



Evaluating the annotation of protein-coding genes in bacterial genomes: *Chloroflexus aurantiacus* strain J-10-fl and *Natrinema* sp J7-2 as case studies

H.X. Zhang, S.J. Li and H.Q. Zhou

School of Life Science and Technology,
Key Laboratory for Neuroinformation of Ministry of Education,
University of Electronic Science and Technology of China, Chengdu, China

Corresponding author: H.X. Zhang
E-mail: johnzhangchina2007@gmail.com

Genet. Mol. Res. 13 (4): 10891-10897 (2014)
Received January 21, 2014
Accepted April 11, 2014
Published December 19, 2014
DOI <http://dx.doi.org/10.4238/2014.December.19.10>

ABSTRACT. Gene annotation plays a key role in subsequent biochemical and molecular biological studies of various organisms. There are some errors in the original annotation of sequenced genomes because of the lack of sufficient data, and these errors may propagate into other genomes. Therefore, genome annotation must be checked from time to time to evaluate newly accumulated data. In this study, we evaluated the gene density of 2606 bacteria or archaea, and identified 2 with extreme values, the minimum value (*Chloroflexus aurantiacus* strain J-10-fl) and maximum value (*Natrinema* sp J7-2), to conduct genome re-annotation. In the genome of *C. aurantiacus* strain J-10-fl, we identified 17 new genes with definite functions and eliminated 34 non-coding open-reading frames; in the genome of *Natrinema* sp J7-2, we eliminated 118 non-coding open reading frames. Our re-annotation procedure may provide a reference for improving the annotation of other bacterial genomes.

Key words: Gene density; Genome re-annotation; Missed genes; Over-annotated genes