Title: Multiple locus variable number tandem repeat analysis (MLVA) of the pathogenic intestinal spirochaete *Brachyspira pilosicoli*

Authors: Eugene Neo, Tom La, Nyree Dale Phillips, David J Hampson

PII: S0378-1135(13)00039-4
DOI: doi:10.1016/j.vetmic.2012.12.030
Reference: VETMIC 6050

To appear in: *VETMIC*

Received date: 8-10-2012
Revised date: 19-12-2012
Accepted date: 20-12-2012

Please cite this article as: Neo, E., La, T., Phillips, N.D., Hampson, D.J., Multiple locus variable number tandem repeat analysis (MLVA) of the pathogenic intestinal spirochaete *Brachyspira pilosicoli*, *Veterinary Microbiology* (2010), doi:10.1016/j.vetmic.2012.12.030

Multiple locus variable number tandem repeat analysis (MLVA) of the pathogenic intestinal spirochaete *Brachyspira pilosicoli*

Eugene Neo, Tom La, Nyree Dale Phillips, and David J Hampson*

School of Veterinary and Biomedical Science, Murdoch University, Murdoch, Western Australia 6150, Australia

Short title: MLVA for *Brachyspira pilosicoli*

* Corresponding author. Tel.: +61 89360 2287; fax: +61 89310 4144.

E-mail address: d.hampson@murdoch.edu.au (D.J. Hampson).

Abstract

Brachyspira pilosicoli is an anaerobic intestinal spirochaete that colonizes the large intestine of various host species, in which it may induce diarrhoea, poor growth rates and a localized colitis known as intestinal (or colonic) spirochaetosis. The spirochaete is considered to be potentially zoonotic. The purpose of the current study was to develop a multiple-locus variable number tandem repeat analysis (MLVA) method as a simple and rapid tool to investigate the molecular epidemiology of *B. pilosicoli*. The genomic sequence of *B. pilosicoli* strain 95/1000 was analysed for potential tandem repeats using the default parameters of the Tandem Repeat Finder program. A total of 22 repeat loci were identified and tested for their presence and variability on a set of 10 *B. pilosicoli* isolates. Five loci that were present in most isolates and that showed evidence of allelic variation were selected and used with a collection of 119 isolates from different host species and geographical locations. Not all the isolates amplified at all loci, but using the available data a total of 103 VNTR profiles were generated. The discriminatory power of this method was 0.976. A phylogenetic tree constructed from the allelic profiles confirmed the diversity of *B. pilosicoli*, and the general lack of clustering of strains based on species of origin or geographic origin. Some isolates with known epidemiological links were found to be identical or highly similar. The MLVA method was simple and easy to use, and could readily differentiate between strains of *B. pilosicoli*. MLVA should prove to be a useful tool for rapid identification of relationships between *B. pilosicoli* isolates in epidemiological investigations.

Keywords: Spirochaete; *Brachyspira pilosicoli*; MLVA; molecular epidemiology; zoonosis.

1. Introduction

The genus *Brachyspira* includes species of anaerobic spirochaetes that colonize the large intestine of animals and birds (Stanton, 2006). The two most important pathogenic species are *Brachyspira hyodysenteriae*, the agent of swine dysentery, and *Brachyspira pilosicoli*, the agent of intestinal spirochaetosis (Hampson, 2012). *B. pilosicoli* has a broader host range than *B. hyodysenteriae*, colonizing various species of animals and birds, as well as human beings (Hampson et al., 2006). Colonized individuals may develop focal colitis, with chronic diarrhoea. Infections with *B. pilosicoli* are particularly common in intensively housed pigs and chickens, in which rates of growth and production may be depressed. Colonization also is common amongst humans living in crowded and unhygienic conditions (Trott et al., 1997a; Nelson et al., 2009). Spirochaetaemia with *B. pilosicoli* has been recorded in immunocompromised or debilitated human beings (Trott et al., 1997b), but not yet in animals. Transmission may be by direct exposure to infected faeces, or by indirect routes; for example, carcasses of spent hens in supermarkets may be contaminated with *B. pilosicoli* (Verlinden et al., 2012), while accessing drinking water from wells has been linked to transmission in humans (Margawani et al., 2004).

