The Journal of Microbiology (2011) Vol. 49, No. 5, pp. 768-775
DOI 10.1007/s12275-011-1084-3

A Genome-Wide Identification of Genes Potentially Associated with Host Specificity of *Brucella* Species

Kyung Mo Kim[1], Kyu-Won Kim[2], Samsun Sung[2], and Heebal Kim[2]*

[1]*Korean Bioinformation Center (KOBIC), Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon 305-806, Republic of Korea*
[2]*Department of Agricultural Biotechnology and Research Institute for Agriculture and Life Sciences, Seoul National University, Seoul 151-742, Republic of Korea*

(Received February 17, 2011 / Accepted May 20, 2011)

***Brucella* species are facultative intracellular pathogenic α-Proteobacteria that can cause brucellosis in humans and domestic animals. The clinical and veterinary importance of the bacteria has led to well established studies on the molecular mechanisms of *Brucella* infection of host organisms. However, to date, no genome-wide study has scanned for genes related to the host specificity of *Brucella* spp. The majority of bacterial genes related to specific environmental adaptations such as host specificity are well-known to have evolved under positive selection pressure. We thus detected signals of positive selection for individual orthologous genes among *Brucella* genomes and identified genes related to host specificity. We first determined orthologous sets from seven completely sequenced *Brucella* genomes using the Reciprocal Best Hits (RBH). A maximum likelihood analysis based on the branch-site test was accomplished to examine the presence of positive selection signals, which was subsequently confirmed by phylogenetic analysis. Consequently, 12 out of 2,033 orthologous genes were positively selected by specific *Brucella* lineages, each of which belongs to a particular animal host. Extensive literature reviews revealed that half of these computationally identified genes are indeed involved in *Brucella* host specificity. We expect that this genome-wide approach based on positive selection may be reliably used to screen for genes related to environmental adaptation of a particular species and that it will provide a set of appropriate candidate genes.**

***Keywords*:** *Brucella*, evolution, genome, host specificity, positive selection

Comparisons of molecular sequences provide valuable information for understanding evolutionary forces (e.g., natural selection and neutral evolution). The ratio of nonsynonymous and synonymous substitution rates (d_N/d_S) has been regarded as the most reliable measure to identify the evolutionary forces acting on a given gene (Kimura, 1983; Ohta, 1992; Yang *et al*., 2000). This ratio can predict whether a gene has been under positive selection ($d_N/d_S>1$), negative selection ($d_N/d_S<1$) or neutral evolution ($d_N/d_S=1$) (Li *et al*., 1985). The d_N/d_S ratio cannot be calculated for individual sequences but is a measure of the whole signal for orthologous sequences of a gene. Therefore, it is not appropriate for estimating the strengths of the selection forces for individual sequences. However, the branch-site test, which was recently developed using maximum-likelihood models, allows codon changes of $d_N/d_S>1$ for a particular sequence of a given alignment. The test allows one to calculate the degree of the selection signals for individual sequences (Zhang *et al*., 2005). Using the likelihood ratio test (LRT) of statistical estimates between null and alternative models for a given alignment, we can evaluate which sequences have evolved under positive selection. Simulation studies showed that this method performed robustly even when few sequences are included in an alignment (Wong *et al*., 2004).

The genus *Brucella* belongs to the order α-Proteobacteria and consists of intracellular pathogenic bacteria that can cause brucellosis in several animals including humans. The 16S ribosomal RNA (rRNA) sequences of the species are nearly or completely identical to one another. Therefore, the genus was regarded as being monospecific according to the *Brucella* Taxonomic Subcommittee of the International Committee on Systematics of Prokaryotes. However, this classification cannot reflect different types of biovars, each of which is associated with a specific animal host (Vizcaino *et al*., 2000; Michaux-Charachon *et al*., 2002; Moreno *et al*., 2002). Despite the complete identity of the 16S rRNA sequences of *Brucella*, based in differences in host preferences, the *Brucella* Taxonomic Subcommittee recently decided that the genus consists of six species (Osterman and Moriyon, 2006): *B. abortus* for cattle, *B. canis* for dogs, *B. melitensis* for goats and sheep, *B. neotomae* for desert wood rats, *B. ovis* for sheep, and *B. suis* for swine, reindeer, and hares. Although some species such as *B. melitensis* and *B. suis* exhibit liquidity of host ranges depending on biovars, each of the *Brucella* species has a representative host preference (Vizcaino *et al*., 2000; Foster *et al*., 2009; Wattam *et al*., 2009). Among the six recognized *Brucella* species, *B. neotomae* is pathogenic to desert wood rats that are not domesticated.

