

BG7: A new approach for bacterial genome annotation designed for Next Generation Sequencing data

Pablo Pareja-Tobes¹, Marina Manrique¹, Eduardo Pareja-Tobes¹, Eduardo Pareja¹
and Raquel Tobes¹,

¹Oh no Sequences! , Era7 Bioinformatics , Granada, Spain.

rtobes@era7.com

Abstract. Next Generation Sequencing technologies are generating genome assemblies with high error rates and fragmented in many contigs. Given the difficulties derived of this type of genome sequence it is mandatory the availability of a system able to do a high quality functional annotation even for very fragmented genomes with multiple sequence errors. BG7 is a new system for de novo bacterial, archaeal and viral genome annotation based on a new approach specifically designed for annotating genomes sequenced with next generation sequencing technologies. The system is versatile and able to annotate genes even in the step of preliminary assembly of the genome. It is especially efficient detecting unexpected genes horizontally acquired distant genomes, phages, plasmids, and mobile elements. BG7 predicts ORFs and infers their function based on protein similarity, integrating ORF prediction and functional annotation phases in just one step. BG7 is tolerant to sequencing errors in start and stop codons, frameshifts, and assembly or scaffolding errors. The new approach has been designed to tolerate all these errors frequently found in genomes obtained by means of NGS technologies. The system is especially tolerant to the fragmentation of genes frequent in not completely assembled genomes. BG7 is a fast, automated and scalable system that can cope with the challenge of analyzing the huge amount of genomes sequenced with NGS technologies. The suitability of BG7 for genome annotation has been proved for Illumina, 454, Ion Torrent, and PacBio sequencing technologies. Its flexibility for dealing with sequencing errors will facilitate its adaptation to future new technologies. Bg7 is an Open Source project licensed under the AGPLV3 license. More information at <http://bg7.ohnosequences.com/>

References

1. BG7: A new approach for bacterial genome annotation designed for Next Generation Sequencing data. Pareja-Tobes P, Manrique M, Pareja-Tobes E, Pareja E., Tobes R. PLOS ONE (November 21, 2012) in press