genomeA
Journals.ASM.org

High-Quality Draft Genome Sequence of *Babesia divergens*, the Etiological Agent of Cattle and Human Babesiosis

Isabel Cuesta,[a] Luis M. González,[b] Karel Estrada,[c] Ricardo Grande,[d] Ángel Zaballos,[e] Cheryl A. Lobo,[f] Jorge Barrera,[a] Alejandro Sanchez-Flores,[c,d] Estrella Montero[b]

Unidad de Bioinformática, Centro Nacional de Microbiología, Majadahonda, Spain[a]; Servicio de Parasitología, Centro Nacional de Microbiología, Majadahonda, Spain[b]; Unidad Universitaria de Apoyo Bioinformático, Instituto de Biotecnología de la UNAM, Morelos, México[c]; Unidad Universitaria de Secuenciación Masiva de DNA, Instituto de Biotecnología de la UNAM, Morelos, México[d]; Unidad de Secuenciación, Centro Nacional de Microbiología, Majadahonda, Spain[e]; Blood Borne Parasites, LFKRI, New York Blood Center, New York, New York, USA[f]

Babesia divergens **causes significant morbidity and mortality in cattle and splenectomized or immunocompromised individuals. Here, we present a 10.7-Mb high-quality draft genome of this parasite close to chromosome resolution that will enable comparative genome analyses and synteny studies among related parasites.**

Received 6 October 2014 **Accepted** 8 October 2014 **Published** 13 November 2014

Citation Cuesta I, González LM, Estrada K, Grande R, Zaballos Á, Lobo CA, Barrera J, Sanchez-Flores A, Montero E. 2014. High-quality draft genome sequence of *Babesia divergens*, the etiological agent of cattle and human babesiosis. Genome Announc. 2(6):e01194-14. doi:10.1128/genomeA.01194-14.

Address correspondence to Alejandro Sanchez-Flores, alexsf@ibt.unam.mx, or Estrella Montero, estrella.montero@isciii.es.

Babesia divergens is a tick-borne, intraerythrocytic apicomplexan parasite that causes severe hemolytic anemia in cattle and recently has become a significant zoonotic human pathogen (1). Infection by blood transfusion is possible as well, as the infected erythrocytes are optimum vehicles for the parasite (2, 3). Erythrocytic invasion is a critical component of the *B. divergens* asexual life cycle, ensuring continual parasite replication in red blood cells (RBCs) (4) and producing a malaria-like disease. Despite the veterinary and zoonotic importance of this parasite, relatively little research has been carried out on *Babesia*, and many questions regarding the parasite's biology remain unanswered. A better understanding of the species' biology and host-parasite interactions may lead to improved control mechanisms and new trends in vaccine and antibabesial drug development. To facilitate these studies, we sequenced the genome of *B. divergens* human isolate (Rouen 1987) combining three different sequencing platforms to produce a high-quality draft genome.

Genomic DNA was isolated (5, 6) from *B. divergens* human RBC cultures (7). The genome was assembled in three stages. First, a total of 34,182,568 reads, representing 310× coverage, produced with the Illumina HiSeq 2000 platform, were assembled using ALLPATHS-LG (8). Second, 829,056 reads, representing a 23× coverage produced by 454, were assembled using Newbler v. 2.8 and adding the Illumina scaffolds as overlapping "fake" Sanger reads. Third, 109,329 PacBio reads from one SMRT cell run (P4/C2 chemistry) were used to close gaps or extend contig ends in the Illumina+454 hybrid assembly using PBJelly (9). Finally, ICORN2 (10) was used for platform-specific base error corrections. The final assembly had 10,797,556 bp in 514 scaffolds, with N_{50}/N_{90} values of 1,084,746/6,917 bp, respectively. The GC content was 40%, similar to other *Babesia* genomes (11, 12). However, the estimated genome size read k-mer content (~11.5 Mbp) exceeded the observed size in other *Babesia* species and in other *B. divergens* strains (~9 Mbp) (11–13). This could be due to a different repeat content (~30%) or small contigs containing genes with several haplotypes that were not resolved by the assemblers. While this work was in progress, the genomes of *B. divergens* strains (Rouen 1987 and 1802 A strains) were published (12). The Rouen 1987 sequence was reported as a very fragmented assembly with an arbitrary scaffold ordering. In contrast, our assembly contained very large scaffolds where a high number of telomeric repeats are found, indicating reconstruction of almost complete chromosomes. Using CEGMA v. 2.5 (14), we obtained 80% completeness of the genome. Protein-coding genes (3,741) predicted using Augustus (15) and the CEGMA completion level obtained were very similar to those for the recently published *B. divergens* genome (12), although the degree of continuity obtained in our study is superior, as presented in the assembly statistics.

Therefore, the *B. divergens* genome presented here can be used as a reference to study its genome structure in order to understand the regulation of its genes and probe shared synteny with other *Apicomplexan* species.

Nucleotide sequence accession numbers. This genome shotgun project has been deposited at the European Nucleotide Archive under the accession numbers CCSG01000001 to CCSG01000514.

ACKNOWLEDGMENTS

This work was supported by Ministerio de Economía y Competitividad grant AGL2010-21774 (to E.M.), by Instituto de Salud Carlos III grant MPY1197/10 from Spain (to E.M.), and by Unidad Universitaria de Secuenciación Masiva de DNA, Instituto de Biotecnología, UNAM.

