Table 2. Lyssavirus isolates used for genetic analysis

Virus name	Host	Country	Year	Accession no.
MGL 2	cattle	Mongolia (Govi-Altai)[3)]	2005	AB570995
MGL 5	red fox	Mongolia (Zavkhan)	2006	AB570996
MGL 10	cattle	Mongolia (Khuvsgul)	2006	AB570997
MGL 11	cattle	Mongolia (Zavkhan)	2006	AB570998
MGL 12	camel	Mongolia (Govi-Altai)	2006	AB570999
MGL 13	cattle	Mongolia (Zavkhan)	2008	AB571000
MGL 17	cattle	Mongolia (Zavkhan)	2006	AB571001
MGL 20	cattle	Mongolia (Zavkhan)	2006	AB571002
MGL 21	red fox	Mongolia (Zavkhan)	2008	AB571003
MGL 22	dog	Mongolia (Tuv)	2007	AB571004
MGL 23	camel	Mongolia (Govi-Altai)	2008	AB571005
MGL 24	cattle	Mongolia (Khuvsgul)	2008	AB571006
MGL 25	dog	Mongolia (Khuvsgul)	2008	AB571007
MGL 26	cattle	Mongolia (Zavkhan)	2006	AB571008
MGL 27	red fox	Mongolia (Zavkhan)	2006	AB571009
MGL 28	cattle	Mongolia (Zavkhan)	2006	AB571010
MGL 29	goat	Mongolia (Govi-Altai)	2008	AB571011
MGL 30	cattle	Mongolia (Khuvsgul)	2007	AB571012
MGL 31	cattle	Mongolia (Zavkhan)	2005	AB571013
MGL 32	red fox	Mongolia (Zavkhan)	2005	AB571014
MGL 33	camel	Mongolia (Govi-Altai)	2005	AB571015
MGL 34	wild cat	Mongolia (Govi-Altai)	2005	AB571016
MGL 35	sheep	Mongolia (Bayan-Ulgii)	2005	AB571017
MGL 36	dog	Mongolia (Ulaanbaatar)	2005	AB571018
Mongolia 3[1)]	cattle	Mongolia (Zavkhan)	2006	EF614257
Mongolia 4[1)]	wolf	Mongolia (Zavkhan)	2006	EF614256
Mongolia 6[1)]	red fox	Mongolia (Govi-Altai)	2006	EF614254
Mongolia 7[1)]	camel	Mongolia (Govi-Altai)	2006	EF614255
857r	raccoon dog	Russia	2004	AY352458
304c	steppe fox	Russia	2004	AY352459
248c	steppe fox	Russia	2004	AY352460
3561d	dog	Russia	2004	AY352481
686cattle	cattle	Russia	2004	AY352482
765w	wolf	Russia	2004	AY352483
409f	red fox	Kazakhstan	2004	AY352489
RV259	red fox	Kazakhstan	2004	AY352491
8684GRO	arctic fox	Greenland	1981	U22654
8618POL	raccoon dog	Poland	1985	U22840
9342EST	raccoon dog	Estonia	1999	U43432
9215HON	human	Hungary	1999	U43025
9202ALL	red fox	Germany	2005	U42701
9223FRA	red fox	France	1999	U43433
86107YOU	red fox	Yugoslavia	1999	U42703
8653YOU	wolf	Yugoslavia	1999	U42704
Guangxi_Yl66	dog	China	2006	DQ666287
Guizhou_A148	dog	China	2006	DQ666291
Henan_Sq10	dog	China	2006	DQ666300
Hunan_Xx33	dog	China	2006	DQ666317
SKRRD0406CC	raccoon dog	South Korea	2004	DQ076126
Yokohama	human	Japan[4)]	2006	AB573763
Kyoto	human	Japan[4)]	2006	AB573762
NNV-RAB-H	human	India	2007	EF437215
SRL1145	buffalo	Sri-Lanka	1996	AB041969
8681IRA	dog	Iran	1995	U22482
9001FRA	dog	Guyana	1990	EU293113
Eth2003	wolf	Ethiopia	2003	AY500827
8721AFS	human	South Africa	1995	U22633
8480FX	red fox	Canada	1993	U03768
3789	skunk	USA	1998	AF461045
Chile	bat	Chile	1999	AF070450
pchm3230	human	Peru	1999	AF045166
Argentina	bat	Argentina	1997	EU293116
PV		Lab strain		M13215
CVS-11		Lab strain		AB069973
ERA		Lab strain		EF206707
ABL[2)]	bat	Australia	2002	NC003243

[1)]: described previously by Botvinkin et al., 2008 (10).
[2)]: Australian bat lyssavirus.
[3)]: Province of Mongolia.
[4)]: Japanese bitten by the dog in the Philippines.

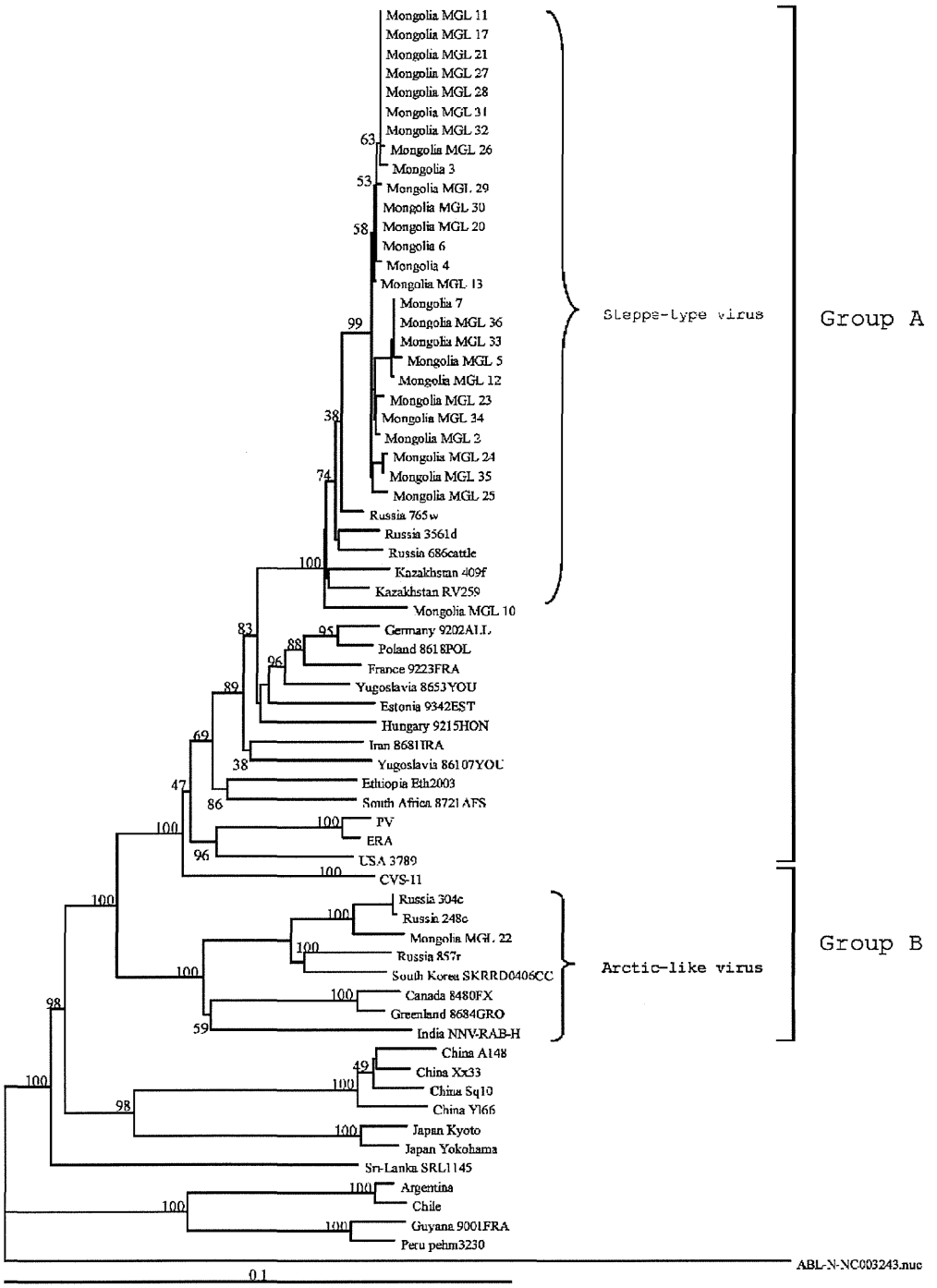

Fig. 2. Phylogenetic tree with complete sequences of N gene of rabies virus from Mongolia and other countries. ClustalX was used for sequence alignment and TreeView was used for drawing phylogenetic tree.

a rabies-related virus (Irkut virus) in bats (*Murina leucogaster*) thriving in Irkutsk adjacent to the Mongolian border (16,17). This bat species has also been found in the southern part of Siberia as well as the western part of Mongolia and Manchuria, which is located in the northern part of China. It therefore seems likely that bats in Mongolia are a possible reservoir of RABVs or related viruses, although further studies are required to confirm this hypothesis.

Acknowledgments The authors would like to thank the staff members of the Mongolian counterparts for their technical supports on the sample collection.

This study was supported by the grant from the Ministry of Health, Labour and Welfare of Japan.

Conflict of interest None to declare.

Table 3. Amino acid alignment of N protein

Virus name	358										367							375									383					389	
MGL 11	R	R	F	F	R	D	E	K	E	L	Q	E	H	E	A	A	E	L	T	K	T	D	V	A	L	A	D	D	G	T	V	H	S	
MGL 17	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
MGL 21	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
MGL 27	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
MGL 28	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
MGL 31	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
MGL 32	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
MGL 26	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
Mongolia 3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
MGL 29	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
MGL 30	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
MGL 20	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
Mongolia 6	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
Mongolia 4	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
MGL 13	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
Mongolia 7	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
MGL 36	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
MGL 33	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
MGL 5	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
MGL 12	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
MGL 23	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
MGL 34	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
MGL 2	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
MGL 24	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
MGL 35	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
MGL 25	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
765w	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
3561d	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
686cattle	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
409f	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
RV259	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
MGL 10	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
9202ALL	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	N	—	
8618POL	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	N	—	
9223FRA	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	N	—	
8653YOU	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	M	—	—	—	—	—	—	—	—	N	—	
9342EST	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	N	—	
9215HON	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	N	—	
8681IRA	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	N	—	
86107YOU	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	N	—	
Eth2003	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	N	—	
8721AFS	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	N	—	
PV	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	N	—	
ERA	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	N	—	
USA 3789	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	N	—	
CVS-11	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	N	—	
304c	—	—	—	—	—	—	—	—	—	—	—	—	Y	T	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	N	—	*
248c	—	—	—	—	—	—	—	—	—	—	—	—	Y	T	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	N	—	*
MGL 22	—	—	—	—	—	—	—	—	—	—	—	—	Y	T	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	N	—	*
857r	—	—	—	—	—	—	—	—	—	—	—	—	Y	T	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	N	—	*
SKRRD0406CC	—	—	—	—	—	—	—	—	—	—	—	—	Y	T	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	N	—	*
8480FX	—	—	—	—	—	—	—	—	—	—	—	—	Y	T	—	—	—	—	—	—	—	—	T	—	—	—	—	—	—	—	—	N	—	*
8684GRO	—	—	—	—	—	—	—	—	—	—	—	—	Y	T	—	—	—	—	—	—	—	—	T	—	—	—	—	—	—	—	—	N	—	*
NNV-RAB-H	—	—	—	—	—	—	—	—	—	—	—	—	Y	T	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	N	—	*
A148	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	L	—	—	—	—	—	—	—	—	N	—	
Xx33	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	L	—	—	—	—	—	—	—	—	N	—	
Sq10	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	L	—	—	—	—	—	—	—	—	N	—	
Yl66	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	L	—	—	—	G	—	—	—	—	N	—	
Kyoto	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	N	—	
Yokohama	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	N	—	
SRL1145	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	—	I	—	—	—	—	—	—	—	—	N	—	
Argentina	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	E	M	—	—	—	—	—	—	—	—	N	—	
Chile	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	—	E	T	—	—	—	—	—	—	—	—	N	—	
9001FRA	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	A	E	T	—	—	—	—	—	—	—	—	N	—	
pehm3230	—	—	—	—	—	—	—	—	—	—	—	—	Y	—	—	—	—	—	—	—	A	E	T	—	—	—	—	—	—	—	—	N	—	
ABL	—	—	—	—	—	—	—	—	—	—	—	D	Y	—	—	—	—	A	M	—	I	—	L	—	—	—	—	—	—	—	—	N	—	

The dashes indicate amino acid identity; the open box shows the antigenic site I.
The gray boxes show the antigenic site IV.
Asterisks show Arctic-like viruses.
389 in boldface type is phosphorylation site.
The virus name was aligned according to Fig 2.

