Draft Genome Assembly of the Sheep Scab Mite, *Psoroptes ovis*

Stewart T. G. Burgess,[a] Kathryn Bartley,[a] Edward J. Marr,[a] Harry W. Wright,[a] Robert J. Weaver,[j] Jessica C. Prickett,[j] Margaret Hughes,[b] Sam Haldenby,[b] Phuong Thi Le,[c,d,e,f,g] Stephane Rombauts,[c,d,e] Thomas Van Leeuwen,[i] Yves Van de Peer,[c,d,e,h] Alasdair J. Nisbet[a]

[a]Moredun Research Institute (MRI), Edinburgh, United Kingdom
[b]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]Nutriomics: Nutrition and Obesity–Systemic Approaches, Hôpital la Pitié Salpêtrière, Paris, France
[g]Integromics, Hôpital la Pitié Salpêtrière, Paris, France
[h]Department of Genetics, University of Pretoria, Pretoria, South Africa
[i]Department of Plants and Crops, Ghent University, Ghent, Belgium
[j]Fera Science Ltd., York, United Kingdom

ABSTRACT Sheep scab, caused by infestation with *Psoroptes ovis*, is highly contagious, results in intense pruritus, and represents a major welfare and economic concern. Here, we report the first draft genome assembly and gene prediction of *P. ovis* based on PacBio *de novo* sequencing. The ~63.2-Mb genome encodes 12,041 protein-coding genes.

Sheep scab, caused by the ectoparasitic mite *Psoroptes ovis*, is characterized by pruritus and skin irritation and is a major welfare and economic concern for the livestock industry (1, 2). Control relies on injectable macrocyclic lactone-based endectocides and organophosphate dips, but concerns over residues, environmental contamination, and the development of resistance threaten the sustainability of this approach and have highlighted interest in developing alternative control methods (3). However, the development of novel vaccines and identification of potential chemotherapeutic targets has been hampered by the lack of transcriptomic and genomic resources for *P. ovis*.

Adult female *P. ovis* mites (~3,200, Moredun isolate) were harvested, as previously described (4), from experimentally infested sheep. Ethical approval was obtained from the Moredun Research Institute's local ethics committee (E03/17). Any contaminating host material was removed by extensive washing before mites were cleaned with 0.01% Triton X-100. Mites were ground under liquid nitrogen in a prechilled (−80°C) pestle and mortar, and genomic DNA (gDNA) was extracted using the SDS-proteinase K method (5). DNA integrity was assessed on an Agilent Bioanalyzer and quantified using a Qubit double-stranded DNA (dsDNA) broad-range (BR) assay kit. PacBio sequencing libraries were generated from high-molecular-weight gDNA using the PacBio SMRTbell template prep kit version 1.0 according to the manufacturer's instructions, and sequencing was performed using 10 single-molecule real-time (SMRT) cells on the PacBio RS II platform. Genomic sequences were assembled using the Hierarchical Genome Assembly Process version 3 (HGAP3) pipeline embedded in PacBio SMRT Portal version 2.3.0.140936.p0 (6). Filtered subreads were aligned to contigs using BLASR (7), and the corrected consensus reads were generated using Quiver (6). The contaminating sequences were identified using BLASTx searches of the NCBI prokaryotic RefSeq data set, combined with tetramer frequencies and read coverage for emergent self-organizing

map (ESOM) clustering, and removed from the final polished assembly, which contained 93 contigs, with an N_{50} value of 2,279,290 bp and an L_{50} value of 8 contigs (8, 9, 10). The largest scaffold was 5,538,194 bp and, as with other closely related mites, the GC content of the coding sequences was 33.6%. The total genome size was estimated to be 63.2 Mb, and 14 Gb of PacBio data provided ~220-fold coverage.

Gene prediction was performed with EUGENE version 4.1 (11), which was pretrained on *Dermatophagoides farinae* (BioProject number PRJNA379991). The extrinsic data used to support the *ab initio* gene models were protein sets from *Tetranychus urticae*, RefSeq data (invertebrates), *D. farinae* predictions, and transcript data from *P. ovis*. The pipeline identified 12,041 predicted protein-coding genes, with an average length 1,928.21 bp, in agreement with the predicted average gene size (~1,800 bp) for a genome of ~63.2 Mb. The presence of core eukaryotic protein-coding genes was assessed with BUSCO version 1.22 (arthropod set) (12) with 91% of arthropod single-copy orthologs present. Predicted genes were functionally annotated by combining InterProScan (13) and reciprocal best BLAST hits, generating sensible gene function descriptions. Potential secretion signals, transmembrane helices, and other domains were predicted in PHOBIUS (14). BLAST hits against the NCBI nonredundant (nr) database (July 2016) were identified for 9,964 genes, and gene ontology (GO) assessment was performed in Blast2GO (15), resulting in the assignment of GO terms for 7,957 genes and functional annotation for 5,217 genes.

Accession number(s). This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number PQWQ00000000. The version described in this paper is the first version, PQWQ01000000.

REFERENCES

1. Kirkwood AC. 1986. History, biology and control of sheep scab. Parasitol Today 2:302–307. https://doi.org/10.1016/0169-4758(86)90124-9.
2. Nieuwhof GJ, Bishop SC. 2005. Costs of the major endemic diseases of sheep in Great Britain and the potential benefits of reduction in disease impact. Anim Sci 81:23–29. https://doi.org/10.1079/ASC41010023.
3. Nisbet AJ, Huntley JF. 2006. Progress and opportunities in the development of vaccines against mites, fleas and myiasis-causing flies of veterinary importance. Parasite Immunol 28:165–172. https://doi.org/10.1111/j.1365-3024.2006.00803.x.
4. Burgess ST, Frew D, Nunn F, Watkins CA, McNeilly TN, Nisbet AJ, Huntley JF. 2010. Transcriptomic analysis of the temporal host response to skin infestation with the ectoparasitic mite *Psoroptes ovis*. BMC Genomics 11:624. https://doi.org/10.1186/1471-2164-11-624.
5. Sambrook J, Russell DW. 2001. Molecular cloning: a laboratory manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
6. 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.
7. Chaisson MJ, Tesler G. 2012. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. BMC Bioinformatics 13:238. https://doi.org/10.1186/1471-2105-13-238.
8. Ultsch A, Moerchen F. 2005. ESOM-Maps: tools for clustering, visualization, and classification with Emergent SOM, technical report no. 46. Department of Mathematics and Computer Science, University of Marburg, Marburg, Germany.

9. Dick GJ, Andersson A, Baker BJ, Simmons SS, Thomas BC, Yelton AP, Banfield JF. 2009. Community-wide analysis of microbial genome sequence signatures. Genome Biol 10:R85. https://doi.org/10.1186/gb-2009-10-8-r85.
10. Kang DD, Froula J, Egan R, Wang Z. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3:e1165. https://doi.org/10.7717/peerj.1165.
11. Foissac S, Gouzy J, Rombauts S, Mathe C, Amselem J, Sterck L, de Peer Y, Rouze P, Schiex T. 2008. Genome annotation in plants and fungi: EuGene as a model platform. Curr Bioinform 3:87–97. https://doi.org/10.2174/157489308784340702.
12. 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.
13. 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.
14. Käll L, Krogh A, Sonnhammer ELL. 2004. A combined transmembrane topology and signal peptide prediction method. J Mol Biol 338:1027–1036. https://doi.org/10.1016/j.jmb.2004.03.016.
15. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676. https://doi.org/10.1093/bioinformatics/bti610.