Brucella species are responsible for reproductive abortion and the restriction of food supply of domestic animals. Additionally, the transmission of *Brucella* pathogens to humans causes a febrile disease (Delvecchio *et al*., 2002; Paulsen *et al*., 2002; Pappas *et al*., 2005). Given the importance of the

* For correspondence. E-mail: heebal@snu.ac.kr; Tel.: +82-2-880-4803; Fax: +82-2-883-8812

Table 1. *Brucella* strains examined in this study

Species	Strain	Accession number[a]	Symbol[b]	Reference[c]
B. abortus	9-941	I : NC_006932 II : NC_006933	*B_cattle_1*	Halling *et al*. (2005)
B. abortus	2308	I : NC_007618 II : NC_007624	*B_cattle_2*	Chain *et al*. (2005)
B. canis	ATCC 23365	I : NC_010103 II : NC_010104	*B_dogs*	Michaux-Charachon *et al*. (2002)
B. melitensis	16M	I : NC_003317 II : NC_003318	*B_goats*	Michaux-Charachon *et al*. (2002)
B. ovis	ATCC 25840	I : NC_009505 II : NC_009504	*B_sheep*	Michaux-Charachon *et al*. (2002)
B. suis	1330	I : NC_004310 II : NC_004311	*B_swine_1*	Michaux-Charachon *et al*. (2002)
B. suis	ATCC 23455	I : NC_010169 II : NC_010167	*B_swine_2*	Lavigne *et al*. (2005)

[a] 'I' and 'II' indicate chromosomes I and II, respectively.
[b] indicates the host of each *Brucella* strain. 'B' indicates *Brucella*. Two strains per host were sorted by '_1' and '_2', respectively.
[c] The host of each strain was determined by referring to the literature.

bacteria to the veterinary economy and public health, tens of complete *Brucella* genome sequences have been generated so far (Wattam *et al*., 2009; Bohlin *et al.*, 2010). Nevertheless, a few studies have been conducted among the *Brucella* genomes and these have largely focused on the comparison of genomic contents (e.g., G+C ratio, SNPs, the presence and absence of ORFs) as well as horizontal gene transfer (Paulsen *et al*., 2002; Chain *et al*., 2005; Halling *et al*., 2005; Wattam *et al*., 2009). To our knowledge, no in-depth study has examined the host specificity of *Brucella* on a genome-wide scale. We here accomplished genome-wide comparisons of seven *Brucella* genomes, identified orthologous genes, performed a

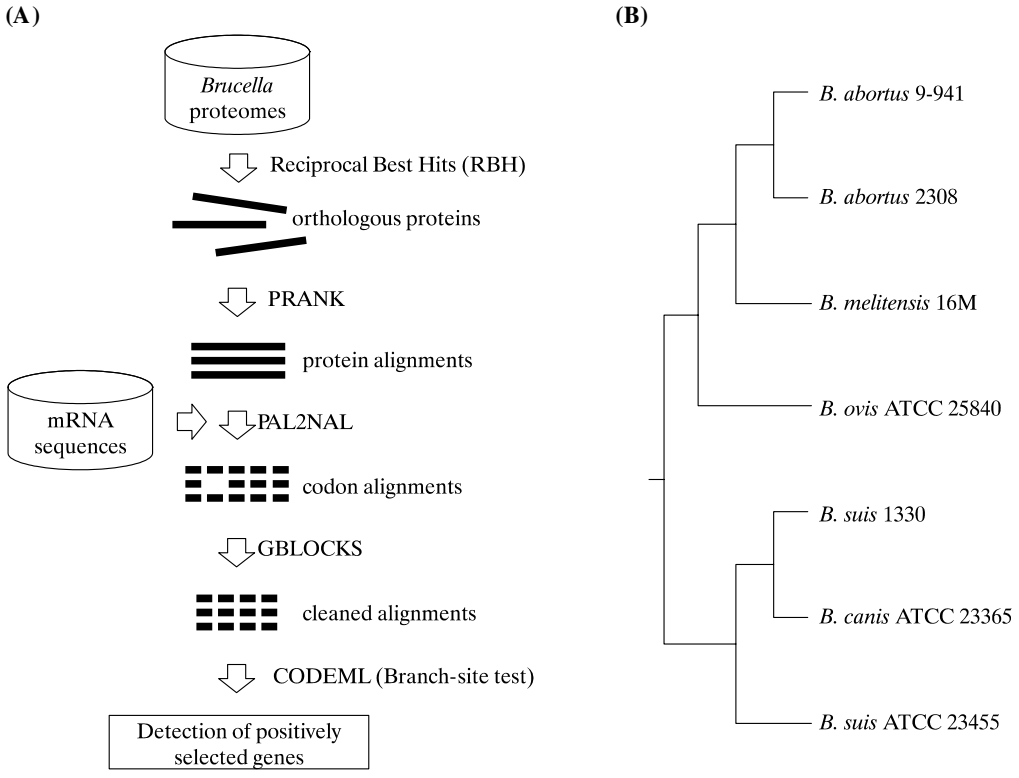

Fig. 1. Schematic representation of a positive selection analysis and the *Brucella* species tree. (A) The chart depicts all of the computational steps involved in the detection of orthologs and positively selected genes. A detailed description can be found in the Methods section of the text. (B) The rooted cladogram was reconstructed by referring to the phylogenetic tree reconstructed by Wattam *et al.* (2009).

positive selection analysis, and attempted a genome-wide screening for positively selected genes by individual animal hosts. An extensive literature review was conducted to determine whether the identified genes are indeed associated with the host specificity of *Brucella*.