REFERENCES

1. Otgonbaatar, D., Botvinkin, A.D., Tserennorov, D., et al. (2005): Rabies in contagious areas of Mongolia and Russia: history and modern situation. Sci. J. Ctr. Infect. Dis. Nat. Foci., 13, 30–38 (in Mongolian).
2. Dean, D.J., Abelseth, M.K. and Atanasiu, P. (1996): The fluorescent antibody test. p. 88–95. *In* F.X. Meslin., M.M. Kaplan and H. Koprowski (ed), Laboratory Techniques in Rabies. World Health Organization, Geneva.
3. Goto, H., Minamoto, N., Ito, H., et al. (1994): Nucleotide sequence of the nucleoprotein gene of the RC.HL strain of rabies virus, a seed strain used for animal vaccine production in Japan. Virus Genes, 8, 91–97.
4. Smith, J.S. (1995). Rabies virus. p. 997–1003. *In* P.R. Murray, E.J. Baron, M.A. Pfaller, et al. (ed.), Manual of Clinical Microbiology. American Society for Microbiology, Washington, D.C.
5. Saitou, N. and Nei, M. (1987): The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol., 4, 406–425.
6. Larkin, M.A., Blackshields, G., Brown, N.P., et al. (2007): Clustal W and Clustal X version 2.0. Bioinformatics, 23, 2947–2948.
7. Dietzschold, B., Lafon, M., Wang, H., et al. (1987): Localization and immunological characterization of antigenic domains of the rabies virus internal N and NS proteins. Virus Res., 8, 103–125.
8. Dandii, D. (1968): Some problems of epidemiology and specific prophylaxis of rabies in MPR during ten years (1957–1966). p. 250. *In* Measles, Enteroviruses, Rabies. Proceedings of 15th scientific session. Sessions of the Institute of Polio and Viral Encephalitis (in Russian).
9. Angar, D. (2001): Trend of plaque, anthrax and rabies morbidity. Sci. J. Ctr. Infect. Dis. Nat. Foci., 9, 20–24 (in Mongolian).
10. Botvinkin, A.D., Otgonbaatar, D., Tsoodol, S., et al. (2008): Rabies in the Mongolian steppes. Dev. Biol. (Basel), 131, 199–205.
11. Zhang, Y.Z., Xiong, C.L., Zou, Y., et al. (2006): Molecular characterization of rabies virus isolates in China during 2004. Virus Res., 121, 179–188.
12. Liu, Q., Xiong, Y., Luo, T.R., et al. (2007): Molecular epidemiology of rabies in Guangxi Province, south of China. J. Clin. Virol., 39, 295–303.
13. Shao, X.Q., Yan, X.J., Luo, G.L., et al. (2010): Genetic evidence for domestic raccoon dog rabies caused by Arctic-like rabies virus in Inner Mongolia, China. Epidemiol. Infect., 14, 1–7.
14. Kuzmin, I.V., Botvinkin, A.D., McElhinney, L.M., et al. (2004): Molecular epidemiology of terrestrial rabies in the former Soviet Union. J. Wildl. Dis., 40, 617–631.
15. Botvinkin, A.D., Savitsky, V.P., Sidorov, G.N., et al. (1980): About natural focality of rabies in east Zabaikalie. p. 52–61. *In* Modern Methods of Investigation of Natural-Focal Diseases. Leningrad (in Russian).
16. Botvinkin, A.D., Poleschuk, E.M., Kuzmin, I.V., et al. (2003): Novel lyssaviruses isolated from bats in Russia. Emerg. Infect. Dis., 9, 1623–1625.
17. Kuzmin, I.V., Botvinkin, A.D., Poleschuk, E.M., et al. (2006): Bat rabies surveillance in the former Soviet Union. Dev. Biol. (Basel), 125, 273–282.
18. Heaton, P.R., Johnstone, P., McElhinney, L.M., et al. (1997): Heminested PCR assay for detection of six genotypes of rabies and rabies-related viruses. J. Clin. Microbiol., 35, 2762–2766.

MOLECULAR ECOLOGY

Molecular Ecology (2010) 19, 2241–2255 doi: 10.1111/j.1365-294X.2010.04646.x

Rapid diversification by recombination in *Bartonella grahamii* from wild rodents in Asia contrasts with low levels of genomic divergence in Northern Europe and America

EVA C. BERGLUND,* KIRSTEN ELLEGAARD,*[1]¶ FREDRIK GRANBERG,*[1] ZHOUPENG XIE,*[1] SOICHI MARUYAMA,† MICHAEL Y. KOSOY,‡ RICHARD J. BIRTLES§ and SIV G. E. ANDERSSON*

*Department of Molecular Evolution, Evolutionary Biology Centre, Uppsala University, SE-752 36 Uppsala, Sweden, †Laboratory of Veterinary Public Health, College of Bioresource Sciences, Nihon University, 1866 Kameino, Fujisawa, Kanagawa 252-0880, Japan, ‡Division of Vector-Borne Infectious Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Fort Collins, CO 80521, USA, §Infection Biology Group, School of Veterinary Sciences, University of Liverpool, Neston CH64 7TE, UK

Abstract

***Bartonella* is a genus of vector-borne bacteria that infect the red blood cells of mammals, and includes several human-specific and zoonotic pathogens. *Bartonella grahamii* has a wide host range and is one of the most prevalent *Bartonella* species in wild rodents. We studied the population structure, genome content and genome plasticity of a collection of 26 *B. grahamii* isolates from 11 species of wild rodents in seven countries. We found strong geographic patterns, high recombination frequencies and large variations in genome size in *B. grahamii* compared with previously analysed cat- and human-associated *Bartonella* species. The extent of sequence divergence in *B. grahamii* populations was markedly lower in Europe and North America than in Asia, and several recombination events were predicted between the Asian strains. We discuss environmental and demographic factors that may underlie the observed differences.**

Keywords: *Bartonella grahamii*, population structure, phylogeography, recombination

Received 2 January 2010; revision received 15 March 2010; accepted 22 March 2010

Introduction

Rodents represent a major reservoir for bacterial infectious diseases; however, we know remarkably little about the ecology and evolution of rodent-adapted bacterial populations. *Bartonella* is a genus of vector-borne bacteria that is particularly abundant in rodent populations all over the world, with an overall prevalence ranging from 6% in the Greater Jakarta area (Winoto *et al.* 2005) to 70% in the Russian far east (Mediannikov *et al.* 2005). Of the almost 30 identified species and subspecies of *Bartonella*, more than half have been isolated from rodents and several of these have been associated with human disease (Boulouis *et al.* 2005).

Bartonella grahamii is one of the most prevalent *Bartonella* species in wild rodents, and has been associated with two cases of human neuroretinitis (Kerkhoff *et al.* 1999; Serratrice *et al.* 2003). It is transmitted by the rodent flea *Ctenophthalmus nobilis* (Bown *et al.* 2004), and has been isolated from the rodent genera *Myodes, Apodemus, Microtus, Mus, Dryomus, Arvicola* and *Rattus* in many countries (Hsieh *et al.* 2009; Inoue *et al.* 2009), indicating a wide host range and a global distribution pattern. In rodents and other natural hosts, *Bartonella* cause asymptomatic infections of red blood cells. The genus also contains two pathogens adapted to humans: *Bartonella bacilliformis*, which causes Carrion's disease, and *Bartonella quintana*, the agent of trench fever. There

Correspondence: Siv Andersson, Fax: +(46) 18 4716404; E-mail: siv.andersson@ebc.uu.se

[1]These authors contributed equally to this work.

are additionally several zoonotic pathogens, for example, the feline-adapted *Bartonella henselae*, which causes cat-scratch disease in incidentally infected humans.

The genome of *B. grahamii* strain as4aup, isolated from a wood mouse (*Apodemus sylvaticus*) in central Sweden (Holmberg *et al.* 2003), was recently sequenced, and found to consist of a 2.3 Mb chromosome and a 28 kb plasmid (Berglund *et al.* 2009). Sixteen genomic islands (BgGI 1–16), mainly containing horizontally transferred genes for surface proteins, secretion systems and phage genes, were identified by comparison to the previously published *Bartonella* genomes (Alsmark *et al.* 2004; Saenz *et al.* 2007; Berglund *et al.* 2009). Among the functionally most important secretion systems are the VirB and Trw type IV secretion systems (T4SSs) and the type V secretion system (T5SS) trimeric autotransporter BadA, which have been shown in other *Bartonella* species to be required for binding to and infect endothelial cells and erythrocytes (Schulein & Dehio 2002; Seubert *et al.* 2003b; Riess *et al.* 2004; Schmid *et al.* 2004; Zhang *et al.* 2004; Schulein *et al.* 2005; Schmid *et al.* 2006; Kaiser *et al.* 2008). Many of the genomic islands are clustered in a genomic region that is amplified by run-off replication and packaged into phage particles (Berglund *et al.* 2009). As this phage packages bacterial DNA randomly it was described as a gene transfer agent (GTA) (Berglund *et al.* 2009).

Previous multilocus sequence typing (MLST) based on six house-keeping genes resulted in 10–16 alleles per locus in 31 *B. grahamii* strains (Inoue *et al.* 2009). Phylogenetic analysis revealed a clustering according to geographic origin, with the Asian isolates clearly distinguished from the American and European (Inoue *et al.* 2009). The Japanese strains were split into two subgroups, one of which was suggested to originate from China and the other from Far Eastern Russia (Inoue *et al.* 2009). In comparison, MLST of more than 200 *B. henselae* strains identified only 14 STs (Sequence Types), with up to four alleles per locus, and no obvious correlation with geographic origin or host species (Iredell *et al.* 2003; Lindroos *et al.* 2006; Arvand *et al.* 2007), although variable number tandem repeat work has suggested that human-infecting strains are overrepresented among strains of genotype II (Bouchouicha *et al.* 2009). Sequencing of highly variable spacer regions (MST) yielded 39 variants in 126 cat isolates (Li *et al.* 2006) and 16 variants in 75 human isolates (Li *et al.* 2007). Microarray comparative genome hybridizations of a global collection of *B. henselae* strains revealed no differences in gene content between human and feline isolates, nor was a geographical pattern evident (Lindroos *et al.* 2006). In *B. quintana*, MST based on as many as 34 spacer regions identified only seven sequence variants in 74 strains (Foucault *et al.* 2005; Woolley *et al.* 2007). This low diversity was attributed to a bottleneck associated with a recent host switch (Foucault *et al.* 2005).