To help answer questions about transmission routes of *B. pilosicoli*, a simple, rapid and discriminating typing system that gives information about genetic relatedness of isolates is needed. Pulsed field gel electrophoresis has been used for this purpose (Atyeo et al., 1996), but it is quite time consuming and requires large quantities of pure DNA. Multiple-locus variable number tandem repeat analysis (MLVA) is a method that uses non-coding regions of genomic DNA, consisting of direct repeats of the same sequence of bases, referred to as tandem repeats. The number of repeats in

the sequence is determined genetically and can vary between different strains. These regions are known as variable number tandem repeats (VNTR) loci. Recently, an MLVA scheme had been developed and used for *B. hyodysenteriae* (Hidago et al., 2010). The purpose of the current study was to develop an MLVA scheme for *B. pilosicoli*, and to use it as an epidemiological tool to compare isolates from different sources.

2. Materials and Methods

2.1. Spirochaete isolates

A total of 119 *B. pilosicoli* isolates and strains recovered from different host species and geographical areas were obtained from the culture collection at the Reference Centre for Intestinal Spirochaetes at Murdoch University (Table 1). These consisted of 45 isolates from pigs, 42 from human beings, 25 from chickens, five from dogs, and two from horses. They originated from Australia (n=66), Papua New Guinea (n=29), the USA (n=8), Canada (n=5), Italy (n=5), the UK (n=3) and New Zealand (n=1). The isolates had been identified as *B. pilosicoli* on the basis of their phenotypic properties and amplification in a species-specific PCR, as previously described (La et al., 2003).

2.2. Spirochaete culture and DNA extraction

The spirochaetes were thawed from frozen stock and subcultured at 37^{o}C in Kunkle's pre-reduced anaerobic broth containing 2 % (vol/vol) foetal bovine serum

and a 1 % (vol/vol) ethanolic cholesterol solution (Kunkle et al., 1986). Cells were harvested from mid-log phase culture by centrifuging at 10,000 x *g*, and counted in a haemocytometer chamber under a phase contrast microscope. For each isolate, 10ml of Trypticase Soy broth containing ~10^8 cells/ml of *B. pilosicoli* was centrifuged at 5000g. The supernatant was discarded and the pellet resuspended in an equal volume of phosphate buffered saline (37 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.47 mM KH_2PO_4, pH 7.4) and heat killed at 95°C for 15 min to release the DNA, before being stored frozen at -20°C. The solution containing the extracted DNA was used as the template for the PCR reactions.

2.2. Selection and PCR amplification of potential VNTR loci

The complete genome sequence of *B. pilosicoli* strain 95/1000 (Wanchananthuek et al., 2010), isolated from a pig, was obtained from GenBank (accession number CP002025; Project ID: 48097) and investigated for potential tandem repeats using the default parameters of the Tandem Repeat Finder program (Benson, 1999), which is available as a Web service (http://tandem.bu.edu/trf/trf.html). The selected tandem-repeat loci were ranked by consensus length, and those between 40 and 300 base pairs were used to design primers, using the flanking regions. These 22 loci were named BPilo, followed by the repeat length ranking number (from 1 to 22), separated by an underscore. The primer sequences are shown in Table 2.

PCR and electrophoresis was carried out on DNA from the *B. pilosicoli* strains. PCR was performed using 0.2µl *Taq* DNA polymerase, 5µl of 10x PCR buffers, 3µl of 25mM $MgCl_2$, 5µl of 8mM dNTP, 5µl of the forward and reverse primers, 12µl of

cresol red solution and 2µl of template, with the reaction mix topped up with PCR-grade water to a final 50µl volume. The PCR conditions were 95°C for 2 min, followed by 33 cycles at 95°C for 30 s, 50°C for 15 s, 72°C for 1 min for every 1kbp of product, and a final extension period of 5 min at 72°C before holding at 14°C. The products were visualized using gel electrophoresis and staining by immersion in 1% ethidium bromide for 30 min. A molecular mass marker (1Kbp Ladder; Axygen Biosciences, California, USA) was placed in a well at each end of the gel. Images of the gel were captured using a Biorad Chem Doc XRS Universal Hood and were used for analysis. On the gel image, lines were drawn from bands of the ladder on one side of the gel, connecting the bands of the same size on the ladder at the other end. The size of each VNTR was estimated using the lines as reference points.

