

Draft *Anaplasma phagocytophilum* Genome Sequences from Five Cows, Two Horses, and One Roe Deer Collected in Europe

Thibaud Dugat,^a Marie-Noëlle Rossignol,^b Olivier Rué,^c Valentin Loux,^c Sylvain Marthey,^b Marco Moroldo,^b Cornelia Silaghi,^{d*} Dirk Höper,^e Julia Fröhlich,^d Martin Pfeffer,^e Erich Zweggarth,^{d*} Anne-Claire Lagrée,^f Henri-Jean Boulouis,^f Nadia Haddad^f

Université Paris-Est, Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail, Laboratoire de santé animale, UMR BIPAR, Maisons-Alfort, France^a; GABI, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France^b; MalAGE, INRA, Université Paris-Saclay, Jouy-en-Josas, France^c; Comparative Tropical Medicine and Parasitology, Ludwig-Maximilians-Universität München, Munich, Germany^d; Institute of Diagnostic Virology, Friedrich-Loeffler-Institut, Federal Research Institute for Animal Health, Insel Riems, Germany^e; Université Paris-Est, École Nationale Vétérinaire d'Alfort, UMR BIPAR ENVA Anses UPEC USC INRA, Maisons-Alfort, France^f

* Present address: Cornelia Silaghi, National Centre for Vector Entomology, Institute of Parasitology, Vetsuisse-Faculty, University of Zurich, Zürich, Switzerland; Erich Zweggarth, Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of Pretoria, Pretoria, South Africa, and Institute of Parasitology and Tropical Veterinary Medicine, Free University of Berlin, Berlin, Germany.

***Anaplasma phagocytophilum* is a zoonotic tick-borne intracellular bacterium responsible for granulocytic anaplasmosis. As it is difficult to isolate and cultivate, only 20 *A. phagocytophilum* genomes have been sequenced to date. Here, we present eight *A. phagocytophilum* genome sequences obtained using alternative approaches based on sequence capture technology.**

Received 19 July 2016 Accepted 7 October 2016 Published 1 December 2016

Citation Dugat T, Rossignol M-N, Rué O, Loux V, Marthey S, Moroldo M, Silaghi C, Höper D, Fröhlich J, Pfeffer M, Zweggarth E, Lagrée A-C, Boulouis H-J, Haddad N. 2016. Draft *Anaplasma phagocytophilum* genome sequences from five cows, two horses, and one roe deer collected in Europe. *Genome Announc* 4(6):e00950-16. doi:10.1128/genomeA.00950-16.

Copyright © 2016 Dugat et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Nadia Haddad, nadia.haddad@vet-alfort.fr.

Anaplasma phagocytophilum is a zoonotic and obligate intracellular bacterium transmitted by hard ticks. It is the causative agent of tick-borne fever in ruminants, a disease which causes significant economic losses in Europe, and of equine granulocytic anaplasmosis (EGA) in horses, of canine granulocytic anaplasmosis (CGA) in dogs, and of human granulocytic anaplasmosis (HGA) in both the United States and Europe (1). *A. phagocytophilum* is difficult to isolate and cultivate, and thus only 20 genomes have been sequenced to date, among which only four originate from European samples. In particular, there are no *A. phagocytophilum* genomes from horse or roe deer samples, even though EGA can have such important economic impact (2), and roe deer are suspected to play a crucial role as reservoir hosts in *A. phagocytophilum* epidemiology (1). Here, we present the draft genome sequences of European *A. phagocytophilum* from four cow and two horse samples, and two strains isolated from one cow and one roe deer which had been maintained in continuous cell cultures (3, 4).

To overcome limitations due to the lack of reliable, easy, and

feasible protocols to isolate *A. phagocytophilum*, we decided to follow a different strategy in order to obtain these genome sequences. We used a whole-genome sequence capture approach, whose efficacy on *A. phagocytophilum* has already been demonstrated (5). For each sample, total genomic DNA was extracted from whole blood of the animal, or from bacterial culture in IDE8 tick cell lines, using the NucleoSpin Blood QuickPure kit (Macherey-Nagel, Bethlehem, USA). The six bacterial genomes from animal blood were then captured and sequenced as previously described (5), but in this case a liquid-phase protocol based on the SeqCap EZ library kit (NimbleGen, Madison, USA) was used instead of solid-phase. All eight samples were sequenced on a single flow cell lane of a HiSeq3000 (Illumina, San Diego, USA) sequencer as paired-end 150 bp reads (insert-size: 280 ± 50 bp).

