Draft Genome Assembly of the Poultry Red Mite, *Dermanyssus gallinae*

Stewart T. G. Burgess,[a] Kathryn Bartley,[a] Francesca Nunn,[a] Harry W. Wright,[a] Margaret Hughes,[b] Matthew Gemmell,[b] Sam Haldenby,[b] Steve Paterson,[b,c,d,e] Stephane Rombauts,[c,d,e] Fiona M. Tomley,[f] Damer P. Blake,[f] James Pritchard,[f] Sabine Schicht,[g] Christina Strube,[g] Øivind Øines,[h] Thomas Van Leeuwen,[i] Yves Van de Peer,[c,d,e,j] Alasdair J. Nisbet[a]

[a]Moredun Research Institute (MRI), Edinburgh, United Kingdom

[b]Centre for Genomic Research, Institute of Integrative Biology, University of Liverpool, Liverpool, United Kingdom

[c]Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium

[d]VIB Center for Plant Systems Biology, Ghent, Belgium

[e]Bioinformatics Institute Ghent, Ghent University, Ghent, Belgium

[f]Department of Pathology and Population Sciences, Royal Veterinary College, Hatfield, Hertfordshire, United Kingdom

[g]Institute for Parasitology, Centre for Infection Medicine, University of Veterinary Medicine Hannover, Hannover, Germany

[h]Norwegian Veterinary Institute, Oslo, Norway

[i]Department of Plants and Crops, Ghent University, Ghent, Belgium

[j]Department of Biochemistry, Genetics, and Microbiology, University of Pretoria, Pretoria, South Africa

ABSTRACT The poultry red mite, *Dermanyssus gallinae*, is a major worldwide concern in the egg-laying industry. Here, we report the first draft genome assembly and gene prediction of *Dermanyssus gallinae*, based on combined PacBio and MinION long-read *de novo* sequencing. The ~959-Mb genome is predicted to encode 14,608 protein-coding genes.

Infestation of hen houses with the poultry red mite, *Dermanyssus gallinae*, causes major health and welfare concerns for the egg-producing industry worldwide (1, 2), costing the European Union poultry industry >€231 million annually (see https://www.pluimveeweb.nl/artikelen/2017/01/schade-bloedluis-21-miljoen-euro/). Control relies on the treatment of premises with acaricide sprays or systemic treatments with isoxazoline-based therapeutics (2, 3). Concerns over residues, environmental contamination, and acaricide resistance threaten sustainability and have highlighted interest in developing alternative control methods (1). These novel approaches require comprehensive genomic information and genome-based tools for gene expression analysis and trait mapping.

Adult female *D. gallinae* mites were harvested from a commercial poultry shed in Scotland, and freshly laid mite eggs were collected over 24 h. Contaminating material was removed by washing in 0.1% benzalkonium chloride before rinsing in double-distilled water (ddH$_2$O). Approximately 900 μl of eggs were gently homogenized in 12 ml of SDS/RNase A/proteinase K buffer, and genomic DNA (gDNA) was extracted using the SDS-proteinase K method (4). DNA integrity was assessed by gel electrophoresis and quantified using a Qubit double-stranded DNA (dsDNA) broad-range (BR) kit. PacBio sequencing libraries were generated from high-molecular-weight gDNA using the PacBio SMRTbell template prep kit v1.0 according to the manufacturer's instructions and sequenced using 10 single-molecule real-time (SMRT) cells on a PacBio RS II instrument. Sequences were assembled using Canu v1.6 (5) with an estimated genome size of 500 Mb. The resulting assembly was scaffolded with low-coverage Oxford

Nanopore Technologies MinION reads (6 gigabases [Gb] of sequence data generated with the 1D ligation kit on an R9.4 flow cell) using PBJelly 2 (6) followed by 8 iterations of genome polishing with Arrow (7). The final assembly contained 7,171 contigs with an N_{50} value of 278,630 bp and an L_{50} value of 800 contigs, the largest scaffold having 3,781,415 bp and an overall genome GC content of 44.6%. The assembled genome size was 959 Mb, and 63.5 Gb of PacBio sequencing data provided ~66× coverage.

Gene prediction employed the MAKER pipeline v2.31.8 (8) with Semi-HMM-based Nucleic Acid Parser (SNAP) *ab initio* gene predictions (9) using proteins from *Metaseiulus occidentalis*, *Ixodes scapularis*, and the UniProt and Swiss-Prot protein databases to support gene models. Full-length transcripts were obtained from the total RNA from mixed-stage *D. gallinae* and were analyzed with PacBio Iso-Seq pipelines within the PacBio SMRT Portal v5.0.1.10424 (7) (minimum Quiver/Arrow accuracy, 0.85; minimum GMAP alignment identity, 0.8), generating 13,612 high-quality and 53,082 low-quality isoforms post-Quiver polishing (mean read length of 1,142 bp). Repeat sequences were detected using RepeatModeler (http://www.repeatmasker.org) and provided to MAKER for repeat masking. This identified 14,608 predicted protein-coding genes with an average length of 1,294 bp. The core eukaryotic protein-coding gene presence was assessed with BUSCO (10) v2.0 (Arthropoda set) with 93% of the single-copy orthologs present (73% single copies, 13% duplicates, 7% fragments). Predicted proteins were annotated with Pfam information using InterProScan v5.22-61.0 (11). Those containing a Pfam domain or an annotation score (AED) of <1 were accepted in the final output of 14,608 genes/transcripts. BLAST hits against the NCBI nonredundant (nr) database (July 2018) were identified for 13,840 genes, and Gene Ontology (GO), performed in Blast2GO (10), resulted in the assignment of GO terms for 11,624 genes and functional annotation of 10,914 genes. Unless otherwise stated above, default algorithm parameters were used.

Data availability. This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number QVRM00000000. The version described in this paper is version QVRM01000000. The mixed-stage *D. gallinae* PacBio Iso-Seq data have been deposited at the NCBI Sequence Read Archive under the accession number PRJNA494800.

REFERENCES

1. Sparagano OA, George DR, Harrington DW, Giangaspero A. 2014. Significance and control of the poultry red mite, *Dermanyssus gallinae*. Annu Rev Entomol 59:447–466. https://doi.org/10.1146/annurev-ento-011613-162101.
2. Sigognault Flochlay A, Thomas E, Sparagano O. 2017. Poultry red mite (*Dermanyssus gallinae*) infestation: a broad impact parasitological disease that still remains a significant challenge for the egg-laying industry in Europe. Parasit Vectors 10:357. https://doi.org/10.1186/s13071-017-2292-4.
3. Thomas E, Chiquet M, Sander B, Zschiesche E, Flochlay AS. 2017. Field efficacy and safety of fluralaner solution for administration in drinking water for the treatment of poultry red mite (*Dermanyssus gallinae*) infestations in commercial flocks in Europe. Parasit Vectors 10:457. https://doi.org/10.1186/s13071-017-2390-3.
4. Sambrook J, Russell DW. 2001. Molecular cloning: a laboratory manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
5. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/gr.215087.116.
6. English AC, Richards S, Han Y, Wang M, Vee V, Qu J, Qin X, Muzny DM, Reid JG, Worley KC, Gibbs RA. 2012. Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. PLoS One 7:e47768.
7. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10:563–569. https://doi.org/10.1038/nmeth.2474.
8. Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 12:491. https://doi.org/10.1186/1471-2105-12-491.
9. Korf I. 2004. Gene finding in novel genomes. BMC Bioinformatics 5:59. https://doi.org/10.1186/1471-2105-5-59.
10. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351.
11. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240. https://doi.org/10.1093/bioinformatics/btu031.