Initially all 22 sets of primers that were identified as being potentially suitable were tested using DNA from *B. pilosicoli* strain 95/1000 as a template. After verifying that the primers yielded the appropriate results, PCRs were carried out on DNA extracted from a subset of 10 randomly selected isolates of *B. pilosicoli* (marked with an asterisk in Table 1). Each PCR reaction set included DNA from *B. pilosicoli* strain 95/1000 as a positive control and distilled water as a negative control. For the loci where the PCR did not produce product, different sets of primers were designed and used in an attempt to detect the predicted tandem repeats. The sequences of these primers are shown in Supplementary Table 1.

To confirm that the PCR products were from the repeat regions, the product from five loci for strains 95/1000, OF 2 and H54 were purified with the Wizard® SV Gel and

PCR Clean-Up System Kit (Promega, Sydney, Australia) in accordance with the manufacturer's instructions, and were sequenced using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, USA). Sequencing was performed with the ABI 373A sequencing system (Applied Biosystems).

2.3. MLVA with B. pilosicoli

PCRs were carried out on DNA from all 119 isolates and strains, sequentially using the five selected primer pairs for the loci indicated in Table 3. The process was repeated using the same samples to assess the reproducibility of the procedure.

The number of tandem repeats was calculated using the estimated size of the resultant band divided by the size of the repeat. Alleles with the same number of tandem repeats were grouped together and assigned an allelic number according to their size, with the larger ones first. Those isolates that consistently produced a negative result with a particular primer pair were assigned the number 99 for the purpose of typing. The allelic number was entered into the START2 program and a rooted phylogenetic tree with 1000 bootstrap replicates was generated from the data matrix using the 'Neighbour-Joining' (NJ) with the 'Maximum likelihood' model (Jolley et al., 2001). Two of the 119 isolates that did not amplify at any of the five loci used were not included in the tree. A separate tree was generated for 20 isolates where all five alleles were amplifiable.

The discriminatory power (D) of the MLVA method was calculated as previously described (Grundmann et al., 2001). D is the average probability that the typing

system will assign a different type to two unrelated strains randomly sampled in the microbial population of a given taxon.

3. Results

Investigation of the genome sequence of *B. pilosicoli* 95/1000 with the Tandem Repeat Finder program identified 404 repeats in tandem. Of these, 22 were suitable candidates for MLVA due to their size and well-separated locations dispersed around the genome.

Using the 22 primer pairs with 10 isolates yielded five sets that gave amplication with most of these isolates and loci, and which produced variations between the amplicons that allowed differentiation between them. The remaining 17 primer pairs either did not produce product or/and identified little or no allelic variation. Use of the alternate primers listed in Supplementary Table 2 gave the same results, so these were not used further. The five repeats that were selected for use in MLVA were located across about two thirds of the genome of strain 95/1000. Sequencing of the products confirmed that the length polymorphism was due to differences in copy number of tandem repeats, and their characteristics as assessed using 95/1000 as the template are shown in Table 3.

A summary of the alleles carried by the 119 isolates at the five loci is presented in Table 4. The number of alleles varied from four for BPilo_14 to nine for BPilo_9. The number of isolates in which a locus did not amplify varied from 16 for BPilo_14 (13.4%) to 79 for BPilo_8 (66.4%). Two isolates (UNL-3 and Q98.0072.31) did not

amplify at any of the five loci. Nevertheless, by including null alleles ("99"), a total of 103 VNTR profiles (VTs) were generated (Supplementary Table 2). The repeated tests yielded the same results with the isolates.

Overall the discriminatory power of this method using the five loci was 0.976.