To learn more about the emergence of infectious diseases, it is important to examine the genomic diversity of bacterial species adapted to wild animals, and the evolutionary, ecological and demographic factors that shape this diversity. The aim of this study was to determine the diversity of the global population of *B. grahamii* infecting wild rodents on a whole-genome scale. We found a much higher genomic diversity in Asia than in Europe and North America, and a larger genome plasticity in *B. grahamii* compared with previous results from *B. henselae* and *B. quintana*. We discuss potential environmental and demographic factors that may underlie the observed differences.

Materials and methods

Bacterial strains and DNA isolation

The *B. grahamii* strains used in this study are listed in Table 1. Bacteria were grown for 4–5 days on chocolate agar plates and DNA was extracted as described previously (Lindroos *et al.* 2005).

PCR amplification and phylogenetic inference

Primers used for PCR amplification are listed in Table 2. In addition, a 379-bp fragment of the *gltA* gene was amplified with primers 5′-GGGGACCAGCTCATGGTGG-3′ and 5′-AATGCAAAAAGAACAGTAAACA-3′. PCRs were performed as described previously (Lindroos *et al.* 2006). Sequences were assembled and edited with Phred, Phrap and Consed (Ewing & Green 1998; Ewing *et al.* 1998; Gordon *et al.* 1998) and trimmed with Lucy (Chou & Holmes 2001). The homologous regions were also sequenced from *Bartonella vinsonii berkhoffii* strain Winnie (Kordick & Breitschwerdt 1998), and extracted from the previously sequenced genomes of *B. grahamii* as4aup and *Bartonella tribocorum* IBS 325 (GenBank entries NC_012846 and NC_010161, respectively).

For loci including a spacer region, the gene borders were identified manually and the sequence was cut at these positions. The different parts were aligned with Prank (Loytynoja & Goldman 2008), concatenated and overhanging ends were cut for each locus. Genic parts were aligned with the translate option. Phylogenetic trees were inferred for each locus, the concatenation of all loci, the concatenation of all gene sequences and the concatenation of all loci after removing all sites with gaps, with RAxML (Stamatakis 2006), substitution model GTRMIX, 1000 bootstrap replicates. A phylogenetic tree was also inferred using ClonalFrame (Didelot & Falush 2007), with the number of burn-in iterations

Table 1 Country of origin, host and date of isolation of the *Bartonella grahamii* strains analysed in this study

Strain	Origin*	Host animal†	Isolation	Reference
as4aup	Sweden	*Ap. sylvaticus*	Sep-1999	Holmberg *et al.* (2003)
AL1714yn	China	*Ap. latronum*	Sep-2000	Ying *et al.* (2002)
Ac1733yn	China	*Ap. chevrieri*	Sep-2000	Ying *et al.* (2002)
Cg4224alb	Canada	*My. gapperi*	Jul-2002	Inoue *et al.* (2009)
Cg4227alb	Canada	*My. gapperi*	Jul-2002	Inoue *et al.* (2009)
Cg4228alb	Canada	*My. gapperi*	Jul-2002	Inoue *et al.* (2009)
Cg4263alb	Canada	*My. gapperi*	Aug-2002	Inoue *et al.* (2009)
Cg4285alb	Canada	*My. gapperi*	Aug-2002	Inoue *et al.* (2009)
Mo12494sd	USA	*Mi. ochrogaster*	Aug-2005	Inoue *et al.* (2009)
Mo12658sd	USA	*Mi. ochrogaster*	Aug-2005	Inoue *et al.* (2009)
PTZA 30/3	Russia	*Ap. flavicollis*	Aug-2006	Markov *et al.* (2006)
PTZB 29/18	Russia	*Ap. uralensis*	Aug-2006	Markov *et al.* (2006)
Hokkaido 29-1	Japan	*Ap. speciosus*	Aug-2005	Inoue *et al.* (2008)
Nagano 14-1	Japan	*Ap. speciosus*	Sep-2003	Inoue *et al.* (2008)
Fuji 4-1	Japan	*Ap. speciosus*	Aug-2003	Inoue *et al.* (2008)
Ehime 5-1	Japan	*Ap. speciosus*	May-2006	Inoue *et al.* (2008)
Nakanoshima 39-1	Japan	*Ap. speciosus*	Oct-2004	Inoue *et al.* (2008)
MAC29	UK	*Ar. terrestris*	Jun-2002	This study
WM11	UK	*Ap. sylvaticus*	Sep-1999	This study
V2	UK	*My. glareolus*	Sep-1999	Birtles *et al.* (1995)
C162	UK	*Mi. agrestis*	Sep-2002	Telfer *et al.* (2007)
J142	UK	*Mi. agrestis*	Sep-2003	Telfer *et al.* (2007)
C066	UK	*Mi. agrestis*	Sep-2002	Telfer *et al.* (2007)
R170	UK	*Mi. agrestis*	Sep-2004	Telfer *et al.* (2007)
J019	UK	*Mi. agrestis*	Sep-2003	Telfer *et al.* (2007)
S116	UK	*Mi. agrestis*	Sep-2004	Telfer *et al.* (2007)

*The collection sites within each country were: as4aup, Håtunaholm in central Sweden; Chinese strains, Yunnan; Canadian strains, Alberta; U.S. strains, South Dakota; Russian strains, Moscow region; Hokkaido 29-1, Hokkaido; Nagano 14-1, Nagano; Fuji 4-1, Shizuoka; Ehime 5-1, Ehime; Nakanoshima 39-1, Kagoshima; MAC29, Reisa Mhic Phaiden; WM11, Brimstage; V2, Shropshire; remaining UK strains, Kielder.
†Host species are abbreviated as follows: *Ap—Apodemus, Ar—Arvicola, Mi—Microtus, My—Myodes* (formerly known as *Clethrionomys*).

set to 50 000, and the number of MCMC iterations after the burn-in period set to 200 000. Two runs were performed and the convergence was tested with the Gelman-Rubin statistic. All parameters were found to be equal to or below 1.2, except delta, which was 2.5.

Phylogenetic trees for *gltA* and for the concatenation of *rrs*, *gltA* and *groEL* sequences from (Inoue *et al.* 2009) were inferred both with RAxML as described above, and with the NJ method with Jukes-Cantor parameters.

Pairwise sequence identities were calculated based on the complete concatenated alignment, excluding sites with gaps in each comparison. Sequence identities between and within groups were defined as the average of all relevant pairwise comparisons.

Detection of recombination

The SuperNetwork method in SplitsTree4 (Huson 1998; Huson & Bryant 2006) with default settings was used to detect conflicts between phylogenetic trees for individual loci. Each locus was tested for recombination using the methods implemented in RDP3 (Martin *et al.* 2005) with default settings. As all predicted recombination events took place between the Asian strains, the analysis was repeated with only these strains, in order to get a better alignment.

The MLST data set from (Inoue *et al.* 2009) was used to estimate the relative contribution of recombination and mutation in *B. grahamii* using the software ClonalFrame (Didelot & Falush 2007). Based on recommendations in (Vos & Didelot 2009), *rrs* sequences were excluded and only one strain of each sequence type was included. The mutational rate θ was set to Watterson's θ and 200 000 iterations were performed, of which 100 000 were discarded. Two runs were performed, and the convergence was tested with the Gelman-Rubin statistic and all parameters were found to be below 1.1.

Microarray comparative genome hybridizations

Microarrays were designed and manufactured as reported previously (Berglund *et al.* 2009) and cross-linked at

Table 2 Characteristics of the sequenced regions and primers used for PCR

No	Locus*	Position† and sequence (5′–3′) of primers	Size (bp)‡	No. of alleles
1	Bgr_05150 (*cycK*)	622901 GCGCTGCTTACTTTTTCCC 623480 TCTTTCCCCATAGATCCGC	580	14
2	Bgr_00610 (*batR*)	82636 CAATGGTGCGATCATCTACG 83190 CGTCTTTATCTTTTGCGCTTG	535	11
3	Bgr_00430	61980 ATGCACAGCTTTCTGGTCG 62569 TCCTGCAATAAAACCATTTGC	590	13
4	Bgr_03900 (*aldA*) Bgr_03910 (*ftsK1*)	473807 TGTTTTCCATTTTTGAAACGC 474531 CTTCTCTTGATGCACCTTTCG	725 (469)	15
5	Bgr_05600 (*cspA*) Bgr_05610 (*carB*)	679952 TGAATCCGAAACCTTTTGTTG 680568 TTGGCTTTTCTGTTGTCGC	617 (245)	9
6	Bgr_06060 (*pssA*) Bgr_06070	734202 TAGGCGCTCTTGGTTTGG 734928 TGGACGAGCCATTCTGTTATC	727 (70)	13
7	Bgr_06380 (*uvrC*) Bgr_06390	776004 CAATCATCCGGTAAACCCC 776771 TGAAATGCGTATCCGAAAAAG	768 (397)	16
8	Bgr_11560 (*maeB2*) Bgr_11570 (*acpP2*)	1305872 TTTTCGTGATCGTGTTTTTCC 1306594 GCCTGTTTTAAGGCAACGAG	723 (189)	10
9	Bgr_18510 (*pgk*) Bgr_18520 (*gap*)	2129539 ACCCCATCACTGCTTCCTC 2130143 CGCGTTTTGGTTTGGTATG	605 (514)	13
10	Bgr_19460 (*dnaJ2*) Bgr_19470 (*cobS*)	2239347 GCAAAGATTCGCTCTGGAAC 2240054 ATAGCCAGAAACCATCACACG	708 (201)	11
11	Bgr_19730 Bgr_19740	2273643 CAAGGATTTCGTGCCCC 2274453 TTATGTTTCGCGGTTGTTCTC	811 (144)	16

*The locus_tag and gene name, if any, for the sequenced regions in *B. grahamii* as4aup. For intergenic regions, both the upstream and the downstream gene are shown.
†Genomic position of primers refer to the *B. grahamii* as4aup genome.
‡Size of the total PCR product refer to the *B. grahamii* as4aup genome. Within parentheses is the spacer size, if any.

250 mJ/cm^2. Genomic DNA from *B. grahamii* as4aup was used as reference, and two or three hybridizations were performed for each strain. Prehybridization, DNA labelling, hybridization, scanning and image analysis were performed as described previously (Berglund *et al.* 2009). The channel used for the reference strain is referred to as Ch1 and the channel used for the test strain is referred to as Ch2, and spots were filtered as described previously (Berglund *et al.* 2010). *M*-values were computed as $\log_2$(Ch2/Ch1) and normalization was performed as described previously (Lindroos *et al.* 2006). Median *M*-values of all replicate spots and arrays were computed for each strain. Probes were ordered according to the position in *B. grahamii* as4aup, classified as absent ($M \leq -2$), uncertain ($-2 > M \leq -1$) or present ($-1 > M$) and sorted into regions of contiguous probes with the same status as described previously (Lindroos *et al.* 2006). A maximum parsimony tree was inferred from this data with the PHYLIP package (Felsenstein 1989), with 100 bootstrap replicates.