First, overlapping paired-end reads were merged with Flash (6) and were trimmed with Sickle (<https://github.com/najoshi/sickle>) (−n −q 24 −l 100). Reads were digitally normalized using Khmer (7) with a k-mer size of 20 and a cutoff of 20. Then, the

TABLE 1 Genome sequence accession numbers

Sample	Nucleotide sequence accession no.	Host	Geographical origin	Genome size (bp)	No. of contigs (scaffolds)	G+C (%)	No. of genes
Cow_1	FLLR01000001–FLLR01000249	Cow	France	1,682,137	249	41.93	1,519
Cow_2	FLMA01000001–FLMA01000230	Cow	France	1,641,350	230	42.23	1,463
Cow_3	FLMB01000001–FLMB01000191	Cow	France	1,561,654	191	42.00	1,417
Cow_4	FLLZ01000001–FLLZ01000230	Cow	France	1,604,419	230	42.12	1,448
Cow_5	FLMD02000001–FLMD02000300	Cow	Germany	2,196,284	300	43.70	1,979
Horse_1	FLMF02000001–FLMF02000300	Horse	France	1,784,294	300	41.90	1,657
Horse_2	FLMC02000001–FLMC02000300	Horse	France	2,191,611	300	41.58	1,997
Roe_deer_1	FLME02000001–FLME02000300	Roe deer	Germany	2,120,290	300	42.72	1,912

remaining reads were mapped to the host genomes [*Bos taurus* (UMD3.1), *Equus caballus* (EquCab2.0), and *Capreolus capreolus* (CCMK000000000)] and to *A. phagocytophilum* HZ (NC_007797.1), using the BWA algorithm (v0.6.1) with default parameters (8). Only those reads mapping to *A. phagocytophilum* or which remained unmapped were retained. Three sets of reads were used for assembly processing: paired-end reads mapped on *A. phagocytophilum*, paired-end unmapped reads, and singleton reads (unmapped or mapped to *A. phagocytophilum*). Those reads were then assembled using Spades (9), with a k-mer value from 81 to 125 with a step of 4, and default parameters. Genome annotation was performed using Prokka. Genome characteristics are summarized in Table 1.

Accession number(s). The eight draft genomes have been deposited in the European Nucleotide Archive (Table 1).

ACKNOWLEDGMENTS

This work was performed within the Laboratory of Excellence (Labex) of Integrative Biology of Emerging Infectious Diseases (IBEID).

This work was supported by the Alfort National Veterinary School (ENVA), by the French National Institute for Agricultural Research (INRA), and by the French Agency for Food, Environmental and Occupational Health and Safety (Anses). Thibaud Dugat was funded by Anses.

REFERENCES

- Dugat T, Lagrée A-C, Maillard R, Boulouis H-J, Haddad N. 2015. Opening the black box of *Anaplasma phagocytophilum* diversity: current situation and future perspectives. *Front Cell Infect Microbiol* 5:61. <http://dx.doi.org/10.3389/fcimb.2015.00061>.
- Dzięgiel B, Adaszek Ł, Kalinowski M, Winiarczyk S. 2013. Equine granulocytic anaplasmosis. *Res Vet Sci* 95:316–320. <http://dx.doi.org/10.1016/j.rvsc.2013.05.010>.
- Silaghi C, Kauffmann M, Passos LM, Pfister K, Zweygarth E. 2011. Isolation, propagation and preliminary characterisation of *Anaplasma phagocytophilum* from roe deer (*Capreolus capreolus*) in the tick cell line IDE8. *Ticks Tick-Borne Dis* 2:204–208. <http://dx.doi.org/10.1016/j.ttbdis.2011.09.002>.
- Wernike K, Silaghi C, Nieder M, Pfeffer M, Beer M. 2014. Dynamics of Schmallenberg virus infection within a cattle herd in Germany, 2011. *Epidemiol Infect* 142:1501–1504. <http://dx.doi.org/10.1017/S0950268813002525>.
- Dugat T, Loux V, Marthey S, Moroldo M, Lagrée A-C, Boulouis H-J, Haddad N, Maillard R. 2014. Comparative genomics of first available bovine *Anaplasma phagocytophilum* genome obtained with targeted sequence capture. *BMC Genomics* 15:973. <http://dx.doi.org/10.1186/1471-2164-15-973>.
- Magoč T, Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics Oxf Engl* 27:2957–2963. <http://dx.doi.org/10.1093/bioinformatics/btr507>.
- Crusoe MR, Alameldin HF, Awad S, Boucher E, Caldwell A, Cartwright R, Charbonneau A, Constantinides B, Edvenson G, Fay S, Fenton J, Fenzl T, Fish J, Garcia-Gutierrez L, Garland P, Gluck J, González I, Guermond S, Guo J, Gupta A, Herr JR, Howe A, Hyer A, Härpfer A, Irber L, Kidd R, Lin D, Lippi J, Mansour T, McA'Nulty P, McDonald E, Mizzi J, Murray KD, Nahum JR, Nanlohy K, Nederbragt AJ, Ortiz-Zuazaga H, Ory J, Pell J, Pepe-Ranney C, Russ ZN, Schwarz E, Scott C, Seaman J, Sievert S, Simpson J, Skennerton CT, Spencer J, Srinivasan R, Standage D, Stapleton JA, Steinman SR, Stein J, Taylor B, Trimble W, Wiencko HL, Wright M, Wyss B, Zhang Q, Zyme E, Brown CT. 2015. The khmer software package: enabling efficient nucleotide sequence analysis. *F1000Res* 4:900. <http://dx.doi.org/10.12688/f1000research.6924.1>.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics Oxf Engl* 25:1754–1760. <http://dx.doi.org/10.1093/bioinformatics/btp324>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <http://dx.doi.org/10.1089/cmb.2012.0021>.