The NJ tree generated using the allelic profile from the MLVA sequencing for 117 isolates is shown as Figure 1, and the identity of the isolates is shown in Table 1. A separate tree for the 20 isolates where all five alleles amplified is shown as a Supplementary Figure. Isolates from the same species of origin (pig, chicken, human) were distributed throughout the tree shown in Figure 1, without any major clustering on a species basis. Similarly, isolates from the same country or region of origin generally were not restricted to one area on the tree. Some clustering of isolates that were likely to be epidemiological linked was observed; for example certain isolates from village 1 and piggery 1 in Papua New Guinea, and certain human isolates collected from an Aboriginal community in Western Australia (marked WA in Table 1; eg the two isolates with VNTR profile 63). On the other hand, there were examples where isolates from different regions and species of origin had identical VNTR profiles; for example VNTR profile 26 was shared by an isolate from a UK pig, an isolate from a human from Australia, and two isolates from humans in Italy; profile 37 was shared by an isolate from a US pig and one from an Australian chicken; profile 54 was shared by an isolate from an Australian chicken and one from an Australian human being.

4. Discussion

Using the genome sequence of *B. pilosicoli* 95/1000, 22 loci were identified as being potentially suitable regions for MLVA analysis. This was very similar to the 23 identified in the MLVA study conducted on *B. hyodysenteriae* (Hidalgo et al., 2010). Following further investigation only five loci were identified which were present in at least half of the tested isolates, and that had at least four variations between the amplified bands. This was in contrast to the situation with *B. hyodysenteriae* where eight suitable loci were identified (Hidalgo et al., 2010). Typically the number of loci used for MLVA for other bacteria has been between 5-8; for example, five loci have been used for *Salmonella* Typhimurium (Lindstedt et al., 2004), six for *Enterococcus faecium* (Top et al., 2004), and eight for *Staphylococcus aureus* (Malachowa et al., 2005). Hence, even with only five suitable loci identified, there was potential for these to be used in MVLA of *B. pilosicoli.* Locus BPilo_8 was the least suitable of the five, since it was only present in around one third of the isolates, but it was retained in the analysis as a number of alleles were detected in the isolates from which it was amplified, and no other more suitable loci were identified.

Using these five loci with 119 isolates, 103 VTs were produced. Two isolates did not amplify with any of the loci, and indeed there were only 16 profiles (20 isolates) that included allleles from all five loci. The presence of such "null alleles" can also act as supporting evidence for strain identification, although clearly they are not as definitive as when alleles are detected. In the previous study on *B. hyodysenteriae* using eight loci (Hidalgo et al., 2010), 44 VTs were obtained from 174 isolates, and the discriminatory power was 0.938. This compared with a D value of 0.976 and 103 VTs for the *B. pilosicoli* strains, which together demonstrate the existence of greater genetic variation in *B. pilosicoli* than in *B. hyodysenteriae*.

241

242 The MLVA method with five loci was a simple, rapid and replicable way to
243 identify different strains of *B. pilosicoli*, and as such could be a very useful tool for
244 molecular epidemiological studies. Evidence for its usefulness for strain typing can be
245 deduced from the situation where pairs or sets of isolates from the same regions and
246 species were shown to be similar or identical. The existence of isolates from different
247 species with the same or similar VTs also points to the likelihood that cross-species
248 transmission of *B. pilosicoli* strains occurs under natural conditions.

249

250 MLVA is likely less reliable than some other methods for assessing broad
251 phylogenetic relationships of *B. pilosicoli* strains, even where all five loci are
252 amplified (supplementary figure). The reason for this is that *B. pilosicoli* has a
253 recombinant population structure (Trott et al., 1998), and strains show evidence of
254 gene rearrangements (Zuerner et al., 2002) and gene loss (Mappley et al., 2012);
255 consequently MLVA loci, which have no known selective advantage, may readily
256 become disrupted. Indirect evidence for this was the presence of the numerous
257 apparent "null" alleles amongst the isolates investigated. Hence these localized
258 changes may blur understanding of the underlying core relationships between
259 different strains as identified by MLVA. Phylogeny is probably better studied by
260 analyzing conserved genes with core metabolic functions ("housekeeping genes"), as
261 used in multilocus enzyme electrophoresis (Trott et al., 1998; Hampson et al., 2006)
262 and multilocus sequence typing (Råsbäck et al., 2007; La et al., 2009).

