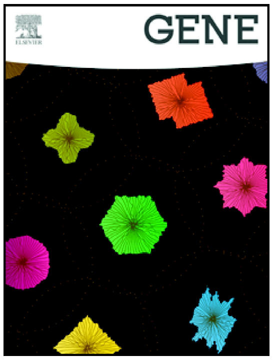

Ortholog-based screening and identification of genes related to intracellular survival

Xiaowen Yang, Jiawei Wang, Guoxia Bing, Pengfei Bie, Yanyan De, Yanli Lyu, Qingmin Wu

PII: S0378-1119(18)30073-8
DOI: https://doi.org/10.1016/j.gene.2018.01.059
Reference: GENE 42509

To appear in: *Gene*

Received date: 10 June 2017
Revised date: 29 October 2017
Accepted date: 17 January 2018

Please cite this article as: Xiaowen Yang, Jiawei Wang, Guoxia Bing, Pengfei Bie, Yanyan De, Yanli Lyu, Qingmin Wu , Ortholog-based screening and identification of genes related to intracellular survival. The address for the corresponding author was captured as affiliation for all authors. Please check if appropriate. Gene(2017), https://doi.org/10.1016/j.gene.2018.01.059

Ortholog-based screening and identification of genes related to intracellular survival

Xiaowen Yang[1], Jiawei Wang[2], Guoxia Bing[3], Pengfei Bie[1], Yanyan De[1], Yanli Lyu[1]* and Qingmin Wu[1]*

1. Key Laboratory of Animal Epidemiology and Zoonosis of the Ministry of Agriculture, College of Veterinary Medicine, China Agricultural University, Beijing 100193, China,
2. Animal Science and Technology College, Beijing University of Agriculture, Beijing 102206, China,
3. China Animal Disease Control Center, Beijing 102600, China.

∗Corresponding author: Key Laboratory of Animal Epidemiology and Zoonosis of the Ministry of Agriculture, College of Veterinary Medicine, China Agricultural University, Beijing 100193, People’s Republic of China.

Tel: +86-10-6273-3901; Fax: +86-10-6273-3901; E-mail: luyanli@cau.edu.cn，wuqm@cau.edu.cn

Abstract: Bioinformatics and comparative genomics analysis methods were used to predict unknown pathogen genes based on homology with identified or functionally clustered genes. In this study, the genes of common pathogens were analyzed to screen and identify genes associated with intracellular survival through sequence similarity, phylogenetic tree analysis and the λ-Red recombination system test method. The total 38,952 protein-coding genes of common pathogens were divided into 19,775 clusters. As demonstrated through a COG analysis, information storage and processing genes might play an important role intracellular survival. Only 19 clusters were present in facultative intracellular pathogens, and not all were present in extracellular pathogens. Construction of a phylogenetic tree selected 18 of these 19 clusters. Comparisons with the DEG database and previous research revealed that seven other clusters are considered essential gene clusters and that seven other clusters are associated with intracellular survival. Moreover, this study confirmed that clusters screened by orthologs with similar function could be replaced with an approved *uvrY* gene and its orthologs, and the results revealed that the *usg* gene is associated with intracellular survival. The study improves the current understanding of intracellular pathogens characteristics and allows further exploration of the intracellular survival-related gene modules in these pathogens.

Abbreviation

HGT, horizontal gene transfer; BLAST, Basic Local Alignment Search Tool; COG, Clusters of Orthologous Group of proteins; CDC, Chinese Centers for Disease Control and Prevention; TSA, tryptic soy agar; TSB, tryptic soy broth; ML, maximum likelihood method; DEG, Database of Essential Genes; FRT, flipase rercognition target; MOI, infection multiplicity; CFU, colony-forming units; ANOVA, one-way analysis of variance

Keywords: orthologs; intracellular survival; comparative genomics; *uvrY*; *usg*

1. Introduction

In the new genomic era, bioinformatics and comparative genomics are used to predict unknown genes of pathogens based on homology with identified or clustered functional genes. The evolutionary relationships among genes mainly included orthologs and paralogs. Orthologs are defined as genes from different species that evolved from a single gene in the last common ancestor, and these genes typically have the same or similar functions (Sonnhammer and Koonin, 2002). The original definition of paralogs was genes that evolved from a single gene and have many copies within a genome produced by duplication, and these genes usually have different functions. The latter definition of paralogs does not specify that they can only be found in a single organism (Gabaldon et al., 2009). Orthologs are usually more conserved in terms of biological structures and functions than paralogs because they have similar gene sequences (Chen et al., 2007). Thus, orthology was particularly useful for extending the biological knowledge of model organisms to