Pulsed-field gel electrophoresis

Bacteria grown on chocolate agar plates for 4–5 days were harvested, suspended and washed twice in TNE buffer (10 mM Tris pH 8.0, 150 mM NaCl and 1 mM EDTA). The bacterial suspension was added to an equal volume of 2% SeaPlaque GTG agarose (Cambrex BioScience, East Rutherford, NJ) in 0.5 × Tris–borate–EDTA (TBE) buffer, and plugs were made using reusable moulds. After solidification, the agarose plugs were transferred into lysozyme solution (10 mM Tris [pH 8], 1 mM EDTA, 100 mM NaCl, 0.5% *N*-lauroyl sarcosine, RNAse A [0.5 mg/mL], Ready-Lyse lysozyme [50 kU/mL, Epicentre Technologies, Madison, WI, USA]) and incubated at 37 °C overnight. The plugs were rinsed with sterile water and incubated in proteinase K solution (10 mM Tris [pH 8], 100 mM EDTA, 1% SDS, 1% *N*-lauroyl sarcosine, 0.25% Triton X-100, proteinase K [1 mg/mL]) at 50 °C overnight. Subsequent to three washes in TE buffer (10 mM Tris [pH 8], 0.5 mM EDTA) at 4°C for 30 min, any remaining proteinase K was inactivated by incubation in TE buffer supplemented with 0.4 mg/mL phenylmethylsulfonyl fluoride at 50°C for 1 h. The ready plugs were either stored in 0.2 M EDTA at 4°C or equilibrated in TE buffer. Enzymatic digestions with 10 U of *Not*I or *Asc*I (New England Biolabs, Ipswich, MA) were performed as previously described (Lindroos *et al.* 2005). The DNA fragments were separated by electrophoresis

in 1.1% agarose (SeaKem Gold; Cambrex Bio Science) in 0.5 × TBE buffer in a GenNavigator System apparatus (Amersham Biosciences, Uppsala, Sweden) at 14 °C and 5.6 V/cm, for a total of 65 h with switch times ramped from 5 to 150 s. The gels were stained with ethidium bromide (0.5 μg/mL) at room temperature for 1 h, destained in 0.5 × TBE buffer for 30 min, and visualized under UV light. The sizes of the fragments were estimated using pulsed-field gel electrophoresis (PFGE) λ-ladder, Low Range PFG marker, and Yeast Chromosome PFG marker (New England Biolabs).

Isolation of phage particles, extraction of phage DNA and gel electrophoresis of phage DNA

Phage particles were isolated and phage DNA was extracted and amplified with the REPLI-g Mini kit (Qiagen, Valencia, CA) as described previously (Berglund *et al.* 2009). PFGE was performed as described above with the following modifications: phage particles in PBS were directly moulded into plugs, no digestion was performed, and the DNA fragments were separated by electrophoresis at 5 V/cm for a total of 29 h.

Nucleotide sequence and microarray accession numbers

The novel sequences have been submitted to EMBL with accession numbers shown in Table S1 (Supporting Information). The microarray data have been deposited in the ArrayExpress database of the European Bioinformatics Institute under the accession numbers A-MEXP-1576 for the array design and E-TABM-875 for the experimental data.

Results

We have analysed 26 strains of *B. grahamii* isolated from mice and voles of four genera (*Apodemus*, *Myodes*, *Microtus* and *Arvicola*) that were captured in seven countries in North America, Europe and Asia (Table 1). These strains were selected from a much larger collection, in order to represent most of the diversity. Studies of *Bartonella* prevalence in rodent populations in continents that are not included in our study (Africa, Australia and South America) have not yielded any *B. grahamii* isolates (Birtles *et al.* 1999; Pretorius *et al.* 2004; Gundi *et al.* 2009). We thus believe that we have a good representation of the known diversity of *B. grahamii*. The relationships of the 26 strains were inferred by sequence analysis of eleven loci, their gene contents were investigated by microarray hybridizations using the sequenced *B. grahamii* strain as4aup as reference and the genome sizes of all strains were estimated by pulsed-field gel electrophoresis.

Geographic population structure

We first sequenced eleven loci (three genes and eight intergenic regions) in all strains, with the number of alleles ranging from 9 to 16 per locus (Table 2). The alignments for all loci were concatenated to a total length of *c.* 7 kb, and maximum likelihood phylogenetic analysis revealed a clustering of *B. grahamii* strains according to geographic origin (Fig. 1). The tree had two major clades: one containing the Asian strains and the other the European and American strains (Fig. 1).

A closer inspection revealed that the Asian strains were divided into a Japanese and a Chinese group, and that all Canadian strains formed a single clade, as did also all European strains. One European subclade, here called Europe 1, contained the Swedish, Russian and two British strains (V2 and WM11) and the other subclade, here called Europe 2, the remaining British isolates. The bootstrap support for each of the described groups was 100%. Finally, the results indicated that the U.S. strain Mo12658sd represents the earliest diverging branch within the European–American clade, with 92% bootstrap support for the cohesion of all other strains to the exclusion of this isolate. As some intergenic regions contained long indels, we also inferred maximum likelihood phylogenies based on only the concatenated genic regions (Fig. S1, Supporting Information) and from the original alignment after removal of all sites with gaps (Fig. S2, Supporting Information). The major clades described above were highly supported also in these two trees.

For comparison with the previously described diversity of *B. grahamii* we inferred a phylogeny of the *gltA* gene, which includes published sequences from other countries (Fig. S3, Supporting Information). Because of the low level of sequence information contained in this gene, the bootstrap support values for the different clades were low, and we could not root the tree reliably. Despite these limitations, the tree reveals that there are no major clades within *B. grahamii* that are absent from our data set. However, the diversity within Europe is larger than revealed by the isolates included in this study, with two strains from Greece and Poland clustering with the other European strains whereas two others cluster with the American strains.

Recombination is extensive in the Asian strains

We observed a high degree of sequence divergence among the Chinese and Japanese strains, with an average pairwise sequence identity of only 95.9% excluding sites with gaps (Table S2, Supporting Information). In contrast, the average sequence identity between European and American strains was as high as 98.2% (Table S2, Supporting Information). Additionally, large inser-

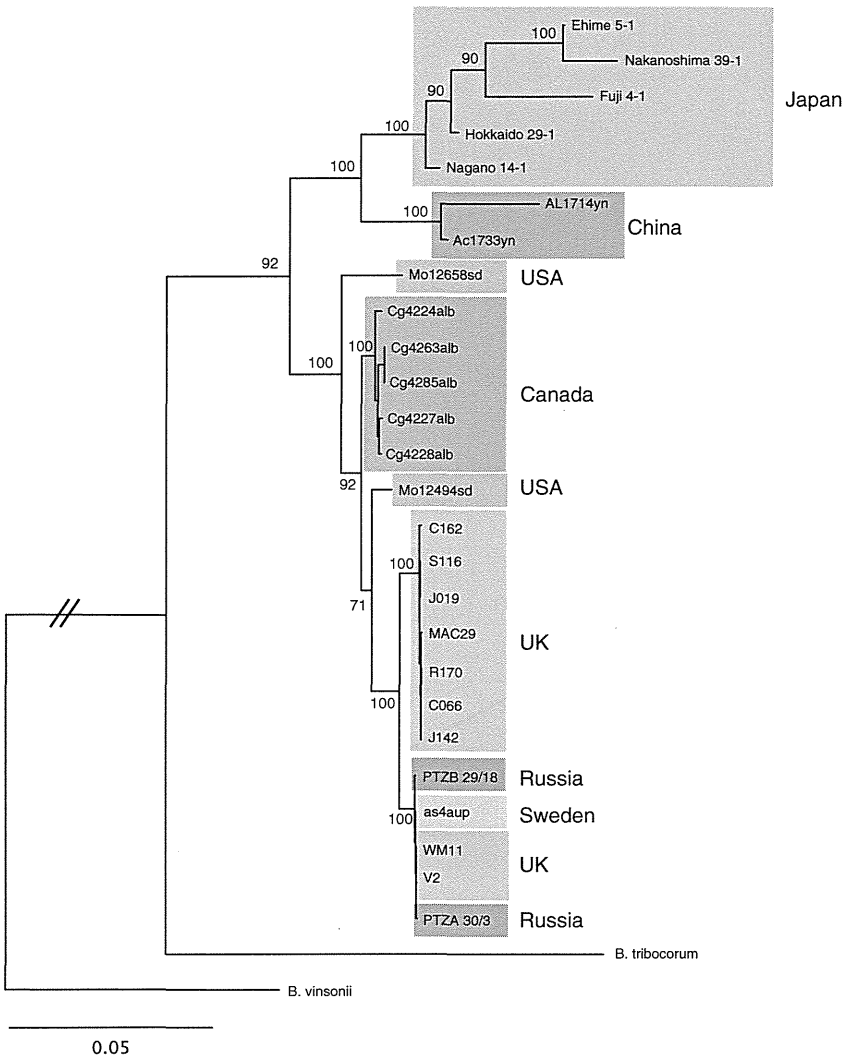

Fig. 1 Phylogenetic tree inferred from a concatenated alignment of all sequenced loci. Branch lengths are according to maximum likelihood, and bootstrap support values for the major clades are shown. Strains are colour-coded according to geographic origin. The branch length of *Bartonella vinsonii* is not according to scale.

tion–deletion events were identified in three intergenic regions in the Asian strains: Fuji 4-1 and AL1714yn shared a deletion in locus 9, Nagano 14-1 and Nakanoshima 39-1 had an insertion in locus 4, and all Asian strains had an insertion in locus 11. Both insertions included short segments with similarity to phage genes in *B. tribocorum* (Btr_0462, E < e^{-8} and Btr_0221, E < e^{-6}, respectively). In the European–American strains, no large indels were observed.

To investigate whether each sequenced region supported the same tree topology, we inferred a phylogenetic tree for each locus, and computed a super-network from these trees using SplitsTree4. Although the network confirmed that the population structure of the *B. grahamii* strains is mostly a function of geography, conflicts were abundant within the Asian clade (Fig. 2). Manual inspection of each tree showed that the conflicts were often well supported by bootstrap values (Fig. 3), indicative of recombination between the Asian strains. Recombination also appears to occur between *Bartonella* species, exemplified by the clustering of the Japanese strain Fuji 4-1 and *B. tribocorum* with a bootstrap support value of 92% in locus 5 (Fig. S4, Supporting Information). For comparison, we also computed a super-network of a MST data set from *B. henselae* (Li *et al.* 2006), which revealed no conflicts between trees (data not shown).

Recombination events among Asian strains were also predicted within five of the sequenced regions (Table S3, Supporting Information), by at least three different algorithms in RDP3 ($P < 0.05$). The predicted recombined segments spanned *c.* 100–600 bp and occurred in genes, spacers as well as across borders of genes and spacers (Table S3, Supporting Information). Of the five loci in which recombination events were detected, four were associated with phage gene remnants (loci 3, 4, 10 and 11). No remnants of phage genes were identified in any of the six loci at which no recombination events were detected.