263

264 Some modifications could be considered to improve MLVA as a typing method.
265 As five different primer sets were used, the possibility of multiplexing the reaction

266 could be explored. However, due to the variation in size between the different bands,
267 it would be difficult to be confident that the bands produced originated from the
268 specific VNTR locus. To get a more accurate prediction of the size of the VNTRs,
269 capillary electrophoresis could be used, or the amplicons could be sequenced, as
270 described in the MLVA study of *B. hyodysenteriae* (Hidalgo et al., 2010). This would
271 be especially appropriate for the VNTRs that had a small repeat size (Bpilo_8 and
272 Bpilo_16); however, to do this the MLVA method would require more sophisticated
273 equipment, be more expensive to run and would have a longer turnover time, and this
274 would defeat the aim of developing MLVA as a rapid and simple method for typing
275 *B. pilosicoli* isolates in epidemiological studies.
276
277 **Conflict of interest**
278 There are no conflicts of interest.
279
280 **References**
281 Atyeo, R.F., Oxberry, S.L., Hampson, D.J., 1996. Pulsed-field gel electrophoresis for
282 sub-specific differentiation of *Serpulina pilosicoli* (formerly *'Anguillina coli'*).
283 FEMS Microbiol. Lett. 141, 77-81.
284 Benson G., 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucl.
285 Acid. Res. 27, 573-580.
286 Grundmann, H., Hori S., Tanner G., 2001. Determining confidence intervals when
287 measuring genetic diversity and the discriminatory abilities of typing methods for
288 microorganisms. J. Clin. Microbiol. 39, 4190–4192.
289 Hampson, D.J., 2012. *Brachyspiral* colitis. In: Zimmerman, J.J., Karriker, L.A.,
290 Ramirez, A., Schwartz, K.J., Stevenson, G.W. (Eds.), Diseases of Swine 10th Ed.

Wiley-Blackwell, Iowa, USA. pp. 680-696.

Hampson, D.J., Oxberry, S.L., La, T., 2006. Potential for zoonotic transmission of *Brachyspira pilosicoli*. Emerg. Infect. Dis. 12, 869–870.

Hidalgo, Á., Carvajal, A. La, T, Naharro,G,. Rubio, P, Phillips, N.D., Hampson, D.J., 2010. A multiple-locus variable-number tandem repeats analysis of the swine dysentery pathogen, *Brachyspira hyodysenteriae.* J. Clin. Microbiol. 48, 2859-2865.

Jolley, K.A., Feil, E.J., Chan, M.S., Maiden, M.C., 2001. Sequence type analysis and recombinational tests (START). Bioinformatics 17, 1230-1231.

Kunkle, R.A., Harris, D.L., Kinyon, J.M., 1986. Autoclaved liquid medium for propagation of *Treponema hyodysenteriae*. J. Clin. Microbiol. 24, 669-671.

La, T., Phillips, N.D., Hampson, D.J., 2003. Development of a duplex PCR assay for the detection of *Brachyspira hyodysenteriae* and *Brachyspira pilosicoli* in pig feces. J. Clin. Microbiol. 41, 3372-3375.

La, T., Phillips, N.D., Harland, B.L., Wanchanthuek, P., Bellgard, M.I., Hampson, D.J., 2009. Multilocus sequence typing as a tool for studying the molecular epidemiology and population structure of *Brachyspira hyodysenteriae*. Vet. Microbiol. 138, 330–338.

Lee, J.I., Hampson, D.J., 1992. Intestinal spirochaetes colonising Aborigines from communities in the remote north of Western Australia. Epidemiol. Infect. 109, 133-141.

Lindstedt B.A., Vardund T., Aas L., Kapperud G., 2004. Multiple-locus variable-number tandem-repeats analysis of *Salmonella enterica* subsp. enterica serovar Typhimurium using PCR multiplexing and multicolor capillary electrophoresis. J. Microbiol. Meth. 59, 163–172.

