Sequence analysis

TTA Lynx: a web-based service for analysis of actinomycete genes containing rare TTA codon

Nestor Zaburannyy*, Bohdan Ostash and Victor Fedorenko

Department of Genetics and Biotechnology, Ivan Franko National University of L'viv, L'viv 79005, Ukraine

Received on April 7, 2009; revised on June 15, 2009; accepted on June 26, 2009

Advance Access publication July 2, 2009

Associate Editor: Limsoon Wong

© The Author 2009. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org

ABSTRACT

Motivation: TTA Lynx is a web-based service for finding, assessing and comparing coding sequences that contain TTA codons. This codon is most notable for being a regulatory switch that governs different aspects of the physiology of several GC-rich, Gram-positive bacteria belonging to genus Streptomyces, prolific producers of clinically important natural products. The ever-increasing pace of genome sequencing is creating a huge volume of data that could be utilized to improve our understanding of rare codons in actinomycete biology (and other biological systems.) The service described here is designed to facilitate analysis of TTA-containing genes and to assess the importance of TTA-mediated regulation in an organism of interest. This service and its database of organisms with well-known or hypothetical TTA-based regulation provides an opportunity for the identification of such regulation on a genome-wide scale.

Availability: http://ttalynx.bio.lnu.edu.ua

Contact: ttas@franko.lviv.ua

Supplementary information: Supplementary data are available at Bioinformatics online.

1 INTRODUCTION

The discovery of new regulatory mechanisms, as well as better defined known ones, is an important and challenging field of research for biologists and computer scientists. Codon-based regulation at the translation level is one area in need of further investigation. In the GC-rich genomes of members of the genus Streptomyces, one of the six available leucine codons—TTA—is very rare and contributes to an explicit regulatory mechanism (Chater, 2006; Chater and Chandra, 2006; Leskiw et al., 1991). Its mRNA equivalent (UUA) is recognized by a developmentally important tRNA which is encoded by the bldA gene in streptomycetes (Lawlor et al., 1987; Piret and Chater, 1985). In wild-type streptomycetes, lack expression of this gene during exponential growth leads to the abortive translation of TTA-containing (TTA$^+$) genes thus blocking antibiotic production and the initiation of morphological differentiation (Champness, 1988; Merrick, 1976). Poor aerial mycelium development and impaired secondary metabolite production throughout the cell cycle are two of the most pronounced phenotypic manifestations of the bldA deficiency. Although TTA-mediated regulation is well documented for the model strain Streptomyces coelicolor and some other streptomycetes (Rebets et al., 2006), it remains debatable for other genera of the order Actinomycetales (Ventura et al., 2007).

Functions of many TTA$^+$ genes are either species-specific or unknown. On the other hand, many regulatory genes for secondary metabolite production by Streptomyces are TTA$^+$ (Chandra and Chater, 2008). The mere presence of TTA in an actinomycete gene does not automatically imply its regulatory role, since either mistranslation or the distal position of TTA (with regard to start codon) within an open reading frame can render this codon ‘regulatory-neutral’ (Trepanier et al., 2002). To assess the possibility of TTA-mediated regulation on a per-gene as well as a genome-wide scale, we developed a comprehensive cross-linked screening service for discovering and displaying connections between TTA$^+$ genes within and across genomes. Although our studies are focused mainly on actinobacteria, the possibility of regulation based on rare codons cannot be ruled out for other organisms. In fact, usage of synonymous codons of different rarity in mammalian cell lines has been reported to result in proteins with modified substrate specificity (Kimchi-Sarfaty et al., 2007). Also, interesting skews in codon usage were revealed with our tool in several bacterial genomes (Supplementary Material Figs 8–10). TTA Lynx can be used to mine various genomes and gain initial clue about possible codon-based regulatory mechanisms.

2 SYSTEM ARCHITECTURE AND FEATURES

The service uses Perl programming language for user interaction, BioPerl (Stajich et al., 2002) modules for data processing, BLAST (Altschul et al., 1990) for sequence comparison, COG (Tatusov et al., 2003) and KAAS (Moriya et al., 2007) databases for functional predictions. The service utilizes a Linux/Unix server with Apache/MySQL support run from our department.