For comparison, we also computed a super-network and applied RDP3 to the *B. grahamii* MLST data set from (Inoue *et al.* 2009). The results showed that

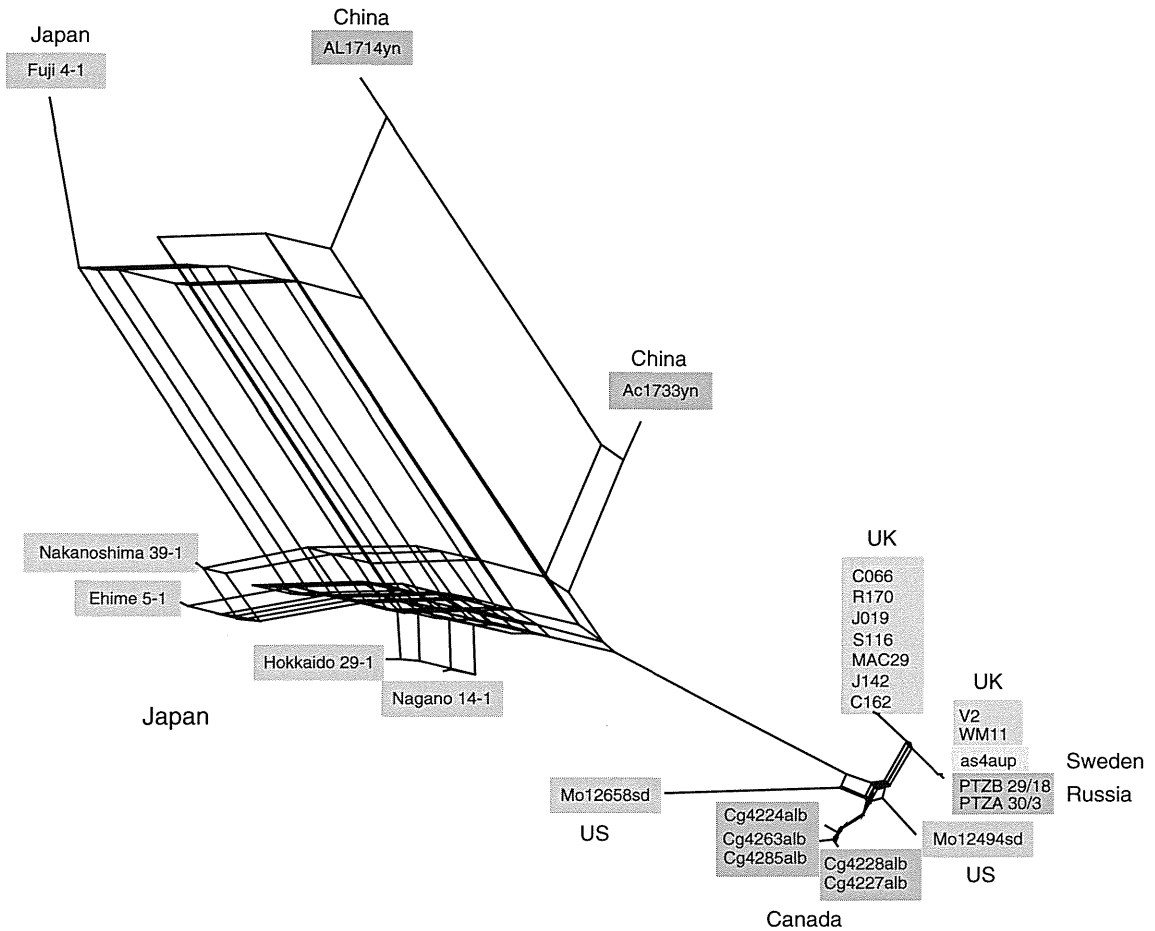

Fig. 2 Super-network of the analysed *Bartonella grahamii* strains, based on the individual trees for each sequenced locus. Conflicts between individual tree topologies, represented as rectangular boxes in the network, are abundant in the Asian group. Strains are colour-coded according to geographic orgin.

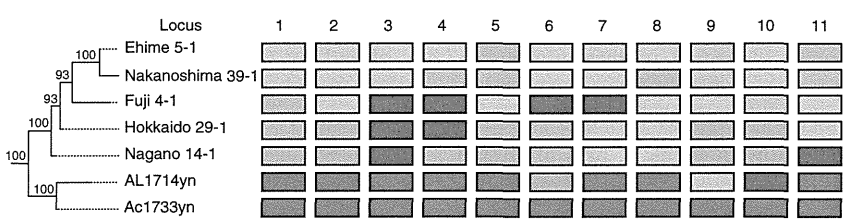

Fig. 3 Schematic representation of highly supported subclades of Asian strains in individual phylogenies of each sequenced locus. The phylogeny of the Asian strains (from Fig. 1) is shown to the left. Each locus (numbered as in Table 2) is represented by one column to the right, and clades in the phylogeny of that locus that are identical or that have at least 85% bootstrap support, are marked in the same colour. Groups that are coloured in different shades of green or purple indicate that these subclades also cluster together with at least 85% support.

conflicts were abundant in the Asian group, and recombination events were detected in three of six genes (*ribC*, *rpoB* and *ftsZ*) (Table S4, Supporting Information). The relative contribution of recombination vs. nucleotide substitution (the *r/m* ratio) in this data set was estimated to 1.7, which means that the probability that a sequence polymorphism is a result of recombination is slightly higher than the probability for nucleotide substitutions. Altogether, these results suggest that recombination is frequent in the Asian population of *B. grahamii*.

To examine possible influences of recombination on the tree topology, and whether the high bootstrap support values in Fig. 1 underestimates the uncertainty, we inferred a phylogenetic tree of the complete data set using ClonalFrame, which takes recombination into

account when building trees. The analysis resulted in a tree with the same major clades, all with high posterior probabilities, as the maximum likelihood tree in Fig. 1 (Fig. S5, Supporting Information). The main difference between the topology of our phylogenetic tree and that inferred by Inoue *et al.* is that all Japanese strains grouped together in our tree whereas they were separated into two groups in theirs (Inoue *et al.* 2008; Inoue *et al.* 2009). Interestingly, in a phylogenetic tree inferred from only the three genes where no recombination events were detected in Inoue's data set, all Japanese strains clustered together in one clade (data not shown). In our data set, removal of the recombined loci did not influence the topology. It thus appears as if the population structure of *B. grahamii* can be reliably inferred if enough loci with enough variation are sequenced, even though individual trees may present conflicts.

CGH data reveal gene content variations in genomic islands and plasmids

Pulsed-field gel electrophoresis (PFGE) with the *Not*I restriction enzyme revealed a great variability in genome structure, with 25 distinct patterns in the 26 strains (Fig. S6, Supporting Information). The genome sizes were estimated to range from less than 2 Mb in the Canadian strains Cg4263alb and Cg4285alb to more than 2.5 Mb in Hokkaido 29-1 (Fig. S6, Supporting Information).

We analysed differences in gene content by comparative genome hybridizations to a microarray designed based on the genome of *B. grahamii* strain as4aup from Sweden. The array contains 4438 probes, which cover 96% of the 1768 annotated genes in this genome. The results showed that most variations in gene content are concentrated to the previously described genomic islands, including the *badA* region in BgGI 1, the prophage regions in BgGI 2 and BgGI 4, the membrane proteins in BgGI 3, the phage-related BgGI 5, and the Vbh T4SS in BgGI 16 (Fig. 4). Variability was also recorded in the so-called *fha*-repeat, which is present in the islands BgGI 4, BgGI 7, BgGI 10, BgGI 11 and BgGI 14 (Fig. 4). The plasmid pBGR3, which encodes a copy of the Vbh T4SS, showed a highly scattered distribution pattern, and appeared to be present in at least one strain from each continent (Fig. S7, Supporting Information). Major deletions were observed in the Canadian strains Cg4263alb and Cg4285alb, in agreement with the small genome size estimate from the PFGE.

In the Asian strains, low hybridization signals were observed for many more genes, including core genes (Fig. 4). To check whether the low signal in these genes could be an effect of high levels of sequence divergence relative to the Swedish reference genome rather than gene loss, we estimated the sequence conservation across the genome by calculating the percentage identity of each probe to the genome of *B. tribocorum*. The results confirmed that many of the core genes that showed a low hybridization signal in the Asian strains are among the most divergent between *B. grahamii* and *B. tribocorum* (Fig. 4).

Clustering patterns of the European–American strains by CGH data

To examine the extent to which gene contents correlate with the population structure as inferred from the sequence data, we divided the genome into regions of present or absent ($M \leq -2$) probes, and inferred a maximum parsimony tree from this data (Fig. 5). Because of their high sequence divergence, the Asian strains were excluded from this analysis. With the exception of the British strain C162, which clusters with the Canadian strains, the resulting tree topology is largely in agreement with the sequence-based tree topology, with a major division between American and European strains, and the same two subgroups of European strains.

As the sequenced strain belongs to Europe 1, it is not surprising that the strains with the fewest number of losses belong to this group. These strains are unique in that the *badA* region showed no decrease in signal, and the few observed losses for this group were mostly confined to phage genes and the *fha*-repeat. In contrast, strains belonging to Europe 2 displayed decreased hybridization signals at several sites, including the *badA* region, BgGI 3, the repeated genes in the *trw* locus for a T4SS as well as in the *bep* genes located immediately downstream of the *virB* locus for another T4SS. Partial losses of the *fha*-repeat were observed in most Europe 2 strains (Fig. S8, Supporting Information).

In comparison, the Canadian strains showed a much more extensive loss of genes, mainly within the *fha*-repeat and in prophage I (Fig. 4). An almost complete loss of the *fha*-repeat was also found in the British strain C162, and this might be the reason why this strain clusters with the Canadian strains. The Canadian strain Cg4228alb is the earliest diverging in the Canadian clade and atypical in that it only showed partial losses of the *fha*-repeat, thus resembling the Europe 2 isolates over these segments (Fig. S8, Supporting Information). The two U.S. strains, which did not cluster together in the sequence-based tree, also displayed major differences in terms of gene content. Mo12658sd was most divergent on the sequence level and also showed more gene loss, including, for example, BgGI 3 and BgGI 5.

Two sets of strains that were identical in the sequence analysis (Cg4263alb and Cg4285alb; C066, J142 and R170) also clustered together in Fig. 5. Thus, the relationships