316 Malachowa N., Sabat A., Gniadkowski M., Krzyszton-Russjan J., Empel J.,
317 Miedzobrodzki J., Kosowska-Shick K., Appelbaum P.C., Hryniewicz W., 2005.
318 Comparison of multiple-locus variable-number tandem-repeat analysis with pulsed-
319 field gel electrophoresis, spa typing, and multilocus sequence typing for clonal
320 characterization of *Staphylococcus aureus* isolates. J. Clin. Microbiol. 43, 3095-
321 3100.
322 Mappley, L.J., Black, M. L., AbuOun, M., Darby, A.C., Woodward, M.J., Parkhill, J.,
323 Turner, A.K., Bellgard, M.I., La, T., Phillips, N.D., La Ragione, R.M., Hampson,
324 D.J., 2012. Comparative genomics of *Brachyspira pilosicoli* strains: genome
325 rearrangements, reductions and correlation of genetic compliment with phenotypic
326 diversity. BMC Genomics 2012 Sep 5;13(1):454.
327 Margawani, K.R., Robertson, I.D., Brooke, C.J., Hampson, D.J., 2004. Prevalence,
328 risk factors and molecular epidemiology of *Brachyspira pilosicoli* in humans on the
329 island of Bali, Indonesia. J. Med. Microbiol. 53, 325–332.
330 Nelson, E.J., Tanudra, A., Chowdhury, A., Kane, A.V., Qadri, F., Calderwood, S.B.,
331 Coburn, J., Camilli, A., 2009. High prevalence of spirochetosis in cholera patients,
332 Bangladesh. Emerg. Infect. Dis. 15, 571–573.
333 Råsbäck, T., Johansson, K.-E., Jansson, D.S., Fellstrom, C., Alikhani, M.Y., La, T.,
334 Dunn, D.S., Hampson, D.J., 2007. Development of a multilocus sequence typing
335 scheme for intestinal spirochaetes within the genus *Brachyspira*. Microbiology 153,
336 4074–4087.
337 Stanton, T.B., 2006. The genus *Brachyspira*. In: Falkow, S., Rosenberg, S.E.,
338 Schleifer, K.H., Stackebrant, E., (Eds.), The Prokaryotes (Volume 7). Springer,
339 New York, USA, pp. 330–356.

340 Top J., Schouls L.M., Bonten M.J.M., Willems R.J.L., 2004. Multiple-locus variable-
341 number tandem repeat analysis, a novel typing scheme to study the genetic
342 relatedness and epidemiology of *Enterococcus faecium* isolates. J. Clin. Microbiol.
343 42, 4503-4511.
344 Trott, D.J., Combs, B.G., Oxberry, S.L., Mikosza, A.S.J., Robertson, I.D., Passey, M.,
345 Taime, J., Sehuko, R., Hampson, D.J., 1997a. The prevalence of *Serpulina*
346 *pilosicoli* in humans and domestic animals in the Eastern Highlands of Papua New
347 Guinea. Epidemiol. Infect. 119, 369–379.
348 Trott, D.J., Jensen, N.S., Saint Girons, I., Oxberry, S.L., Stanton, T.B., Lindquist, D.,
349 Hampson, D.J., 1997b. Identification and characterization of *Serpulina pilosicoli*
350 isolates recovered from the blood of critically ill patients. J. Clin. Microbiol. 35,
351 482-485.
352 Trott, D.J., Mikosza, A.S.J., Combs, B.G., Oxberry, S.L., Hampson, D.J., 1998.
353 Population genetic analysis of *Serpulina pilosicoli* and its molecular epidemiology
354 in villages in the Eastern Highlands of Papua New Guinea. Int. J. Syst. Bacteriol.
355 48, 659–668.
356 Verlinden, M., Pasmans, F., Garmyn, A., De Zutter, L., Haesebrouck, F., Martel, A.,
357 2012. Occurrence of viable *Brachyspira* spp. on carcasses of spent laying hens from
358 supermarkets. Food Microbiology 32, 321-324.
359 Wanchanthuek, P., Bellgard, M.I., La, T., Ryan, K., Moolhuijzen, P., Chapman, B.,
360 Black, M., Schibeci, D., Hunter, A., Barrero, R., Phillips, N.D., Hampson, D.J.,
361 2010. The complete genome sequence of the pathogenic intestinal spirochete
362 *Brachyspira pilosicoli* and comparison with other *Brachyspira* genomes PLoS One
363 5(7): e11455.
364 Zuerner, R.L., Stanton, T.B., Minion, F.C., Li, C., Charon, N.W., Trott, D.J.,

365 Hampson, D.J., 2004. Genetic variation in *Brachyspira*: chromosomal
366 rearrangements and sequence drift distinguish *B. pilosicoli* from *B. hyodysenteriae.*
367 Anaerobe 10, 229-237.
368

Figure legends

Fig. 1. Phylogenetic tree for 117 *B. pilosicoli* isolates constructed using the NJ method generated with the allelic profile obtained from MLVA. The names of the isolates and the numbers on the branches correspond with those shown in Table 1. The length of the scale bar represents one locus difference per unit.

