[S11-3]

Detection of Ortholog Sequences and Identification of Host-Specific Positive Selection in *Brucella* spp.

Heebal Kim* and Kyung Mo Kim

Laboratory of Bioinformatics and Population Genetics, Department of Agricultural Biotechnology, Seoul National University, Seoul 151-921

The understanding of complex traits related to various domesticated genes makes it possible to examine economically important genes, which can directly lead to mass production of domestic animals and plants. Furthermore, recent researches give us the possibility that the same technique of domesticated gene identification can be fully applied to the development of antibiotics related to human and animal pathogens. Here, we show that the defining orthologous proteins and host-specific positive selection in *Brucella* spp. is a good model example for the identification of domestication genes in animals and plants.

After the concept of homology was first introduced by Owen [1], it has been adopted as the basis of phylogenetics and comparative biology. Although there have been many arguments in interpreting the concept for decades [2] the term homology can be agreeably categorized into orthology and paralogy, which are due to gene duplication and speciation [3]. Orthology indicates the relationships among genetic elements diverged from their common ancestor along with speciation, whereas paralogy means the relationships between descendants produced by gene duplication in a single species [3-6]. Recent gene duplication without any further speciation produces co-orthologs, which can also be orthologous to genes in other species [6]. Because of the slight sequence dissimilarity between orthologs and paralogs, it is prone to regarding paralogs as orthologs. For this reason, several methods based on evolutionary distance, phylogeny and BLAST have tried to detect orthologs. Depending on their own algorithms, each of previous methods still causes the increase of false negative or false positive rates. For a new attempt, we developed a novel algorithm as a distance method using phylogenetic information based on the concept of minimum evolution. Our algorithm assumes that the sequences only consisting of orthologs requires evolutionary cost less than those including paralogous relationships. To calculate the evolutionary cost, it requires the reconstruction of a neighbor-joining (NJ) tree, but the calculated evolutionary cost is not affected by the topology of a given NJ tree. Unlike the method of tree reconciliation, our algorithm could be free from

the problem of incorrect topologies of species and gene trees.

As pathogenic microorganisms, *Brucella* species can cause brucellosis in several animals including human. There are some arguments for the species concepts of *Brucella*, but it has been generally accepted that the genus contains six species: *B. abortus*, *B. canis*, *B. melitensis*, *B. neotomae*, *B. ovis*, and *B. suis* [7, 8, 10]. Among six species, *B. abortus*, *B. melitensis*, and *B. suis* can be infectious to various animal hosts within the level of a species [e.g. *B. melitensis* of biotype 1-3: goats, sheep, camels in the references 8, 9]. In spite of the liquidity of host range under biotypes, there is a host preference for each Brucella species [10]. Furthermore, the host specificity of each *Brucella* species has been exactly defined in the level of the strain (Table 1). Although molecular mechanisms for behavior of Brucella infection into its hosts have studied well, there is no consideration for genes determining host-specificities of Brucella species. To screen these genes from the genomic data, we attempted to screen these genes under the assumption that genes adapted to a specific host could have signals for positive selection. For this work, we extracted 2,033 putative ortholog sets from all published Brucella genomes based on the reciprocal best hits. Among the ortholog sets, our maximum likelihood-based and phylogenetic analyses generated 10 gene sets as the datasets including positively selected sequence(s) related to host-specificity. We believe that the positively selected genes discovered here could be high-qualified source for experimental verification of Brucella host-specificity. Furthermore, computational procedures used in this study may be applied well to comparative genomic approaches that screen genes related to environmental adaptation of a specific organism.

Reference

1. Owen R, Lectures on the comparative anatomy and physiology of the invertebrate animals Longman, *Brown, Green & Longmans*, 1843.
2. Fitch WM. Homology a personal view on some of the problems. *Trends Genet.*, **16**, 227, 2000.
3. Graur D, Li WH. Fundamentals of molecular evolution, 2nd ed. - Dan Graur, Wen-Hsiung Li, Eds. Sunderland, Mass. *Sinauer Associates*, 2000.
4. Fitch WM, Distinguishing homologous from analogous proteins. *Syst. Zool.*, **19**, 99, 1970.
5. Sonnhammer ELL, Koonin EV. Orthology, paralogy and proposed classification for paralog subtypes. *Trends Genet*, **18**, 619, 2002.
6. Koonin EV. Orthologs, Paralogs, and Evolutionary Genomics. *Annu. Rev. Genet.*, **39**, 309, 2005.
7. Michaux-Charachon, S., E. Jumans-Bilak, A. Allardet-Servent, G. Bourg, M. L. Boschiroli, M. Ramuz, and D. O'Callaghan. The Brucella genome at the beginning of the post-genomicera. *Vet. Microbiol*, **90**, 581, 2002.

8. Moreno E, A. Cloeckaert, and I. Moriyon. Brucella evolution and taxonomy. *Vet. Microbiol*. **90**, 209, 2002.
9. Pappas, G., N. Akritidis, M. Bosilkovski, and E. Tsianos. Brucellosis. *N. Engl. J. Med*. **352**, 2325, 2005.
10. Vizcaino, N., A. Cloeckaert, J.-M. Verger, M. Grayon, and L. Fernandez-Lago. DNA polymorphism in the genus Brucella. *Microbes Infect*. **2**, 1089, 2000.