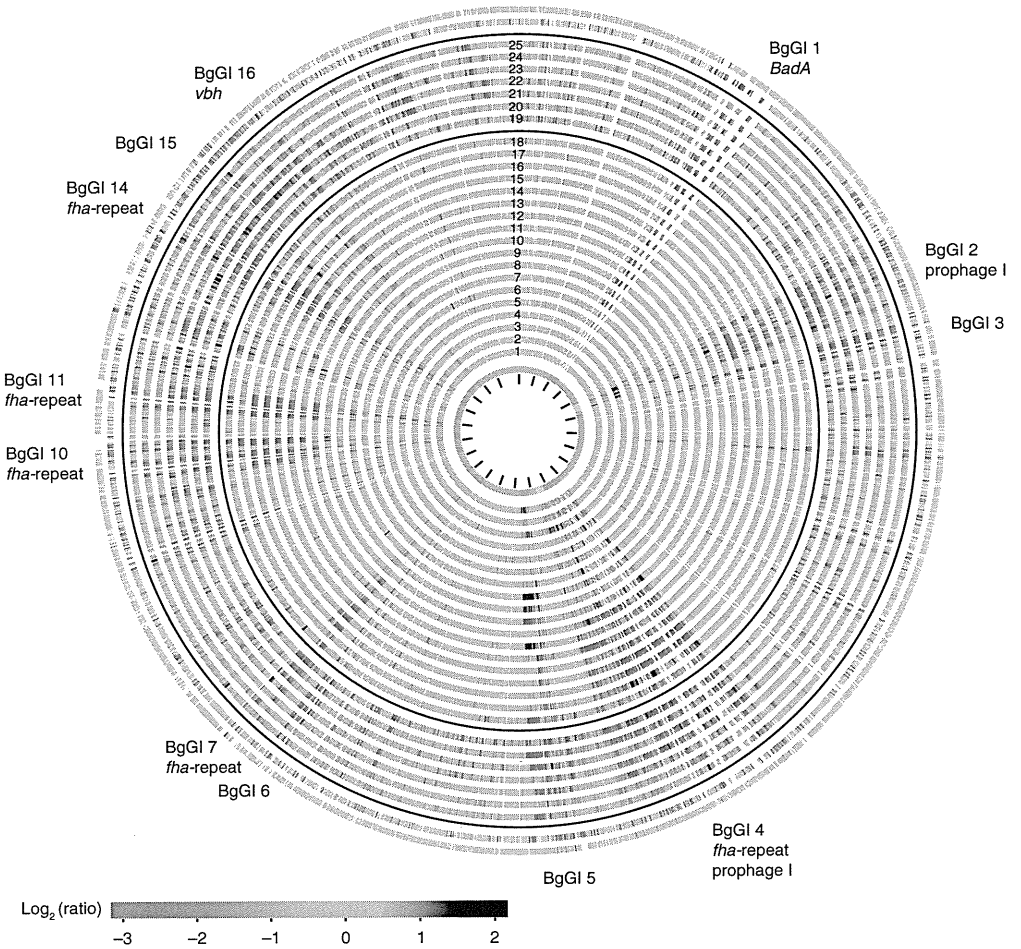

Fig. 4 Circular representation of the microarray data for the *Bartonella grahamii* chromosome, ordered according to the genome position in the sequenced strain. The innermost circle shows the genomic islands in magenta and prophages in yellow. The black lines within this circle indicate the genome position, with 100 kb between each line. Each other circle, except the two outermost, shows the microarray results for one strain, with the colour corresponding to the hybridization signal relative to as4aup (red for lower and blue for higher signal). The strains are: 1, V2; 2, WM11; 3, PTZA 30/3; 4, PTZB 29/14; 5, J142; 6, C066; 7, R170; 8, J019; 9, S116; 10, MAC29; 11, C162; 12, Cg4228alb; 13, Cg4224alb; 14, Cg4227alb; 15, Cg4263alb; 16, Cg4285alb; 17, Mo12494sd; 18, Mo12658sd; 19, Ac1733yn; 20, AL1714yn; 21, Ehime 5-1; 22, Nakanoshima 39-1; 23, Nagano 14-1; 24, Fuji 4-1; 25, Hokkaido 29-1. The second outermost circle shows the nucleotide identity in a global alignment of each probe to the *Bartonella tribocorum* genome, with probes with less than 75% identity in red. The 75% cutoff is based on previous control hybridizations with *B. henselae* and *B. quintana* (Lindroos *et al.* 2005). The outermost circle shows the predicted presence (grey) (Blastp, cutoff $E < e^{-10}$) or absence (red) of each *B. grahamii* gene in the genome of *B. tribocorum*. The black circles show the border between European-American strains, Asian strains and predictions from the *B. tribocorum* genome.

of strains as inferred by their deletion–insertion patterns were broadly consistent with the sequence-based tree topology, and both support relationships governed by geographic vicinity. There were however several examples of independent gains or losses of genes in strains from different countries, for example, the *fha*-repeat (lost from the British strain C162 and most Canadian strains), BgGI 3 (at least partial losses in all Europe two strains except J019 and all American strains except Mo12494sd) and BgGI 5 (lost from the British isolate MAC29 and the US isolate Mo12658sd). Most of the variably present genes are repeated in the genome of the sequenced strain, and the lower microarray signal can be due either to a complete loss, a lower copy number or a high sequence divergence.

Duplication and replication of phage-related genomic islands

Several strains displayed increased signals in prophage I and around BgGI 15, where the putative origin of run-off replication is located. The island BgGI 5, which contains phage genes homologous to those that encode the newly described *Bartonella* GTA, also displayed

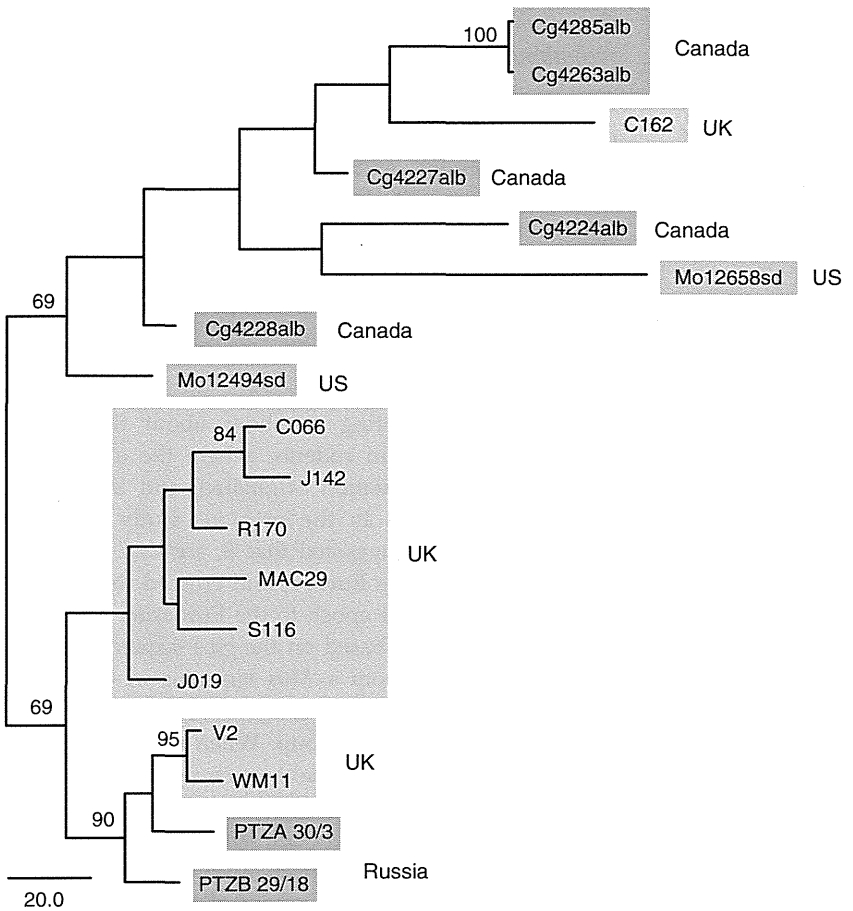

Fig. 5 Maximum parsimony tree of European and American strains based on microarray data. Bootstrap support values over 65 are shown. Strains are colour-coded according to geographic orgin.

increased signal in several strains (Fig. 5). To investigate whether this island is replicated and packaged into phage particles we isolated phage DNA from the strains Cg4228alb and PTZA 30/3 and hybridized to the array against total bacterial DNA from the same strain. The results revealed run-off replication and a slight over-representation of prophage I in the phage DNA, but no increase in signal in BgGI 5. Gel electrophoresis of the phage DNA also did not reveal any other bands than the previously observed 14 and 45 kb bands (data not shown). These observations suggest that the increase in signal is more likely due to duplication(s) than replication of BgGI 5. There was no sign of gene duplication in Hokkaido 29-1, which had an estimated genome size of more than 2.5 Mb, suggesting that additional genes are present in this strain.

Discussion

Influence of host ecology and population structures on the genomic diversity of Bartonella populations

Our study of the genomic diversity in a global *B. grahamii* population revealed a stronger geographic pattern and higher levels of sequence and gene content divergence than has been observed previously in *B. henselae* and *B. quintana* (Iredell *et al.* 2003; Foucault *et al.* 2005; Li *et al.* 2006, 2007; Lindroos *et al.* 2006; Arvand *et al.* 2007). The strong geographic pattern of *B. grahamii* is in agreement with previous MLST analysis (Inoue *et al.* 2009), and may be related to different migration patterns of the hosts. All *B. grahamii* isolates analysed in this study were taken from wild mice and voles, which migrate slowly and typically have a home area of less than 1 ha (Corbet & Harris 1991). In contrast, most of the previously analysed *B. henselae* isolates were derived from domestic cats or humans, which travel much longer distances. Although a geographic pattern is also evident among different cat breeds (Lipinski *et al.* 2008), the presence of the same breeds in different parts of the world has obscured traces of their natural migration patterns.

The near-identity of all *B. quintana* strains is presumably caused by its recent emergence in the human population. Likewise, a possible explanation to the lower sequence divergence in *B. henselae* as compared with *B. grahamii* is that the *B. henselae* population emerged more recently. A much larger population size

and population turnover for rodents compared with cats will lead to a larger population size of *B. grahamii*, a higher accumulation of mutations per unit time and thereby a higher sequence diversity. It is however also possible that domestic cats only harbour a fraction of all diversity within *B. henselae*, and it would be interesting to determine whether strains from wild felines have the same relatively low levels of sequence divergence.

Another remarkable finding was that recombination frequencies appear to differ between *Bartonella* populations. We observed well-supported conflicts across trees for individual loci in *B. grahamii*, but not in *B. henselae*, in agreement with an estimated recombination over mutation ratio in the higher range ($r/m = 1.7$) for *B. grahamii* as compared with $r/m = 0.1$ for *B. henselae* (Vos & Didelot 2009). This difference may be due to the higher density of wild rodents in nature, more frequent co-infections of different *Bartonella* species, and a higher exposure to vectors than for domestic cats. Typically, intracellular bacteria have low rates of recombination and horizontal gene acquisition (Vos & Didelot 2009), probably because of their isolated growth niche, lack of mobile elements and limited exposure to foreign DNA (Moran & Plague 2004). A notable exception is the A-group strains of *Wolbachia pipientis*, where it is estimated that up to 75% of all genes might have undergone recent recombination events, possibly mediated by transfer of DNA via bacteriophages across strains that infect the same insect host (Klasson *et al.* 2009). The high recombination frequencies in *B. grahamii* may be explained by the prevalence of plasmids in this species (Seubert *et al.* 2003a; Berglund *et al.* 2009), or the recently described GTA that packages random bacterial DNA (Berglund *et al.* 2009). With the aid of the GTA, large-scale gene transfer and sequence exchange is possible not only between strains that infect the same host or vector simultaneously but also between strains infecting at different time points.

Also in terms of gene content and genome size, *B. grahamii* displays more variation than what was previously observed in *B. henselae* (Lindroos *et al.* 2006). Still, homologous genes are variable in both the *B. grahamii* and the *B. henselae* populations, including genes for T4SSs, T5SSs, membrane proteins and phage genes. The genome sizes of several *B. grahamii* strains were estimated to be larger than that of the sequenced strain, suggesting that additional genes are present in the *B. grahamii* pan-genome. The identification of phage gene remnants in several sequenced spacer regions further indicates a continuous flux in the mobile gene pool of *B. grahamii*. A higher exposure to bacteriophages in the rodent-associated lineages may also contribute to elevated recombination frequencies.

Geographic variations in the genomic diversity: age, recombination and/or recent bottlenecks?

Our results revealed a striking difference in sequence diversity between Asian and European-American *B. grahamii* strains, with the Asian strains being much more divergent and showing more extensive recombination. As in the comparison of diversity among species, we can think of several possible reasons for this: the Asian population is older, larger and/or subjected to increased recombination frequencies. An alternative explanation is that the American–European population has undergone a recent bottleneck.

This raises interesting questions about the age of *Bartonella* infections in rodents, where the ancestor of the *B. grahamii* population originated, and how it has migrated since then. In the previous study by Inoue *et al.* (2009) it was suggested that *B. grahamii* existed in North America before Europe, and crossed the Bering's strait during a glacial epoch to the Euroasian continent. This hypothesis was based on the *gltA* sequence from a *B. grahamii* isolate from a *Mus musculus* in California, which appeared to be the earliest diverging lineage in the European–American group. When we repeated this analysis, including additional sequences, we found that the American strain Mo12658sd and a strain from Greece appear to represent the earliest diverging lineage in the European–American group, suggesting that we cannot say whether the origin of this group was in Europe or America. However, it is important to recall that the limited sequence information contained within the *gltA* gene is not enough to infer reliable phylogenies, that the diversity that we know today is probably only a fraction of what is out there, and that the earliest diverging lineage may not be there anymore.

Although it is tempting to speculate that the continent with the highest diversity represents the source of a population, we have conclusively shown that the higher genetic diversity in Asia does not by necessity imply that the ancestor was derived from this continent. If so, we would expect the American–European strains to represent an expansion from within the Asian clade. Rather, we observed two well-supported clades, suggesting that the origin could be at either side of the Bering's strait. Hence, we have no evidence to say that one population is older than the other.