Supplementary Figure. Phylogenetic tree for 20 *B. pilosicoli* isolates with all five alleles available, constructed using the NJ method generated with the allelic profile obtained from MLVA. The names of the isolates and the numbers on the branches correspond with those shown in Table 1 and Figure 1. The length of the scale bar represents one locus difference per unit.

Table 1

Names of the 119 isolates, the host species from which they were isolated, location where they were isolated, their position in Figure 1 and their VNTR profile

Strain name	Position on Figure 1	Species of origin	Location of isolation[a]	VNTR profile
V1 H 106	1	Human	PNG (village 1)	70
GP 6	2	Pig	PNG (piggery 1)	70
MC F1	3	Human	Australia (WA)	75
D17	4	Dog	Australia	16
V1 H 11	5	Human	PNG (village 1)	65
W7	6	Pig	Australia	51
IMR 48	7	Human	PNG	14
Q98.0026.12	8	Chicken	Australia	15
Q98.0027.36	9	Chicken	Australia	18
V1 H 103	10	Human	PNG (village 1)	12
16242-94	11	Dog	USA	82
Q97.2224.3.1	12	Chicken	Australia	92
IMR 2	13	Human	PNG	73
IMR 39	14	Human	PNG	77
2152	15	Horse	Australia	102
Q98.000.6.1	16	Chicken	Australia	56
*H21	17	Human	Australia	61
*OF 15	18	Pig	Australia	68
Q98.0228.5.2	19	Chicken	Australia	95
V1 H 126	20	Human	PNG (village 1)	80
GP 24	21	Pig	PNG (piggery 1)	89
HRM7	22	Human	Italy	27
*HRM7A	23	Human	Italy	62
GP 17	24	Pig	PNG (piggery 1)	81
Q94.0354.0.6	25	Chicken	Australia	59
GP 20	26	Pig	PNG (piggery 1)	76
V1 H 116	27	Human	PNG (village 1)	6
V1 D 1	28	Dog	PNG (village 1)	7
95/1000	29	Pig	Australia	88
GP 28	30	Pig	PNG (piggery 1)	72

GP 36	31	Pig	PNG (piggery 1)	71
Gap 418	32	Human	Australia	1
V1 H 120	33	Human	PNG (village 1)	8
Q98.0228.5.7	34	Chicken	Australia	74
EM1	35	Human	Australia (WA)	83
H4-2	36	Human	Australia	83
Gap 51.2	37	Human	Australia	83
V1 H 141	38	Human	PNG (village 1)	84
SC1	39	Human	Australia (WA)	63
JF2	40	Human	Australia (WA)	63
B1	41	Human	Australia (WA)	64
IMR 81	42	Human	PNG	11
NZ 91/31349	43	Pig	NZ	87
TH NF	44	Human	Australia (WA)	3
Q97.000.6.10	45	Chicken	Australia	38
*K1	46	Human	Australia (WA)	66
H43-2	47	Human	Australia	67
GP 32	48	Pig	PNG (piggery 1)	9
IMR49	49	Human	PNG	10
*GP 14	50	Pig	PNG (piggery 1)	5
V1 H 12	51	Human	PNG (village 1)	20
*QF1	52	Human	Australia (WA)	22
N1	53	Human	Australia (WA)	24
Q96.1037.0	54	Chicken	Australia	45
*WIA8	56	Pig	Australia	50
WW25	55	Pig	Australia	46
OF 11	57	Pig	Australia	39
89/1069	58	Pig	Canada	40
*P43/6/78^{T}	59	Pig	UK	29
GP 49	60	Pig	PNG	35
Q97.3008.4.2	61	Chicken	Australia	37
D9201243	62	Pig	USA	37
*W015/C138	63	Pig	Australia	44
JJ1	64	Human	Australia (WA)	25
M1	65	Human	Australia (WA)	36
PWS/B	66	Pig	UK	26
RV1	67	Human	Australia (WA)	26
HRM 4B	68	Human	Italy	26