Data to be submitted to TTA Lynx should be in FASTA or GenBank (recommended) formats and must contain the predicted boundaries of coding sequences (CDS). Some precomputed replicons can be chosen from the database, such as chromosomes and plasmids of Actinobacteria. The analysis process consists of three main stages: calculating codon-wise statistics of the replicon as a whole; functional prediction of TTA$^+$ coding sequences; and scanning for homologues of the latter in our database. The main results table and codon distribution plots generated by TTA Lynx are displayed below (Fig. 1).

After submission, initial data, including codon usage and distribution, is presented in the form of tables and figures.
TTA codon exclusively in non-essential genes strongly implies its biological system where unexplained intragenic codon skews exist towards the beginning of the CDS. This is proposed to be due to the increased energetic expense of creating longer translated products as well as the potential toxicity of these longer aborted peptides (Fuglsang, 2005). Our service illustrates this skew for any codon in a given replicon in the form of a cumulative diagram with cumulative frequency plotted against relative position in ORFs (Supplementary Material Figs 6–10). This data can be accessed in numeric format as sequence alignments with coordinated TTA codons. (Supplementary Material Figs 1–5). For visualization of sequences and features we use GBrowse viewer (Stein et al., 2002). The bldA-like regulation can be predicted since several factors besides its rarity contribute to its regulatory role. Additionally, presence of TTA codon exclusively in non-essential genes strongly implies its possible regulatory role for actinobacteria. Finally, if the codon is acting in a regulatory role, its relative intragenic position is skewed towards the beginning of the CDS. This is proposed to be due to the increased energetic expense of creating longer translated products as well as the potential toxicity of these longer aborted peptides (Fuglsang, 2005). Our service illustrates this skew for any codon in a given replicon in the form of a cumulative diagram with cumulative frequency plotted against relative position in ORFs (Supplementary Material Figs 6–10). This data can be accessed in numeric format and plotted with other software. At the moment, no simple rule or statistical measure are available to reliably assess the regulatory function of TTA for any gene/genome, mainly because of limited dataset (three Streptomyces genomes). Nevertheless, our service uncovers the possibility of such regulation (Fig. 1B), and, with more data, will help define rules which could be used to study other biological systems where unexplained intragenic codon skews exist (Supplementary Material Figs 8–10).

Functional assessment of TTA\textsuperscript{+} sequences provides an insight into possible regulatory roles of these genes as well as the metabolic pathways involved. Any TTA\textsuperscript{+} CDS is analyzed against the COG database for general functional prediction and then submitted to KAAS for pathway identification. This allows users to facilely identify the specific involvement of TTA\textsuperscript{+} sequences in possible regulatory networks and biosynthetic pathways. The best hit of a COG comparison is considered to be reliable with an e-value <10\textsuperscript{-10}, protein sequence identity above 50% and alignable region covering >50% of a longer sequence (Li et al., 2007). The preferences of KAAS can be found elsewhere (Moriya et al., 2005). Assigned functions and classes are listed in the main feature table (Supplementary Material Fig. 2).

For every coding sequence containing a TTA codon a BLAST search is performed using our embedded database featuring genomes of Actinobacteria collected from individual genome annotations. The best hit from the TTA Lynx database is considered significant when sequence identity is above 50%, the e-value <10\textsuperscript{-10} and alignable region covers >50% of longer sequence. Results are shown as sequence alignments with coordinated TTA codons. Conserved TTA codon can provide direct evidence of evolutionary pressure to be retained across species and, therefore, a present or past regulatory role. Putative orthologs from COG and KAAS predictions are displayed in the main feature table (Supplementary Material Figs 2, 13).

**Funding:** Ministry of Education and Science of Ukraine grant Bg-01F (to V.F.).

**Conflict of Interest:** none declared.

**REFERENCES**


**Fig. 1.** (A) The main results table of the service with 10 replicons listed as an example (seven chromosomes and three plasmids). (B) Differences in distribution of leucine codons in chromosomes of Streptomyces griseus NBRC 13 350 (left) and Bifidobacterium longum DJ010A (right).