

Draft Genome Sequence of *Caedibacter varicaedens*, a Kappa Killer Endosymbiont Bacterium of the Ciliate *Paramecium biaurelia*

Haruo Suzuki,^{a,b} Amy L. Dapper,^a Craig E. Jackson,^{c,d} Heewook Lee,^c Vikas Pejaver,^c Thomas G. Doak,^a Michael Lynch,^a John R. Preer, Jr.^a

Department of Biology, Indiana University, Bloomington, Indiana, USA^a; Institute for Advanced Biosciences, Keio University, Fujisawa, Kanagawa, Japan^b; Department of Computer Science and Informatics, School of Informatics and Computing, Indiana University, Bloomington, Indiana, USA^c; School of Public and Environmental Affairs, Indiana University, Bloomington, Indiana, USA^d

***Caedibacter varicaedens* is a kappa killer endosymbiont bacterium of the ciliate *Paramecium biaurelia*. Here, we present the draft genome sequence of *C. varicaedens*.**

Received 21 September 2015 Accepted 25 September 2015 Published 5 November 2015

Citation Suzuki H, Dapper AL, Jackson CE, Lee H, Pejaver V, Doak TG, Lynch M, Preer JR, Jr. 2015. Draft genome sequence of *Caedibacter varicaedens*, a kappa killer endosymbiont bacterium of the ciliate *Paramecium biaurelia*. *Genome Announc* 3(6):e01310-15. doi:10.1128/genomeA.01310-15.

Copyright © 2015 Suzuki et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to Haruo Suzuki, haruo@g-language.org.

The genus *Caedibacter* comprises species “*Candidatus Caedibacter acanthamoebae*” (endosymbiont of acanthamoebae), *Caedibacter caryophilus* (*Paramecium caudatum* symbiont), *Caedibacter macronucleorum* (*Paramecium duboscqui* symbiont), *Caedibacter paraconjugatus* (*Paramecium biaurelia* symbiont), *Caedibacter pseudomutans* (*Paramecium tetraurelia* symbiont), *Caedibacter taeniospiralis* (*Paramecium caudatum* symbiont), and *Caedibacter varicaedens* (*Paramecium biaurelia* symbiont) (1). Of the *Caedibacter* species described to date, only the draft genome sequence of “*Ca. Caedibacter acanthamoebae*” has been determined (2). *Caedibacter varicaedens* is a kappa killer endosymbiont bacterium of the ciliate *Paramecium biaurelia* (3). To study the evolution of the *Caedibacter-Paramecium* symbiosis, we determined the draft genome sequence of *C. varicaedens*.

Whole-genome sequencing was performed using the Roche 454 pyrosequencing platform at the Center for Genomics and Bioinformatics (CGB), Indiana University. The sequencing reads were assembled using Newbler version 2.5.3. After removal of possible contaminant sequences, i.e., contigs with unusual values of statistics such as %GC content and coverage (sequencing depth), the resulting assembly contains 142 contigs consisting of 1,686,852 bp, with a G+C content of 42.1%. Genome annotation was performed using Prokka v1.11 (4), yielding 41 tRNA genes, 1 copy of a 16S-23S-5S rRNA gene, and 1,726 protein-coding sequences (CDS), of which 588 (34%) are hypothetical proteins and 8 contain a protein motif (Pfam PF11747.2) related to the killing trait deposited in Pfam: the protein families database (5). SignalP 4.1 (6) predicted 86 signal peptides. The genome sequence was analyzed using G-language Genome Analysis Environment version 1.9.0 (7), available at <http://www.g-language.org>.

We performed similarity searches of the 1,726 protein sequences against the UniRef90 sequence databases (8) using BLASTP (9) with the E value cutoff of 1e-20 and assigned the most similar (best hit) protein sequence information. Of the 1,726 proteins, 1,222 (71%) had matches with 1,090 unique records in the UniRef90 database, of which 100 (8%) are transposase family proteins. The species distribution of the BLAST best hits in the Uni-

Ref90 database showed that 565 (46%) of the 1,222 hits had top matches with sequences from “*Ca. Caedibacter acanthamoebae*,” followed by “*Candidatus Paracaedibacter symbiosus*” with 36 (2.9%) best BLAST hits. Both of the species, as well as *C. varicaedens*, belong to the order *Rickettsiales* of *Alphaproteobacteria* (1). The genome sequence data in this study will provide useful information for understanding the evolution of symbiosis between *C. varicaedens* and the host *P. biaurelia*.

Nucleotide sequence accession numbers. The sequence has been deposited as a whole-genome shotgun project at DDBJ/EMBL/GenBank under the accession number [BBVC00000000](https://www.ncbi.nlm.nih.gov/nuccore/BBVC00000000). The version described in this paper is the first version, BBVC01000000 (BBVC01000001 to BBVC01000142).

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation, grant MCB-1050161.

This genome was sequenced and analyzed as part of a graduate class at Indiana University. We thank Amanda Kyle Gibson for her help in the class.

REFERENCES

1. Beier CL, Horn M, Michel R, Schweikert M, Gortz HD, Wagner M. 2002. The genus *Caedibacter* comprises endosymbionts of *Paramecium* spp. related to the *Rickettsiales* (*Alphaproteobacteria*) and to *Francisella tularensis* (*Gammaproteobacteria*). *Appl Environ Microbiol* 68:6043–6050. <http://dx.doi.org/10.1128/AEM.68.12.6043-6050.2002>.
2. Wang Z, Wu M. 2015. An integrated phylogenomic approach toward pinpointing the origin of mitochondria. *Sci Rep* 5:7949. <http://dx.doi.org/10.1038/srep07949>.
3. Preer JR, Jr, Preer LB, Jurand A. 1974. Kappa and other endosymbionts in *Paramecium aureliae*. *Bacteriol Rev* 38:113–163.
4. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <http://dx.doi.org/10.1093/bioinformatics/btu153>.
5. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M. 2014. Pfam: the protein families database. *Nucleic Acids Res* 42:D222–D230. <http://dx.doi.org/10.1093/nar/gkt1223>.

6. Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8:785–786. <http://dx.doi.org/10.1038/nmeth.1701>.
7. Arakawa K, Mori K, Ikeda K, Matsuzaki T, Kobayashi Y, Tomita M. 2003. G-language genome analysis environment: a workbench for nucleotide sequence data mining. *Bioinformatics* 19:305–306. <http://dx.doi.org/10.1093/bioinformatics/19.2.305>.
8. UniProt Consortium. 2014. Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res* 42:D191–D198. <http://dx.doi.org/10.1093/nar/gkt1140>.
9. Altschul S, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and psi-blast: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402. <http://dx.doi.org/10.1093/nar/25.17.3389>.