Materials and Methods

Data preparation and determination of orthologs

From the RefSeq database of the NCBI, entire sets of protein sequences of seven *Brucella* genomes, each of which consists of two chromosomes, were retrieved as follows: 1) *B. abortus* 9-941 (NC_006932, NC_006933; Halling *et al*., 2005), 2) *B. abortus* 2308 (NC_007618, NC_007624; Chain *et al*., 2005), 3) *B. canis* ATCC 23365 (NC_010103, NC_010104; Wattam *et al.*, 2009), 4) *B. melitensis* 16M (NC_003317, NC_003318; Delvecchio *et al*., 2002), 5) *B. ovis* ATCC 25840 (NC_009505, NC_009504; J. Craig Venter Institute, unpublished data), 6) *B. suis* 1330 (NC_004310, NC_004311; Paulsen *et al*., 2002), and 7) *B. suis* ATCC 23455 (NC_010169, NC_010167; Wattam *et al*., 2009) (Table 1). We also collected mRNA sequences corresponding to the proteins in the NCBI database. Reciprocal Best Hits (RBH) was used to identify orthologous proteins among the seven *Brucella* genomes (Fig. 1A). For all possible pairwise combinations between protein sequences, all proteins of a *Brucella* strain were searched against those of the other strain using the stand-alone NCBI version of BLASTP ver. 2.2.16 with an *E* value cutoff of 10^{-7}. This threshold represents a stringent cutoff to minimize false positive hits in the BLAST analysis (Kim *et al.*, 2008; Moreno-Hagelsieb and Latimer, 2008)

Positive selection analysis

For individual ortholog sets resulting from the RBH, multiple protein sequence alignments were accomplished using the PRANK with the Hasegawa-Kishino-Yano (HKY) model with empirical base frequencies and kappa of 2 (Loytynoja and Goldman, 2005). Note that other alignment programs such as MUSCLE, MAFFT, and CLUSTAL W are not appropriate for positive selection analysis because they produce a problematic alignment that includes numerous insertions and deletions, which represent false positives (Fletcher and Yang, 2010). We then obtained multiple codon alignments of mRNA corresponding to protein sequence alignments using PAL2NAL ver. 12 (Suyama *et al*., 2006). Ambiguous sites were removed using GBLOCKS (Castresana, 2000). We then obtained the species tree of *Brucella* spp. that was reconstructed using a maximum-likelihood analysis of a concatenated alignment of 2,246 protein families (Wattam *et al*., 2009; Fig. 1B). Foreground sequences (branches) were determined based on the given tree. Using both the alignment and the tree, the branch-site test, which is sometimes referred to as 'test 2' was performed using CODEML of the PAML package ver. 4.3 (Yang, 1997) with the following options: the codon frequency model of F3X4, model *A* of Nssites=2, fix_omega=0, and omega=1; and model *A null* of Nssites=2, fix_omega=1, and omega=1. Test 2 of CODEML calculates the maximum-likelihood estimates of models *A* and *A null* for a given sequence dataset. Because only model *A* provides site parameters of $d_N/d_S > 1$ for given foreground sequences, we evaluated whether selected foreground sequences for each dataset have evolved under positive selection at a 5% significance level. *P* values were determined from the LRT scores $[2 \times (\ln L_{\text{modelA}} - \ln L_{\text{modelAnull}})]$ using a module Chi^2 P of the PAML package. The false discovery rate (FDR) was tested using the R package (Benjamini and Hochberg, 1995). A schematic representation that describes the entire workflow can be found in Fig. 1A.

Results and Discussion

Brucella genomes are more homogeneous than expected

The RBH analysis among proteins of the seven strains resulted in 2,033 protein ortholog sets. The average number of protein-coding genes in *Brucella* genomes is approximately 3,300; therefore, about 64% of the genes in a *Brucella* genome were orthologous to genes in another genome. The number of orthologous protein sets discovered here is in accordance with a recent analysis of ten *Brucella* genomes that produced 2,377 orthologous gene families (Wattam *et al.*, 2009). The protein sequences and the accession numbers of the orthologs are available in our own database (http://snugenome.snu.ac.kr/brucella).

Four genomes of *B. abortus*, *B. melitensis*, and *B. suis* listed in Table 1 have been reported to share 94% genomic DNA similarity (Chain *et al*., 2005). This has been supported by several *Brucella* studies (Paulsen *et al*., 2002; Halling *et al*., 2005; Wattam *et al*., 2009). The high degree of conservation among *Brucella* genomes is unexpected given that some saprophytic bacteria of the same genus share a small number of orthologs (e.g., 40% for three *Bacillus* genomes; Rey *et al*., 2004). Given this statistic, closely related pathogenic genomes such as *Brucella* species appear to have diverged less than closely related free-living genomes. A comparison of three genomes in the genus *Leptospira* revealed that the saprophyte *L. biflexa* genome exhibited 38% unique genes, while the parasites *L. borgpetersenii* and *L. interrogans* showed 8% and 18% unique genes, respectively (Picardeau *et al*., 2008). Parasites like *Brucella* species have more limited ecological niches than other free-living bacteria. With the exception of a few genes related to pathogenicity, most genes in parasite genomes are involved in core cellular processes and primary metabolism. Conversely, free-living bacteria require lineage-specific genes by extensive horizontal gene transfer to adapt to diverse environments (Dobrindt *et al*., 2004). Thus, free-living bacterial genomes appear to have evolved more diversely than parasite genomes, which led to homogenization of the genomes of parasites such as *Brucella*.