We thank the Unidad Universitaria de Apoyo Bioinformático, Instituto de Biotecnología, UNAM, for advice and bioinformatics analysis; Unidad de Bioinformática, Centro Nacional de Microbiología, Spain, for assistance; and Aurelio Velasco for reviewing the manuscript.

REFERENCES

1. **Hunfeld KP, Hildebrandt A, Gray JS.** 2008. Babesiosis: recent insights into an ancient disease. Int. J. Parasitol. 38:1219–1237. http://dx.doi.org/10.1016/j.ijpara.2008.03.001.
2. **Cursino-Santos JR, Alhassan A, Singh M, Lobo CA.** 2014. *Babesia*:

impact of cold storage on the survival and the viability of parasites in blood bags. Transfusion **54:**585–591. http://dx.doi.org/10.1111/trf.12357.
3. **Castro E, Gonzalez LM, Rubio JM, Ramiro R, Girones N, Montero E.** 2014. The efficacy of the ultraviolet C pathogen inactivation system in the reduction of *Babesia divergens* in pooled buffy coat platelets. Transfusion. **54:**2207–2216. http://dx.doi.org/10.1111/trf.12598.
4. **Lobo CA, Rodriguez M, Cursino-Santos JR.** 2012. *Babesia* and red cell invasion. Curr. Opin. Hematol. **19:**170–175. http://dx.doi.org/10.1097/MOH.0b013e328352245a.
5. **Ahmad S, Ghosh A, Nair DL, Seshadri M.** 2007. Simultaneous extraction of nuclear and mitochondrial DNA from human blood. Genes Genet. Syst. **82:**429–432. http://dx.doi.org/10.1266/ggs.82.429.
6. **Lahiri DK, Nurnberger JI, Jr.** 1991. A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. Nucleic Acids Res. **19:**5444.
7. **Gorenflot A, Brasseur P, Precigout E, L'Hostis M, Marchand A, Schrevel J.** 1991. Cytological and immunological responses to *Babesia divergens* in different hosts: ox, gerbil, man. Parasitol. Res. **77:**3–12.
8. **Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB.** 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc. Natl. Acad. Sci. U. S. A. **108:**1513–1518. http://dx.doi.org/10.1073/pnas.1017351108.
9. **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. http://dx.doi.org/10.1371/journal.pone.0047768.
10. **Otto TD, Sanders M, Berriman M, Newbold C.** 2010. Iterative Correction of Reference Nucleotides (iCORN) using second generation sequencing technology. Bioinformatics **26:**1704–1707. http://dx.doi.org/10.1093/bioinformatics/btq269.
11. **Brayton KA, Lau AO, Herndon DR, Hannick L, Kappmeyer LS, Berens SJ, Bidwell SL, Brown WC, Crabtree J, Fadrosh D, Feldblum T, Forberger HA, Haas BJ, Howell JM, Khouri H, Koo H, Mann DJ, Norimine J, Paulsen IT, Radune D, Ren Q, Smith RK, Jr, Suarez CE, White O, Wortman JR , Knowles DP, Jr, McElwain TF, Nene VM.** 2007. Genome sequence of *Babesia bovis* and comparative analysis of apicomplexan hemoprotozoa. PLoS Pathog. **3:**1401–1413. http://dx.doi.org/10.1371/journal.ppat.0030148.
12. **Jackson AP, Otto TD, Darby A, Ramaprasad A, Xia D, Echaide IE, Farber M, Gahlot S, Gamble J, Gupta D, Gupta Y, Jackson L, Malandrin L, Malas TB, Moussa E, Nair M, Reid AJ, Sanders M, Sharma J, Tracey A, Quail MA, Weir W, Wastling JM, Hall N, Willadsen P, Lingelbach K, Shiels B, Tait A, Berriman M, Allred DR, Pain A.** 2014. The evolutionary dynamics of variant antigen genes in *Babesia* reveal a history of genomic innovation underlying host-parasite interaction. Nucleic Acids Res. **42:**7113–7131. http://dx.doi.org/10.1093/nar/gku322.
13. **Cornillot E, Hadj-Kaddour K, Dassouli A, Noel B, Ranwez V, Vacherie B, Augagneur Y, Brès V, Duclos A, Randazzo S, Carcy B, Debierre-Grockiego F, Delbecq S, Moubri-Ménage K, Shams-Eldin H, Usmani-Brown S, Bringaud F, Wincker P, Vivarès CP, Schwarz RT, Schetters TP, Krause PJ, Gorenflot A, Berry V, Barbe V, Ben Mamoun C.** 2012. Sequencing of the smallest apicomplexan genome from the human pathogen *Babesia microti*. Nucleic Acids Res. **40:**9102–9114. http://dx.doi.org/10.1093/nar/gks700.
14. **Parra G, Bradnam K, Ning Z, Keane T, Korf I.** 2009. Assessing the gene space in draft genomes. Nucleic Acids Res. **37:**289–297. http://dx.doi.org/10.1093/nar/gkn916.
15. **Stanke M, Schoffmann O, Morgenstern B, Waack S.** 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics **7:**62. http://dx.doi.org/10.1186/1471-2105-7-62.