HRM 2B	69	Human	Italy	26
L72	70	Pig	Australia	49
GP 35	71	Pig	PNG (piggery 1)	57
Q98.0072.37	72	Chicken	Australia	69
WG6	73	Pig	Australia	30
Q95.3281.0	74	Chicken	Australia	85
GP 3	75	Pig	PNG (piggery 1)	53
Q98.0033.72	76	Chicken	Australia	55
WesB	77	Human	Australia (WA)	54
Q98.0072.08	78	Chicken	Australia	54
89-2005A	79	Pig	Canada	60
KP5	80	Pig	Australia	41
KP1	81	Pig	Australia	90
Q97.000.6.2	82	Chicken	Australia	91
89-2005B	83	Pig	Canada	13
9803.1	84	Pig	Australia	21
Meyers K-9 12	85	Dog	USA	94
WF1	86	Pig	Australia	58
Q1588.5	87	Pig	Australia	32
Q98.0028.3	88	Chicken	Australia	32
Cof 10	89	Pig	Australia	34
WG	90	Pig	Australia	34
OF 2	91	Pig	Australia	33
Q98.0062.14	92	Chicken	Australia	47
Q97.000.6.8	93	Chicken	Australia	48
3295.90B	94	Pig	Australia	42
Q98.0026.11	95	Chicken	Australia	31
*Q97.000.6.4	96	Chicken	Australia	43
Q97.000.6.7	97	Chicken	Australia	43
H54	98	Human	Australia	28
H60-2	99	Human	Australia	19
GP 44	100	Pig	PNG (piggery 1)	3
HRM 2A	101	Human	Italy	52
Q97.000.6.22	102	Chicken	Australia	99
Q97.2110.4.1	103	Chicken	Australia	99
Ard 127	104	Pig	UK	93
04-2152	105	Horse	Australia	97

28/94	106	Human (blood)	France	97
88-3769	107	Pig	Canada	2
GP 42	108	Pig	PNG (piggery 1)	17
89-223A	109	Pig	Canada	98
B1555A	110	Pig	USA	100
BR81/80	111	Human (blood)	France	78
GP 5	112	Pig	PNG (piggery 1)	86
24072-93A	113	Dog	USA	96
42167	114	Chicken	USA	101
H38-2	115	Human	Australia	79
UNL-5	116	Dog	USA	4
V1 H 117	117	Human	PNG (village 1)	88
UNL-3	-	Pig	USA	103
Q98.0072.31	-	Chicken	Australia	103

[a] PNG, Papua New Guinea (Trott et al., 1997a); WA, Western Australia - marked for human isolates from an Aboriginal community (Lee and Hampson, 1992); NZ, New Zealand

*indicates the 10 isolates used in the initial screening

Isolates in adjacent shaded rows have identical VNTR types

Table 2

Alignment score, period size of the tandem repeat, the number of tandem repeats, consensus size of the VNTR, and the size of the amplicon using the corresponding primers with the template strain *B. pilosicoli* 95/1000

Loci	Alignment score	Period size	Copy no.	Consensus size	Sequencing size
BPilo_4	1262	102	7.7	785	1780
BPilo_8	451	75	3.1	231	1226
BPilo_9	803	159	2.9	461	1456
BPilo_14	278	102	2.7	280	1275
BPilo_16	806	72	6.2	478	1444

Table 3

Numbers of isolates carrying each allele at five loci

		Number of isolates carrying each allele									
Locus	Total number of alleles	1	2	3	4	5	6	7	8	9	No amplification
BPilo_4	8	2	12	3	1	2	37	8	12	-	42
BPilo_8	5	1	3	13	18	5	-	-	-	-	79
BPilo_9	9	2	3	16	2	15	11	30	10	4	26
BPilo_14	4	1	12	72	18	-	-	-	-	-	16
BPilo_16	5	20	13	17	15	12	-	-	-	-	42