Few positively selected genes are involved in host specificity of *Brucella*

For each of the 2,033 orthologous sets, we computationally simulated which sequences in a set are positively selected. We first selected foreground sequences per ortholog set. Given that the sequences of *B_dogs*, *B_goats*, and *B_sheep* are present only once in a set, they were individually regarded as foreground branches. In the case of *B_cattle_1* and *_2*, a pair of two sequences was defined as a foreground branch because the two lineages of *B. abortus* form a clade in the provided *Brucella* species tree. The species tree, however, showed that the two *B. suis* lineages are not monophyletic with the *B. canis* branch. The phylogenetic relationship of the two species is questionable. However, a group of *B. suis* genomes has been reported to exhibit substantially greater genetic diversity than those of other *Brucella* genomes (Wattam *et al.*, 2009). In addition, a recent phylogenetic study of *Brucella*

Table 2. Positively selected genes with functional annotations

Dataset	Protein sequence[a]	Sequence homology[b]	Host	Function	2ΔL[c]	*P* value[d]	FDR[e]
7.1246	YP_222616.1, YP_415309.1, NP_698950.1, YP_001628401.1, YP_001259792.1, **NP_539010.1,** YP_001593780.1	(98, 2)	Goats	adenosylcobalamin-dependent diol dehydratase gamma subunit	25.3	4.83E-07	0.0002
7.150	YP_220885.1, YP_413604.1, NP_697152.1, YP_001626790.1, YP_001258150.1, **NP_540754.1,** YP_001591986.1	(99, 23)	Goats	glycosyltransferase	50.1	1.5E-12	2.93E-9
7.1520	YP_222627.1, YP_415320.1, NP_698961.1, YP_001628412.1, YP_001259803.1, **NP_538999.1,** YP_001593791.1	(99, 3)	Goats	Methyltransferase	34.1	5.2E-09	4.06E-06
7.156	YP_220949.1, YP_413669.1, NP_697219.1, YP_001626854.1, YP_001258212.1, **NP_540683.1,** YP_001592056.1	(98, 6)	Goats	sulfite reductase (ferredoxin)	18.5	1.73E-05	0.0059
7.1574	**YP_222446.1, YP_415141.1,** NP_698767.1, YP_001623018.1, YP_001259626.1, NP_539181.1, YP_001593596.1	(96, 14)	Cattle	branched-chain amino acid ABC transporter periplasmic substrate-binding protein	36.6	1.43E-09	1.24E-6
7.1710	YP_223503.1, YP_418924.1, NP_699670.1, YP_001622301.1, YP_001257458.1, **NP_541761.1,** YP_001594435.1	(75, 80)	Goats	glyoxalase	27.7	1.44E-07	8.64E-5
7.1882	YP_221326.1, YP_414036.1, NP_697581.1, YP_001627246.1, YP_001258564.1, **NP_540285.1,** YP_001592426.1	(93, 11)	Goats	2-haloalkanoic acid dehalogenase I	59.1	1.52E-14	3.97E-11
7.1914	YP_223571.1, YP_418991.1, NP_699577.1, YP_001622215.1, YP_001257380.1, **NP_541860.1,** YP_001594342.1	(98, 9)	Goats	nitrogen fixation protein VnfA	31.5	2E-08	1.42E-05
7.2	YP_221362.1, YP_414073.1, NP_697621.1, YP_001627286.1, YP_001258600.1, NP_540251.1, **YP_001592468.1**	(97, 8)	Dogs	cytochrome *c*-type biogenesis protein	38.3	6.15E-10	6.87E-7
7.296	YP_222399.1, YP_415097.1, NP_698718.1, YP_001622966.1, YP_001259585.1, **NP_539222.1,** YP_001593547.1	(96, 7)	Goats	DeoR family transcriptional regulator	41.4	1.24E-10	1.62E-7
7.512	YP_223179.1, YP_418598.1, NP_700012.1, YP_001622617.1, YP_001257781.1, **NP_541427.1,** YP_001594785.1	(98, 9)	Goats	type I restriction-modification system restriction subunit	41.4	1.23E-10	1.62E-7
7.661	YP_221391.1, YP_414101.1, NP_697651.1, **YP_001627315.1,** YP_001258626.1, NP_540223.1, YP_001592497.1	(76, 84)	Swine	Outer membrane protein IIIA precursor	29.3	6.22E-08	4.05E-5

[a] The sequences are listed in the order of *B_cattle_1*, *B_cattle_2*, *B_swine_1*, *B_swine_2*, *B_sheep*, B_goats, and *B_dogs*. Positively selected sequences are in bold.
[b] Minimum (identity %, number of substitutions and gaps) of two sequences.
[c] $2(\ln L_{\text{alternative hypothesis}} - \ln L_{\text{null hypothesis}})$, where alternative and null models are model *A* and model *A null* for the branch-site test of positive selection, respectively (Zhang *et al*., 2005).
[d] *P* value was determined from a chi-square distribution with one degree of freedom.
[e] False discovery rate.