Higher recombination frequencies would be a simple explanation for the observed higher diversity in the Asian population. Recombination frequencies across strains are likely to increase with the opportunities for multiple infections of the same host individual, which in turn will depend on the extent of host-specificity and the density of hosts and vectors. The overall prevalence of *Bartonella* in the rodent populations we have studied ran-

ged from 17% in Sweden (Holmberg *et al.* 2003) to 64% in Russia (Markov *et al.* 2006). However, large variations in abundance have been observed within countries, for example, in Japan where the *Bartonella* prevalence was over 80% in two of the geographic regions (Nagano and Ehime) from which the isolates in this study were sampled (Inoue *et al.* 2008). Furthermore, co-infections of different species within a single host have been demonstrated in Japan (Inoue *et al.* 2008), suggesting that these strains have indeed had opportunities for recombination. However, we cannot exclude that recombination occurs with similarly high frequencies also within the European and American populations, but has escaped detection in this analysis because of high sequence similarity across strains and limited sampling.

Another possible reason to why American and European populations are less diverse is that they can have experienced bottlenecks associated with the latest glacial period, which ended approximately 10 000 years ago (Ehlers & Gibbard 2007). Indeed, most of the samples from these two continents are derived from sites covered by ice at the glacial maximum, *c.* 18 000 years ago (Ehlers & Gibbard 2007) (Fig. S9, Supporting Information). Numerous studies of genetic diversity among rodents in Europe and North America have identified areas that were used as refuges during this time, and shown that genetic diversity decreases with distance from these refuges (Lessa *et al.* 2003; Michaux *et al.* 2003; Rowe *et al.* 2004; Kotlik *et al.* 2006). If only a limited number of rodents made it back to the previously ice-covered areas, then this is likely to have reduced the variability also of their blood-borne inhabitants. Interestingly, our local populations in UK and Canada display approximately the same degree of diversity as has been observed for human pathogens inferred to have emerged within the past 20 000 years (Achtman 2004, 2008), while the diversity of *B. grahamii* in the entire European–American group is much greater (Table S5, Supporting Information). Thus, if diversity in these regions was reduced as an effect of the ice age, it seems likely that rodents infected with genetically different *Bartonella* strains initiated the re-colonization.

Other factors influencing the infection dynamics such as population size variation, climate differences and seasonal sampling time may also contribute to the observed sequence divergence variation across continents. More ecological data and an ecological model are clearly needed to identify the selective and ecological factors underlying the apparent high recombination frequencies among *B. grahamii* strains in Asia. In the future, systematic surveys of the genomic diversity of rodent-associated bacterial populations along with collections of demographic data on the abundance, migration and diversity of rodents and their vectors will be invaluable for elucidating the historical and ecological factors that influence the genomic diversity of zoonotic bacterial species.

Acknowledgements

We thank Annelie Waldén, Max Käller and Anna Westring at the Royal Institute of Technology, Stockholm, Sweden for microarray printing, Kristina Näslund at Uppsala University, Sweden, for sequencing and Björn Nystedt and Lisa Klasson for helpful discussions. This study was supported by grants to S.G.E.A. from the Swedish Research Council, the Göran Gustafsson Foundation, the Swedish Foundation for Strategic Research and the Knut and Alice Wallenberg Foundation.

References

Achtman M (2004) Population structure of pathogenic bacteria revisited. *International Journal of Medical Microbiology*, **294**, 67–73.

Achtman M (2008) Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annual Review of Microbiology*, **62**, 53–70.

Alsmark CM, Frank AC, Karlberg EO *et al.* (2004) The louse-borne human pathogen *Bartonella quintana* is a genomic derivative of the zoonotic agent *Bartonella henselae*. *Proceedings of the National Academy of Sciences, USA*, **101**, 9716–9721.

Arvand M, Feil EJ, Giladi M, Boulouis HJ, Viezens J (2007) Multi-locus sequence typing of *Bartonella henselae* isolates from three continents reveals hypervirulent and feline-associated clones. *Public Library of Science ONE*, **2**, e1346.

Berglund EC, Frank AC, Calteau A *et al.* (2009) Run-off replication of host-adaptability genes is associated with gene transfer agents in the genome of mouse-infecting *Bartonella grahamii*. *PLoS Genetics*, **5**, e1000546.

Berglund EC, Ehrenborg C, Vinnere Pettersson O *et al.* (2010) Genome dynamics of *Bartonella grahamii* in micro-populations of woodland rodents. *BMC Genomics*, **11**, 152.

Birtles RJ, Canales J, Ventosilla P *et al.* (1999) Survey of *Bartonella* species infecting intradomicillary animals in the Huayllacallan Valley, Ancash, Peru, a region endemic for human bartonellosis. *American Journal of Tropical Medical Hygiene*, **60**, 799–805.

Birtles RJ, Harrison TG, Saunders NA, Molyneux DH (1995) Proposals to unify the genera *Grahamella* and *Bartonella*, with descriptions of *Bartonella talpae* comb. nov., *Bartonella peromysci* comb. nov., and three new species, *Bartonella grahamii* sp. nov., *Bartonella taylorii* sp. nov., and *Bartonella doshiae* sp. nov. *International Journal of Systematic Bacteriology*, **45**, 1–8.

Bouchouicha R, Durand B, Monteil M *et al.* (2009) Molecular epidemiology of feline and human *Bartonella henselae* isolates. *Emerging Infectious Diseases*, **15**, 813–816.

Boulouis HJ, Chang CC, Henn JB, Kasten RW, Chomel BB (2005) Factors associated with the rapid emergence of zoonotic *Bartonella* infections. *Veterinary Research*, **36**, 383–410.

Bown KJ, Bennet M, Begon M (2004) Flea-borne *Bartonella grahamii* and *Bartonella taylorii* in bank voles. *Emerging Infectious Diseases*, **10**, 684–687.

Chou HH, Holmes MH (2001) DNA sequence quality trimming and vector removal. *Bioinformatics*, **17**, 1093–1104.

Corbet GB, Harris S (1991) *The Handbook of British Mammals*, 3rd edn. Blackwell Scientific Publications, Oxford.

Didelot X, Falush D (2007) Inference of bacterial microevolution using multilocus sequence data. *Genetics*, **175**, 1251–1266.

Ehlers J, Gibbard PL (2007) The extent and chronology of Cenozoic Global Glaciation. *Quaternary International 164*, **165**, 6–20.

Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. *II. Error probabilities. Genome Research*, **8**, 186–194.

Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. *I. Accuracy assessment. Genome Research*, **8**, 175–185.

Felsenstein J (1989) PHYLIP—Phylogeny Inference Package (Version 3.2). *Cladistics*, **5**, 164–166.

Foucault C, La Scola B, Lindroos H, Andersson SG, Raoult D (2005) Multispacer typing technique for sequence-based typing of *Bartonella quintana*. *Journal of Clinical Microbiology*, **43**, 41–48.

Gordon D, Abajian C, Green P (1998) Consed: a graphical tool for sequence finishing. *Genome Research*, **8**, 195–202.

Gundi VA, Taylor C, Raoult D, La Scola B (2009) *Bartonella rattaustraliani* sp. nov., *Bartonella queenslandensis* sp. nov. and *Bartonella coopersplainsensis* sp. nov., from Australian rats. *International Journal Systematic and Evolutionary Microbiology*, **52**, 1043–4017.

Holmberg M, Mills JN, McGill S, Benjamin G, Ellis BA (2003) *Bartonella* infection in sylvatic small mammals of central Sweden. *Epidemiology of Infection*, **130**, 149–157.

Hsieh JW, Tung KC, Chen WC *et al.* (2009) Epidemiology of *Bartonella* infection in rodents and shrews in Taiwan. *Zoonoses Public Health*, Epub ahead of print.

Huson DH (1998) SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics*, **14**, 68–73.

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254–267.

Inoue K, Kabeya H, Kosoy MY *et al.* (2009) Evolutional and geographical relationships of *Bartonella grahamii* isolates from wild rodents by multi-locus sequencing analysis. *Microbial Ecology*, **57**, 534–541.

Inoue K, Maruyama S, Kabeya H *et al.* (2008) Prevalence and genetic diversity of *Bartonella* species isolated from wild rodents in Japan. *Applied Environmental Microbiology*, **74**, 5086–5092.

Iredell J, Blanckenberg D, Arvand M *et al.* (2003) Characterization of the natural population of *Bartonella henselae* by multilocus sequence typing. *Journal of Clinical Microbiology*, **41**, 5071–5079.

Kaiser PO, Riess T, Wagner CL *et al.* (2008) The head of *Bartonella* adhesin A is crucial for host cell interaction of *Bartonella henselae*. *Cell Microbiology*, **10**, 2223–2234.

Kerkhoff FT, Bergmans AM, van Der Zee A, Rothova A (1999) Demonstration of *Bartonella grahamii* DNA in ocular fluids of a patient with neuroretinitis. *Journal of Clinical Microbiology*, **37**, 4034–4038.

Klasson L, Westberg J, Sapountzis P *et al.* (2009) The mosaic genome structure of the Wolbachia wRi strain infecting Drosophila simulans. *Proceedings of the National Academy of Sciences, USA*, **106**, 5725–5730.

Kordick DL, Breitschwerdt EB (1998) Persistent infection of pets within a household with three *Bartonella* species. *Emerging Infectious Diseases*, **4**, 325–328.

Kotlik P, Deffontaine V, Mascheretti S *et al.* (2006) A northern glacial refugium for bank voles (Clethrionomys glareolus). *Proceedings of the National Academy of Sciences, USA*, **103**, 14860–14864.

Lessa EP, Cook JA, Patton JL (2003) Genetic footprints of demographic expansion in North America, but not Amazonia, during the Late Quaternary. *Proceedings of the National Academy of Sciences, USA*, **100**, 10331–10334.

Li W, Chomel BB, Maruyama S *et al.* (2006) Multispacer typing to study the genotypic distribution of *Bartonella henselae* populations. *Journal of Clinical Microbiology*, **44**, 2499–2506.

Li W, Raoult D, Fournier PE (2007) Genetic diversity of *Bartonella henselae* in human infection detected with multispacer typing. *Emerging Infectious Disease*, **13**, 1178–1183.

Lindroos H, Vinnere O, Mira A *et al.* (2006) Genome rearrangements, deletions, and amplifications in the natural population of *Bartonella henselae*. *Journal of Bacteriology*, **188**, 7426–7439.

Lindroos HL, Mira A, Repsilber D *et al.* (2005) Characterization of the genome composition of *Bartonella koehlerae* by microarray comparative genomic hybridization profiling. *Journal of Bacteriology*, **187**, 6155–6165.

Lipinski MJ, Froenicke L, Baysac KC *et al.* (2008) The ascent of cat breeds: genetic evaluations of breeds and worldwide random-bred populations. *Genomics*, **91**, 12–21.

Loytynoja A, Goldman N (2008) Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science*, **320**, 1632–1635.

Markov AP, Lopyrev IV, Irkhin AI *et al.* (2006) Wild small mammals are the reservoir hosts of the *Bartonella* genus bacteria in the south of Moscow region. *Molecular Genetics, Microbiol. Virology (Moscow)*, **4**, 8–13.

Martin DP, Williamson C, Posada D (2005) RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics*, **21**, 260–262.

Mediannikov O, Ivanov L, Zdanovskaya N *et al.* (2005) Molecular screening of *Bartonella* species in rodents from the Russian Far East. *Annals of New York Academy of Science*, **1063**, 308–311.

Michaux JR, Magnanou E, Paradis E, Nieberding C, Libois R (2003) Mitochondrial phylogeography of the Woodmouse (*Apodemus sylvaticus*) in the Western Palearctic region. *Molecular Ecology*, **12**, 685–697.

