Complete Genome Sequence of the Endosymbiotic Bacterium "*Candidatus* Riesia pediculicola"

Jongsun Park,[a,b] Si Hyeock Lee,[c,d] Ju Hyeon Kim[c]

aInfoBoss, Inc., Gangnam-gu, Seoul, Republic of Korea
bInfoBoss Research Center, Gangnam-gu, Seoul, Republic of Korea
cResearch Institute of Agriculture and Life Sciences, Seoul National University, Seoul, Republic of Korea
dDepartment of Agricultural Biotechnology, Seoul National University, Seoul, Republic of Korea

ABSTRACT Human head and body lice host the obligate endosymbiotic bacterium "*Candidatus* Riesia pediculicola." In this announcement, we describe the complete genome sequence of a "*Ca.* Riesia pediculicola" strain isolated from the human head louse, *Pediculus humanus* subsp. *capitis*. The inter- and intraspecific variations of endosymbiont genomes were investigated, and this strain was found to display high-level variations in its genome.

The human head louse, *Pediculus humanus* subsp. *capitis*, is an obligate hematophagous ectoparasite living on human hair and scalp. It hosts the endosymbiotic bacterium "*Candidatus* Riesia pediculicola," which is associated with the stomach disc and provides vitamin B, which is lacking in human blood (1).

P. humanus subsp. *capitis* specimens were collected in Bristol, United Kingdom, and reared on an *in vitro* rearing system (2). Total DNA was extracted from approximately 500 newly hatched first-instar nymphs before their first blood meal using a DNeasy blood and tissue kit (Qiagen, Hilden, Germany). Genome sequencing was performed on the extracted DNA using the Illumina GAIIx platform at the National Instrumentation Center for Environmental Management (Seoul, South Korea). A total of 114 million 101-bp raw reads (11.5 Gbp) were generated using TruSeq DNA sample preparation kits based on 350-bp DNA fragments. Bioinformatic analyses were performed using default parameters except where otherwise noted. *De novo* assembly of the bacterial genome sequence was performed using Velvet v1.2.10 (3) with the minimum depth option set to 30×, after filtering the raw reads using Trimmomatic v0.33 (4) in the Genome Information System (GeIS) environment. Contigs obtained from the Velvet assembly results were selected if their sequences were matched to those of "*Ca.* Riesia pediculicola" according to BLASTN results against the nonredundant (NR) data set. The gaps were closed using GapCloser v1.12 (5). As the previously sequenced genome (GenBank accession number CP001085) (6) is linear, displaying both end sequences would be repetitive; thus, genome assembly was performed only when the previously assembled genome sequence showed the same phenomenon. After that, all bases were confirmed based on the raw read alignment against the assembled genome using BWA v0.7.17 (7) and SAMtools v1.9 (8). The coverage of this assembled genome sequence is 296.57×. Genome annotation was conducted using the NCBI Prokaryotic Genome Annotation Pipeline (9).

The complete genome sequence of "*Ca.* Riesia pediculicola" is 574,999 bp long, with a GC content of 28.5%. It is slightly larger than that of the previous genome sequence (GenBank accession number CP001085; 574,390 bp). This genome sequence contains 469 protein-coding genes (PCGs), 34 tRNAs, 6 rRNAs, and 1 transfer-messenger RNA

(tmRNA). One additional tRNA, *tRNA-Met*, exists in our genome sequence, while the number of PCGs is much smaller because of the putative proteins predicted by GLIMMER.

Based on the pairwise alignment of both genome sequences using MAFTT v7.450 (10), 784 single nucleotide polymorphisms (SNPs) (0.136%) and 80 indel regions covering 669 bp (0.116%) were identified, showing one small inversion in the signal recognition particle receptor FtsY and 16S rRNA (guanine^{966}-N^2)-methyltransferase. The number of SNPs is smaller but the coverage of indels is higher than for *Buchnera aphidicola* (*Aphis gossypii*). The proportions of SNPs and indels identified in the two host species genome sequences (1.367% and 0.893%, respectively) (11) are higher than those in the endosymbiotic bacterial genome sequences, suggesting that the pressures of natural selection in the endosymbiont environment may be less than that for the host species. This is similar to the case between the mitogenomes of *Aphis gossypii* (12, 13) and its endosymbiont.