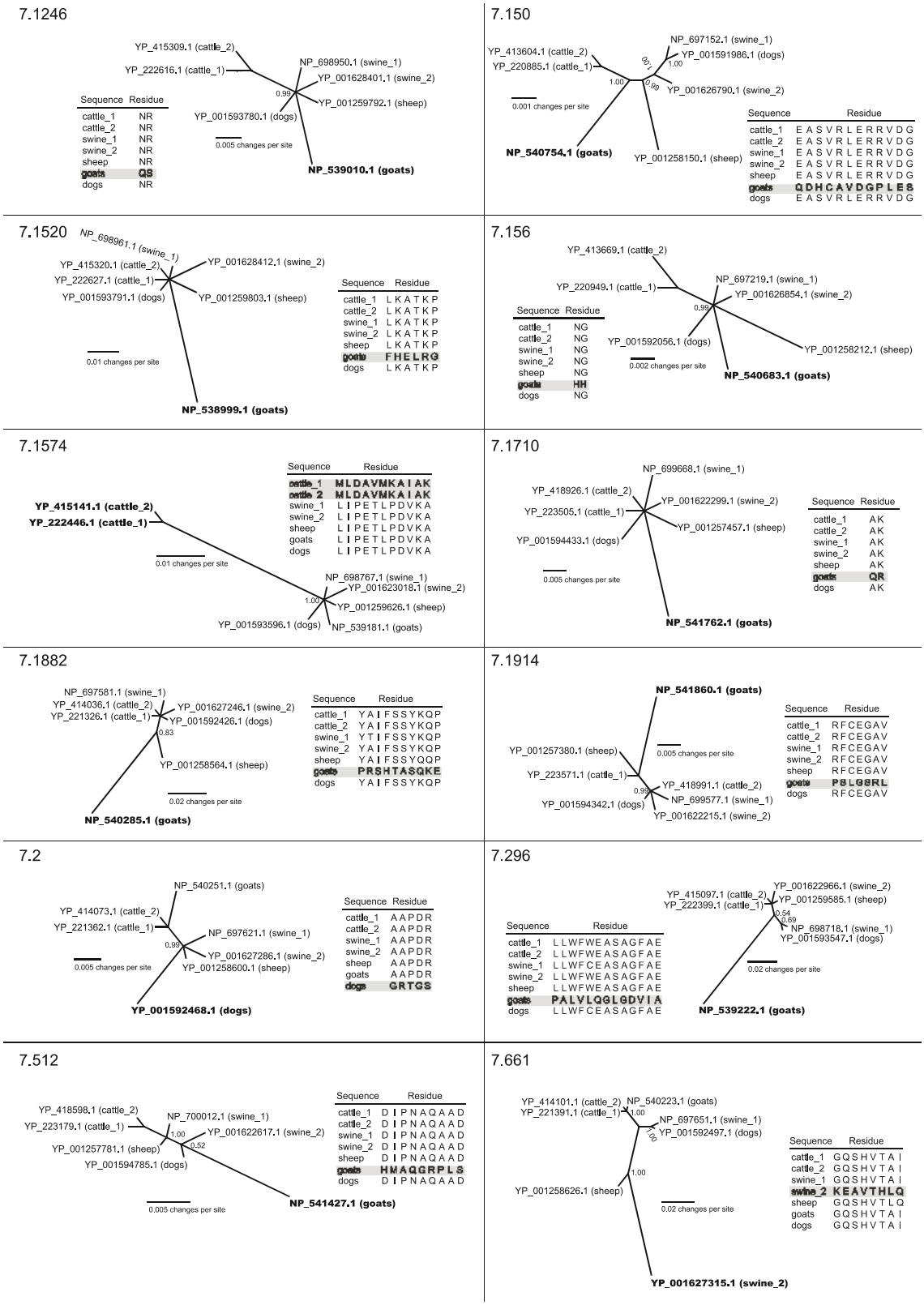

Fig. 2. Unrooted phylograms from Bayesian analysis of 12 orthologous sequence sets, each of which contains positively selected lineages. The taxon names that represent positively selected lineages are indicated in bold. Posterior probabilities with greater than 0.5 supports are shown above or below the branches. We collected positively selected sites from protein sequence alignments of each of the 12 datasets, and the amino-acid residues of positively selected lineages are labeled in gray.

genomes showed that *B. suis* 1330 is more closely clustered with *B. canis* ATCC 23365 than with *B. suis* ATCC 23445, concluding that *B. canis* was recently derived from one of the highly diverse *B. suis* strains (Foster *et al*., 2009). We thus dealt with the sequences of *B_swine_1* and *_2* separately as foreground sequences. Consequently, we found six different foreground sequences per set. For each of these, we accomplished branch-site test with the corresponding codon alignment and the given *Brucella* species tree. As a result, 14 gene sets with LRT scores greater than 3.8415 (5% significance level; degree of freedom=1) were chosen for the datasets, each of which exhibited significant evidence of positive selection (FDR<0.05 in Table 2). Two genes that are not annotated functionally were excluded from the study. To confidently check whether the 12 remaining genes are indeed under positive selection pressure, positively selected sites and phylogenetic trees were examined for individual gene sets (Fig. 2). All of the phylogenetic trees except for that of dataset 7.156 (sulfite reductase) showed that positively selected lineages determined by the statistical method are longer than other lineages, supporting that the statistical analysis successfully detect lineages in which recent mutations have been rapidly accumulating (Fig. 2). In addition, multiple protein sequence alignments that only display positively selected sites showed that nearly all amino-acid substitutions have occurred in the sequences of positively selected lineages (Fig. 2). These results indicate that the statistical method we used here is a reliable and powerful approach for detecting positive selection signature in a given sequence dataset. Consequently, we confirmed that only 12 out of 2,033 orthologs (0.5%) have positively selected lineages (Table 2; more detailed information at http://snugenome.snu.ac.kr/brucella/11_POSITIVE/). This fraction is similar to the result (0.6%; 23 out of 3,757 orthologs) of a positive selection analysis of genomes of virulent *Escherichia coli* and avirulent *Shigella flexneri* (Petersen *et al*., 2007). In addition, only a few orthologous genes have been suggested as being responsible for the host specificity of *Staphylococcus aureus* to several mammalian species including humans (Sung *et al*., 2008), indicating that bacterial adaptation to host environments is not likely to require complex repertoires and interactions of genes. Note, however that lineage-specific genes resulting from recent horizontal gene transfer and convergent can drive adaptation. Loss or pseudogenization of *Brucella* genes by reductive evolution also contributes to adaptation to intracellular lifestyle (Chain *et al*., 2005). Therefore, a need exists for more comprehensive studies to examine how different positively selected orthologs, species-specific genes, and reductive evolution act on the host-specific adaptation of *Brucella* species.

