



VERSION 1

FEB 27, 2023

OPEN ACCESS

DOI:
dx.doi.org/10.17504/protocols.io.bc38iyrw

Protocol Citation: Qiye Li, Qunfei Guo, Yang Zhou, Huishuang Tan, Terry Bertozzi, Yuanzhen Zhu, Ji Li, Stephen Donnellan, Guojie Zhang 2023. Protein-coding gene annotation protocol for the eastern banjo frog.

protocols.io
<https://dx.doi.org/10.17504/protocols.io.bc38iyrw>

License: This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working

Created: Feb 28, 2020

Last Modified: Feb 27, 2023

PROTOCOL integer ID:
 33632

🌐 Protein-coding gene annotation protocol for the eastern banjo frog V.1

📁 In 1 collection

Qiye Li^{1,2}, Qunfei Guo^{1,3}, Yang Zhou¹, Huishuang Tan^{1,4}, Terry Bertozzi^{5,6}, Yuanzhen Zhu^{1,7}, Ji Li^{2,8}, Stephen Donnellan⁵, Guojie Zhang^{2,8,9,10}

¹BGI-Shenzhen, Shenzhen 518083, China;

²State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China;

³College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China;

⁴Center for Informational Biology, University of Electronic Science and Technology of China, Chengdu 611731, China;

⁵South Australian Museum, North Terrace, Adelaide 5000, Australia;

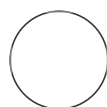
⁶School of Biological Sciences, University of Adelaide, North Terrace, Adelaide 5005, Australia;

⁷School of Basic Medicine, Qingdao University, Qingdao 266071, China;

⁸China National Genebank, BGI-Shenzhen, Shenzhen 518120, China;

⁹Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, 650223, Kunming, China;

¹⁰Section for Ecology and Evolution, Department of Biology, University of Copenhagen, DK-2100 Copenhagen, Denmark



Qunfei Guo

ABSTRACT

This pipeline is the protein-coding gene annotation by using homology-based and *de novo* predictions to build gene models for the *Limnodynastes dumerilii dumerilii* genome assembly.

Keywords: the eastern banjo frog, gene annotation, homology-based, de novo

Gene prediction preparation

- 1 Download protein sequences of diverse vertebrate species (*Danio rerio* (Ensembl, release-98), *Xenopus tropicalis* (Ensembl, release-98), *Xenopus laevis* (NCBI, GCF_001663975.1), *Nanorana parkeri* (NCBI, GCF_000935625.1), *Microcaecilia unicolor* (NCBI, GCF_901765095.1), *Rhinatrema bivittatum* (NCBI, GCF_901001135.1), *Anolis carolinensis* (Ensembl, release-98), *Gallus gallus* (Ensembl, release-98) and *Homo sapiens* (Ensembl, release-98)).

Homology-based prediction

- 2 Align the protein sequences from the previous step to the *Limnodynastes dumerilii dumerilii* genome assembly by using TBLASTN (blast-2.2.26).

Extract the genomic sequences of the candidate loci together with 5 kb flanking sequences.

Align the homologous proteins to the extracted genomic sequences from the previous step by using GeneWise (wise-2.2.0) for exon-intron structure determinations.

de novo prediction

- 3 Randomly pick 1,000 homology-derived gene models of *Limnodynastes dumerilii dumerilii* with complete open reading frames (ORFs) and reciprocal aligning rates exceeding 90% against the *X. tropicalis* proteins to train AUGUSTUS (v3.3.1, RRID:SCR_008417).

Use the obtained gene parameters by AUGUSTUS to predict protein-coding genes on the repeat-masked *Limnodynastes dumerilii dumerilii* genome assembly.

Gene models combination

- 4 Combine gene models derived from the above two methods into a non-redundant gene set using a similar strategy to Xiong *et al.* (2016).

Note

Xiong *et al.* (2016) : Xiong Z, Li F, Li Q, Zhou L, Gamble T, Zheng J, et al. Draft genome of the leopard gecko, *Eublepharis macularius*. Gigascience. 2016;5 1:47. doi:10.1186/s13742-016-0151-4.

Gene filtration

- 5 Remove genes showing BLASTP (blast-2.2.26) hits to transposon proteins in the UniProtKB/Swiss-Prot database (v2019_11), or with more than 70% of their coding regions overlapping repetitive sequences from the combined gene set.

Note

Use the parameters "-F F -e 1e-5" when running BLASTP.

Functional annotation

- 6 Finally, perform functional annotation by searching the *Limnodynastes dumerilii dumerilii* proteins against public databases of UniProtKB/Swiss-Prot (v2019_11), NCBI nr (v20191030), and KEGG (v93.0) with BLASTP (blast-2.2.26;).

Note

Use the parameters "-F F -e 1e -5" when running BLASTP.