Our "*Ca.* Riesia pediculicola" genome sequence can provide insight into the intra-specific variation of endosymbiont genomes, as well as clues to understanding their evolutionary histories, with additional endosymbiont genome sequences forthcoming.

Data availability. The "*Ca.* Riesia pediculicola" whole-genome sequencing project has been deposited at DDBJ/ENA/GenBank under accession number CP062474, BioProject accession number PRJNA665755, and BioSample accession number SAMN16268218. The raw sequences were deposited under accession number SRR12717033.

REFERENCES

1. Puchta O. 1955. Experimentelle Untersuchungen über die Bedeutung der Symbiose der Kleiderlaus Pediculus vestimenti Burm. Zeitschrift für Parasitenkunde 17:1–40. (In German.)
2. Yoon KS, Strycharz JP, Gao J-R, Takano-Lee M, Edman JD, Clark JM. 2006. An improved in vitro rearing system for the human head louse allows the determination of resistance to formulated pediculicides. Pestic Biochem Physiol 86:195–202. https://doi.org/10.1016/j.pestbp.2006.04.001.
3. Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res 18:821–829. https://doi.org/10.1101/gr.074492.107.
4. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
5. Zhao Q-Y, Wang Y, Kong Y-M, Luo D, Li X, Hao P. 2011. Optimizing *de novo* transcriptome assembly from short-read RNA-Seq data: a comparative study. BMC Bioinformatics 12:S2. https://doi.org/10.1186/1471-2105-12-S14-S2.
6. Kirkness EF, Haas BJ, Sun W, Braig HR, Perotti MA, Clark JM, Lee SH, Robertson HM, Kennedy RC, Elhaik E, Gerlach D, Kriventseva EV, Elsik CG, Graur D, Hill CA, Veenstra JA, Walenz B, Tubío JMC, Ribeiro JMC, Rozas J, Johnston JS, Reese JT, Popadic A, Tojo M, Raoult D, Reed DL, Tomoyasu Y, Kraus E, Mittapalli O, Margam VM, Li H-M, Meyer JM, Johnson RM, Romero-Severson J, VanZee JP, Alvarez-Ponce D, Vieira FG, Aguadé M, Guirao-Rico S, Anzola JM, Yoon KS, Strycharz JP, Unger MF, Christley S, Lobo NF, Seufferheld MJ, Wang N, Dasch GA, Struchiner CJ, et al. 2010. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. Proc Natl Acad Sci U S A 107:12168–12173. https://doi.org/10.1073/pnas.1003379107.
7. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv 1303.3997 [q-bio.GN].
8. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
9. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.
10. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010.
11. Kang JS, Cho Y-J, Kim JH, Kim SH, Yoo S, Noh S-J, Park J, Yoon KS, Marshall Clark J, Pittendrigh BR, Chun J, Lee SH. 2015. Comparison of the genome profiles between head and body lice. J Asia Pac Entomol 18:377–382. https://doi.org/10.1016/j.aspen.2015.04.010.
12. Park J, Xi H, Kim Y, Park J, Lee W. 2019. The complete mitochondrial genome of *Aphis gossypii* Glover, 1877 (Hemiptera: Aphididae) collected in Korean peninsula. Mitochondrial DNA B Resour 4:3007–3009. https://doi.org/10.1080/23802359.2019.1666051.
13. Bae Y, Park J, Lee W. 2020. The complete mitochondrial genome of *Aphis gossypii* Glover, 1877 (Hemiptera: Aphididae) isolated from *Plantago asiatica* in Korea. Mitochondrial DNA B Resour 5:2878–2880. https://doi.org/10.1080/23802359.2020.1792366.