Genes detected by positive selection analysis are related to host specificity

Although the molecular mechanisms of *Brucella* infection into animal host cells have been well studied (Delrue *et al*., 2004; Roop II *et al*., 2009), little is known about the biological basis of *Brucella* host specificity. Several researchers concur that the difference in host specificity among *Brucella* species is the result of inactivation and loss of genes involved in cell surface structure, transport, and transcriptional regulation (Chain *et al*., 2005; Tsolis *et al*., 2009; Wattam *et al*., 2009). Thus, gene pseudogenization may be largely responsible for host specificity. In general, pseudogenes have weaker functional constraints than their counterparts and have evolved more rapidly. Therefore, a positive selection analysis based on a branch-site test can easily detect genes of this kind. Indeed, two out of the 14 genes discovered here have no known molecular function and were annotated as hypothetical proteins (Table 2). Even for the remaining 12 genes, a positively selected gene in one *Brucella* species may exhibit lower enzymatic activity than its orthologs in another *Brucella* species, depending on the extent of pseudogenization (Chain *et al*., 2005; Tsolis *et al*., 2009).

Note that our approach detected the outer membrane protein IIIA precursor (Table 2). The outer membrane structure of *Brucella* species is closely related to the processes of host-specific adaptation (Fernandez-Prada *et al*., 2003; Porte *et al*., 2003; Chain *et al*., 2005). Indeed, lipopolysaccharide (LPS) is an important counterpart of the outer membrane proteins because both play critical roles in determining the membrane structure. For this reason, it is unfortunate that our analysis regarded GDP-D-mannose 4,6-dehydratase as a false positive. GDP-D-mannose is catalyzed to GDP-4-keto-6-deoxy-D-mannose by the enzyme GDP-D-mannose 4,6-dehydratase. This serves as a starting point to produce GDP-6-deoxy-D-talose and GDP-D-perosamine in the mannose metabolic pathway. Given that talose and perosamine are components of LPS, a possible association exists between the positive selection of GDP-D-mannose 4,6-dehydratase and the host-specificity of *Brucella* species. Glycosyltransferase was also identified as a true positive. This enzyme could affect LPS structures and is involved in peptidoglycan metabolism, which determines the cell surface structure of individual *Brucella* species (Boschiroli *et al.*, 2001; Seleem *et al.*, 2008). Given that *Brucella* host specificity is partially attributable to variations in cell surface structures (Chain *et al*., 2005; Tsolis *et al*., 2009; Wattam *et al*., 2009), this glycosyltransferase could represent a true positive in this analysis. Our analysis also recognized periplasmic substrate-binding ABC transporter as a candidate gene (Table 2). Notably, ABC transporter is pseudogenized in the *B. ovis* genome (Tsolis *et al*., 2009). For this reason, *B. ovis* cannot utilize ribose and its derivatives. The metabolic defect can reduce the fitness of the species for various host animals because carbon sources that can be utilized by intracellular parasites often depend on the nature of the host animal cells. Consequently, *B. ovis* has a narrow host range, only infecting to sheep. On the other hand, the periplasmic substrate-binding ABC transporter detected in this study is embedded in the genomes of *B. melitensis* (goats) and *B. abortus* (cattle). Despite differences in animal hosts and permease substrates (e.g., ribose for *B. ovis*, glycine for *B. melitensis*), this transporter may also be responsible for *Brucella* host-specificity. Of the seven gene products not discussed, transcriptional regulators of the DeoR family and type I restriction-modification system restriction subunits are suspected to be true positives. The bacterial homologs of the two proteins are involved in the pathogenicity mechanisms (e.g., *Salmonella typhi*; Haghjoo and Galán, 2007) and maintenance of host specificity against horizontally transferred foreign genes (e.g., *S. aureus*; Waldron and Lindsay, 2006). Finally, methyltransferase, which is known to be related to

the attenuation of intracellular replication in macrophages of host animals of *Brucella* (Boschiroli *et al.*, 2001), could be a candidate enzyme involved in host specificity. However, the remaining six genes appear to be false negatives, or their biological significance is currently unknown.