Moran NA, Plague GR (2004) Genomic changes following host restriction in bacteria. *Current Opinion Genetics and Development*, **14**, 627–633.

Pretorius AM, Beati L, Birtles RJ (2004) Diversity of bartonellae associated with small mammals inhabiting Free State province, South Africa. *International Journal of Systematics and Evolutionary Microbiology*, **54**, 1959–1967.

Riess T, Andersson SG, Lupas A *et al.* (2004) Bartonella adhesin a mediates a proangiogenic host cell response. *Journal of Experimental Medicine*, **200**, 1267–1278.

Rowe KC, Heske EJ, Brown PW, Paige KN (2004) Surviving the ice: Northern refugia and postglacial colonization.

Proceedings of the National Academy of Sciences, USA, **101**, 10355–10359.

Saenz HL, Engel P, Stoeckli MC *et al.* (2007) Genomic analysis of Bartonella identifies type IV secretion systems as host adaptability factors. *Nature Genetics*, **39**, 1469–1476.

Schmid MC, Scheidegger F, Dehio M *et al.* (2006) A translocated bacterial protein protects vascular endothelial cells from apoptosis. *Public Library of Science Pathogens*, **2**, e115.

Schmid MC, Schulein R, Dehio M *et al.* (2004) The VirB type IV secretion system of *Bartonella henselae* mediates invasion, proinflammatory activation and antiapoptotic protection of endothelial cells. *Molecular Microbiology*, **52**, 81–92.

Schulein R, Dehio C (2002) The VirB/VirD4 type IV secretion system of *Bartonella* is essential for establishing intraerythrocytic infection. *Molecular Microbiology*, **46**, 1053–1067.

Schulein R, Guye P, Rhomberg TA *et al.* (2005) A bipartite signal mediates the transfer of type IV secretion substrates of *Bartonella henselae* into human cells. *Proceedings of the National Academy of Sciences, USA*, **102**, 856–861.

Serratrice J, Rolain JM, Granel B *et al.* (2003) Bilateral retinal artery branch occlusions revealing *Bartonella grahamii* infection. *Rev Med Interne*, **24**, 629–630.

Seubert A, Falch C, Birtles RJ, Schulein R, Dehio C (2003a) Characterization of the cryptic plasmid pBGR1 from *Bartonella grahamii* and construction of a versatile *Escherichia coli–Bartonella* spp. shuttle cloning vector. *Plasmid*, **49**, 44–52.

Seubert A, Hiestand R, de la Cruz F, Dehio C (2003b) A bacterial conjugation machinery recruited for pathogenesis. *Molecular Microbiology*, **49**, 1253–1266.

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–2690.

Telfer S, Begon M, Bennett M *et al.* (2007) Contrasting dynamics of *Bartonella* spp. in cyclic field vole populations: the impact of vector and host dynamics. *Parasitology*, **134**, 413–425.

Vos M, Didelot X (2009) A comparison of homologous recombination rates in bacteria and archaea. *ISME Journal*, **3**, 199–208.

Winoto IL, Goethert H, Ibrahim IN *et al.* (2005) *Bartonella* species in rodents and shrews in the greater Jakarta area. *Southeast Asian Journal of Tropical Medicine Public Health*, **36**, 1523–1529.

Woolley MW, Gordon DL, Wetherall BL (2007) Analysis of the first Australian strains of *Bartonella quintana* reveals unique genotypes. *Journal of Clinical Microbiology*, **45**, 2040–2043.

Ying B, Kosoy MY, Maupin GO, Tsuchiya KR, Gage KL (2002) Genetic and ecologic characteristics of Bartonella communities in rodents in southern China. *American Journal of Tropical Medicine Hygiene*, **66**, 622–627.

Zhang P, Chomel BB, Schau MK *et al.* (2004) A family of variably expressed outer-membrane proteins (Vomp) mediates adhesion and autoaggregation in *Bartonella quintana*. *Proceedings of the National Academy of Sciences, USA*, **101**, 13630–13635.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Phylogenetic tree inferred from the concatenation of all gene sequences. Branch lengths are according to maximum likelihood, and bootstrap support values for the major clades are shown. The branch length of *B. vinsonii* is not according to scale.

Fig. S2 Phylogenetic tree inferred from the concatenation of all sequenced loci after removing all sites with gaps. Branch lengths are according to maximum likelihood, and bootstrap support values for the major clades are shown. The branch length of *B. vinsonii* is not according to scale.

Fig. S3 Phylogenetic tree based on partial *gltA* sequences (269 bp) of the *B. grahamii* strains analyzed in this study, and additional sequences from other parts of the world. Branch lengths are according to maximum likelihood, and bootstrap support values for the major clades are shown. Origin and accession numbers are shown for genes not sequenced as part of this study.

Fig. S4 Phylogenetic tree for locus 5. Branch lengths are according to maximum likelihood, and bootstrap support values over 75 are shown. The branch length of *B. vinsonii* is not according to scale.

Fig. S5 Consensus tree from ClonalFrame. Posterior probabilities for the major clades are shown.

Fig. S6 Schematic representation of PFGE results with the NotI enzyme for all strains (left). Since the size of the largest NotI band is difficult to estimate accurately, we also applied another enzyme, AscI, to a selected set of strains (right). The country of origin and estimated genome size for each strain are shown. The large difference in estimated genome size of strain Ehime 5-1 is probably due to an undetected band or a double band in the NotI digest.

Fig. S7 Circular representation of the microarray results for the plasmid pBGR3. The innermost circle shows the plasmid genes (magenta indicates the *vbh* genes). The strains are: 1, V2; 2, WM11; 3, PTZA 30/3; 4, PTZB 29/18; 5, J142; 6, C066; 7, R170; 8, J019; 9, S116; 10, MAC29; 11, C162; 12, Cg4228alb; 13, Cg4224alb; 14, Cg4227alb; 15, Cg4263alb; 16, Cg4285alb; 17, Mo12494sd; 18, Mo12658sd; 19, Ac1733yn; 20, AL1714yn; 21, Ehime 5-1; 22, Nakanoshima 39-1; 23, Nagano 14-1; 24, Fuji 4-1; 25, Hokkaido 29-1.

Fig. S8 Hybridization results of the *fha*-repeat in BgGI 4 (top, two of the three *fha/hec* operons in this island are shown), BgGI 14 (bottom, left) and BgGI 7 (bottom, right). In each image, the panels show from the top: the frequency of repeats in the *B. grahamii* as4aup genome (the number at which each sequence occurs in the genome with more than 80% sequence identity over 100 bp); the *B. grahamii* as4aup genes on the two strands with phage genes shown in orange; hybridization results for each strain. Rhombi below a gene indicate an integrase gene or remnant.

Fig. S9 Map showing the approximate extension of the latest glacial epoch. The sites where the bacterial isolates included in this study were sampled are indicated (only one site is shown for each country). Within parenthesis is the number of strains from each country.

Table S1 EMBL accession numbers for the eleven sequenced loci

Table S2 Sequence similarities within and between phylogenetic or geographic groups

Table S3 Predicted recombination events within sequenced loci

Table S4 Predicted recombination events in the MLST dataset from (Inoue *et al.* 2009)

Table S5 Sequence divergence and estimated divergence time of genetically monomorphic pathogens and our *B. grahamii* populations

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

International Journal of Systematic and Evolutionary Microbiology (2010), **60**, 759–763 DOI 10.1099/ijs.0.011528-0

Bartonella japonica sp. nov. and *Bartonella silvatica* sp. nov., isolated from *Apodemus* mice

Kai Inoue,[1] Hidenori Kabeya,[1] Hatsumi Shiratori,[2] Kenji Ueda,[2] Michael Y. Kosoy,[3] Bruno B. Chomel,[4] Henri-Jean Boulouis[5] and Soichi Maruyama[1]

Correspondence
Soichi Maruyama
maruyama.soichi@nihon-u.ac.jp

[1]Laboratory of Veterinary Public Health, Department of Veterinary Medicine, College of Bioresource Sciences, Nihon University, 1866 Kameino, Fujisawa, Kanagawa 252-8510, Japan

[2]Life Science Research Center, College of Bioresource Sciences, Nihon University, Fujisawa, Kanagawa 252-8510, Japan

[3]Division of Vector-Borne Infectious Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Fort Collins, CO 80521, USA

[4]Department of Population Health and Reproduction, School of Veterinary Medicine, University of California, Davis, CA 95616, USA

[5]UMR BIPAR/AFSSA, Institut National de la Recherche Agronomique, Ecole Nationale Vétérinaire d'Alfort/UVPM, 7 avenue du Général de Gaulle, 94704 Maisons-Alfort Cedex, France

Two bacterial strains, Fuji 18-1^T and Fuji 23-1^T, were isolated from the blood of the small Japanese field mouse (*Apodemus argenteus*) and the large Japanese field mouse (*Apodemus speciosus*), respectively, specimens of which were captured in the forest of Mount Fuji, Japan. Phenotypic characterization (growth conditions, incubation periods, biochemical properties and cell morphologies), DNA G+C contents (40.1 mol% for strain Fuji 18-1^T and 40.4 mol% for strain Fuji 23-1^T) and sequence analyses of the 16S rRNA genes indicated that both strains were members of the genus *Bartonella*. Using *rpoB* and *gltA* sequencing analysis, the highest sequence similarities between strains Fuji 18-1^T, Fuji 23-1^T and other recognized species of the genus *Bartonella* showed values considerably lower than 91.4 % and 89.9 % in the *rpoB* gene and 89.1 % and 90.4 % in the *gltA* gene, respectively. It is known that similarities of 95.4 % for the *rpoB* gene and 96.0 % for the *gltA* gene can be applied as cut-off values for the designation of novel species of the genus *Bartonella*. In a phylogenetic tree based on the merged set of concatenated sequences of seven loci [16S rRNA, *ftsZ*, *gltA*, *groEL*, *ribC* and *rpoB* genes and the intergenic spacer region (ITS)], strains Fuji 18-1^T and Fuji 23-1^T formed a distinct clade from other recognized species of the genus *Bartonella*. These data support the classification of strains Fuji 18-1^T and Fuji 23-1^T as novel species of the genus *Bartonella*. The names *Bartonella japonica* sp. nov. and *Bartonella silvatica* sp. nov. are proposed for these novel species. The type strains of *Bartonella japonica* sp. nov. and *Bartonella silvatica* sp. nov. are Fuji 18-1^T (=JCM 15567^T=CIP 109861^T) and Fuji 23-1^T (=JCM 15566^T=CIP 109862^T), respectively.

Abbreviations: ITS, intergenic spacer region; MP, maximum-parsimony; NJ, neighbour-joining.

The GenBank accession numbers for the 16S rRNA, *ftsZ*, *gltA*, *groEL*, *ribC* and *rpoB* genes and the 16S–23S rRNA intergenic spacer region (ITS) from *Bartonella japonica* sp. nov. Fuji 18-1^T are AB440632, AB440633, AB242289, AB440634, AB440635, AB242288, and AB498007, respectively. GenBank accession numbers for the 16S rRNA, *ftsZ*, *gltA*, *groEL*, *ribC* and *rpoB* genes and ITS sequences from *Bartonella silvatica* sp. nov. Fuji 23-1^T are AB440636, AB440637, AB242287, AB440638, AB440639, AB242292, and AB498008, respectively.

Sequence similarities of the 16S rRNA, *ftsZ*, *gltA*, *groEL*, *ribC* and *rpoB* genes and ITS region sequences between strains Fuji 18-1^T, Fuji 23-1^T and other species of the genus *Bartonella*, a maximum-parsimony phylogenetic tree based on the concatenated sequence of the seven loci and neighbour-joining trees of the seven loci are available as supplementary material with the online version of this paper.