas, M. Martinez-Haro, J. Viñuela, Effect of game management on wild red-legged partridge abundance. PLoS One 8 (2013), e66671.
[5] B. Langmead, C. Trapnell, M. Pop, S.L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10 (2009) R25.
[6] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, et al., 1000 genome project data processing subgroup (2009): the sequence alignment/map format and SAMtools. Bioinformatics 25 (2010) 2078–2079.
[7] Q. Liu, Y. Guo, J. Li, J. Long, B. Zhang, Y. Shyr, Steps to ensure accuracy in genotype and SNP calling from Illumina sequencing data. BMC Genomics 8 (2012) S8.
[8] Magrane M. & Consortium U. (2011) UniProt Knowledgebase: a hub of integrated protein data. Database (Oxford) 29, bar009.
[9] J. Millán, Diseases of the red-legged partridge (*Alectoris rufa* L.): a review. Wildl. Biol. Pract. 5 (2009) 70–88.
[10] G. Rocamora, D. Yeatman-Berthelot, Oiseaux menacés et à surveiller en France. Listes rouges et recherches de priorités, Societé d'Études Ornithologiques de France/Ligue pour la Protection des Oiseaux, Paris, 1999.
[11] M.H. Schulz, D.R. Zerbino, M. Vingron, E. Birney, Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28 (2012) 1086–1092.
[12] N. Sevane, J. Cañon, I. Gil, S. Dunner, Transcriptomic characterization of innate and acquired immune responses in red-legged partridges (*Alectoris rufa*): a resource for immunoecology and robustness selection. PlosOne 10 (2015), e0136776.
[13] A. Strauss, A. White, M. Boots, Invading with biological weapons: the importance of disease-mediated invasions. Funct. Ecol. 26 (2012) 1249–1261.
[14] X. Wang, P. Lu, Z. Luo, GMATo: A novel tool for the identification and analysis of microsatellites in large genomes. Bioinformation 9 (2013) 541–544.