Conclusions

We identified 12 genes that were positively selected after *Brucella* speciation. However, the lack of experimentally verified reference genes makes testing the reliability of the positive selection analysis for detecting genes of this kind difficult. Nevertheless, our literature reviews supported that half of the genes computationally discovered herein are either strongly or weakly involved in the host specificity of *Brucella* species, supporting the usefulness of the positive selection analysis. In addition, our results suggest that the evolutionary analysis presented here can be used to mine for genes responsible for the adaptation of bacteria species to specific environments. Note, however, that the generation of false positives and false negatives is inevitable given that our approach is based on statistical analyses. This may explain why our analysis did not detect the type IV secretion system, which is essential for *Brucella* virulence and plays important roles in intracellular replication and growth in macrophages of animal host cells (Watarai *et al*., 2002; Tsolis *et al*., 2009). Our positive selection analysis could be used as a preliminary screening tool to reduce trial and error and to increase the reliability of wet-lab experiments, although establishing a consensus is important in terms of which statistical significance cutoffs are appropriate for accurately detecting positively selected genes. In conclusion, the method presented here enables computational screening on a genome-wide scale for genes that have adapted to specific environments, and it can be used to avoid experimental labor by reducing the size of a gene set for further verification.

Acknowledgements

We are sincerely grateful to Dr. Ajith Harish for providing valuable comments. This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2010-013-C00027).

References

Benjamini, Y. and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)*. 57, 289-300.

Bohlin, J., L. Snipen, A. Cloeckaert, K. Lagesen, D. Ussery, A.B. Kristoffersen, and J. Godfroid. 2010. Genomic comparisons of *Brucella* spp. and closely related bacteria using base compositional and proteome based methods. *BMC Evol. Biol*. 10, 249.

Boschiroli, M.L., V. Foulongne, and D. O'Callaghan. 2001. Brucellosis: a worldwide zoonosis. *Curr. Opin. Microbiol*. 4, 58-64.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540-552.

Chain, P.S.G., D.J. Comerci, M.E. Tolmasky, F.W. Larimer, S.A. Malfatti, L.M. Vergez, F. Aguero, M.L. Land, R.A. Ugalde, and E. Garcia. 2005. Whole-genome analyses of speciation events in pathogenic Brucellae. *Infect. Immun.* 73, 8353-8361.

Delrue, R.M., P. Lestrate, A. Tibor, J.J. Letesson, and X.D. Bolle. 2004. *Brucella* pathogenesis, genes identified from random large-scale screens. *FEMS Micriobiol. Lett*. 231, 1-12.

Delvecchio, V.G., V. Kapatral, R.J. Redkar, G. Patra, C. Mujer, T. Los, N. Ivanova, and *et al*. 2002. The genome sequence of the facultative intracellular pathogen *Brucella melitensis*. *Proc. Natl. Acad. Sci. USA* 99, 443-448.

Dobrindt, U., B. Hochhut, U. Hentschel, and J. Hacker. 2004. Genomic islands in pathogenic and environmental microorganisms. *Nat. Rev. Microbiol*. 2, 414-424.

Fernandez-Prada, C.M., E.B. Zelazowska, M. Nikolich, T.L. Hadfield, R.M. Roop II, G.L. Robertson, and D.L. Hoover. 2003. Interactions between *Brucella melitensis* and human phagocytes: bacterial surface O-polysaccharide inhibits phagocytosis, bacterial killing, and subsequent host cell apoptosis. *Infect. Immun*. 71, 2110-2119.

Fletcher, W. and Z. Yang. 2010. The effect of insertions, deletions, and alignment errors on the branch-site test of positive selection. *Mol. Biol. Evol*. 27, 2257-2267.

Foster, J.T., S.M. Beckstrom-Sternberg, T. Pearson, J.S. Beckstrom-Sternberg, P.S. Chain, F.F. Roberto, J. Hnath, T. Brettin, and P. Keim. 2009. Whole-genome-based phylogeny and divergence of the genus *Brucella*. *J. Bacteriol.* 191, 2864-2870.

Haghjoo, E. and J.E. Galán. 2007. Identification of a transcriptional regulator that controls intracellular gene expression in *Salmonella typhi*. *Mol. Microbiol*. 64, 1549-1561.

Halling, S.M., B.D. Peterson-Burch, B.J. Bricker, R.L. Zuerner, Z. Qing, L.L. Li, V. Kapur, D.P. Alt, and S.C. Olsen. 2005. Completion of the genome sequence of *Brucella abortus* and comparison to the highly similar genomes of *Brucella melitensis* and *Brucella suis*. *J. Bacteriol*. 187, 2715-2726.

Kim, K.M., S. Sung, G. Caetano-Anollés, J.Y. Han, and H. Kim. 2008. An approach of orthology detection from homologous sequences under minimum evolution. *Nucleic Acids Res*. 36, e110.

Kimura, M. 1983. The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge, United Kingdom.

Lavigne, J.P., A.C. Vergunst, G. Bourg, and D. O'Callaghan. 2005. The IncP island in the genome *Brucella suis* 1330 was acquired by site-specific integration. *Infect. Immun.* 73, 7779-7783.

Li, W.H., C.I. Wu, and C.C. Luo. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol. Biol. Evol*. 2, 150-174.

Loytynoja, A. and N. Goldman. 2005. An algorithm for progressive multiple alignment of sequences with insertions. *Proc. Natl. Acad. Sci. USA* 102, 10557-10562.

Michaux-Charachon, S., E. Jumans-Bilak, A. Allardet-Servent, G. Bourg, M.L. Boschiroli, M. Ramuz, and D. O'Callaghan. 2002. The *Brucella* genome at the beginning of the post-genomic era. *Vet. Microbiol*. 90, 581-585.

Moreno, E., A. Cloeckaert, and I. Moriyon. 2002. *Brucella* evolution and taxonomy. *Vet. Microbiol*. 90, 209-227.

Moreno-Hagelsieb, G. and K. Latimer. 2008. Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics* 24, 319-324.

Ohta, T. 1992. The nearly neutral theory of molecular evolution. *Ann. Rev. Ecol. Syst*. 23, 263-286.

Osterman, B. and I. Moriyon. 2006. International Committee on Systematics of Prokaryotes; Subcommittee on the taxonomy of *Brucella*: Minutes of the meeting, 17 September 2003, Pamplona, Spain. *Int. J. Syst. Evol. Microbiol.* 56, 1173-1175.

Pappas, G., N. Akritidis, M. Bosilkovski, and E. Tsianos. 2005. Brucellosis. *N. Engl. J. Med*. 352, 2325-2336.

Paulsen, I.T., R. Seshadri, K.E. Nelson, J.A. Eisen, J.F. Heidelberg, T.D. Read, R.J. Dodson, and *et al*. 2002. The *Brucella suis* genome reveals fundamental similarities between animal and plant pathogens and symbionts. *Proc. Natl. Acad. Sci. USA* 99, 13148-13153.

Petersen, L., J.P. Bollback, M. Dimmic, M. Hubisz, and R. Nielsen.

2007. Genes under positive selection in *Escherichia coli*. *Genome Res*. 17, 1336-1343.

Picardeau, M., D.M. Bulach, C. Bouchier, R.L. Zuerner, N. Zidane, P.J. Wilson, S. Creno, and *et al*. 2008. Genome sequence of the saprophyte *Leptospira biflexa* provides insights into the evolution of *Leptospira* and the pathogenesis of leptospirosis. *PLoS ONE* 3, e1607.

Porte, F., A. Naroeni, S. Ouahrani-Bettache, and J.P. Liautard. 2003. Role of the *Brucella suis* lipopolysaccharide O antigen in phagosomal genesis and in inhibition of phagosome-lysosome fusion in murine macrophages. *Infect. Immun*. 71, 1481-1490.

Rey, M.W., R. Ramaiya, B.A. Nelson, S.D. Brody-Karpin, E.J. Zaretsky, M. Tang, A.L. de Leon, and *et al*. 2004. Complete genome sequence of the industrial bacterium *Bacillus* licheniformis and comparisons with closely related *Bacillus* species. *Genome Biol*. 5, r77.

Roop II, R.M., J.M. Gaines, E.S. Anderson, C.C. Caswell, and D.W. Martin. 2009. Survival of the fittest: how *Brucella* strains adapt to their intracellular niche in the host. *Med. Microbiol. Immunol*. 198, 221-238.

Seleem, M.N., S.M. Boyle, and N. Sriranganathan. 2008. *Brucella*: a pathogen without classic virulence genes. *Vet. Microbiol*. 129, 1-14.

Sung, J.M.L., D.H. Lloyd, and J.A. Lindsay. 2008. *Staphylococcus aureus* host specificity: comparative genomics of human versus animal isolates by multi-strain microarray. *Microbiology* 154, 1949-1959.

Suyama, M., D. Torrents, and P. Bork. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res*. 34, W609-612.

Tsolis, R.M., R. Seshadri, R.L. Santos, F.J. Sangari, J.M.G. Lobo, M.F. de Jong, Q. Ren, and *et al*. 2009. Genome degradation in *Brucella ovis* corresponds with narrowing of its host range and tissue tropism. *PLoS ONE* 4, e5519.

Vizcaino, N., A. Cloeckaert, J.M. Verger, M. Grayon, and L. Fernandez-Lago. 2000. DNA polymorphism in the genus *Brucella*. *Microbes Infect*. 2, 1089-1100.

Waldron, D.E. and J.A. Lindsay. 2006. Sau1: a novel lineage-specific type I restriction-modification system that blocks horizontal gene transfer into *Staphylococcus aureus* and between *S. aureus* isolates of different lineages. *J. Bacteriol.* 188, 5578-5585.

Watarai, M., H.L. Anderws, and R.R. Isberg. 2002. Formation of a fibrous structure on the surface of *Legionella pneumophila* associated with exposure of DotH and DotO proteins after intracellular growth. *Mol. Microbiol*. 39, 313-329.

Wattam, A.R., K.P. Williams, E.E. Snyder, N.F. Almeida, M. Shukla, A.W. Dickerman, O.R. Crasta, and *et al*. 2009. Analysis of ten *Brucella* genomes reveals evidence for horizontal gene transfer despite a preferred intracellular lifestyle. *J. Bacteriol*. 191, 3569-3579.

Wong, W.S., Z. Yang, N. Goldman, and R. Nielsen. 2004. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* 168, 1041-1051.

Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci*. 13, 555-556.

Yang, Z., R. Nielsen, N. Goldman, and A.M.K. Pedersen. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155, 431-449.

Zhang, J., R. Nielsen, and Z. Yang. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol*. 22, 2472-2479.