

# Genome mining for novel bioactive peptides

Dmitrii Travin  
Center of Life Sciences, Skolkovo  
Institute of Science and Technology,  
Moscow, Russia  
dmitrii.travin@skoltech.ru

Dmitry Bikmetov  
Institute of Molecular Genetics RAS,  
Moscow, Russia  
bikdm12@gmail.com

Konstantin Severinov  
Center of Life Sciences, Skolkovo  
Institute of Science and Technology,  
Moscow, Russia  
k.severinov@skoltech.ru

**Abstract** — Genome mining approach was applied to the search of novel bioactive peptide compounds sharing a functionally important post-translational modification – azol(in)e cycles. It revealed the great diversity of uncharacterized biosynthetic gene clusters, some of which likely encode the biosynthetic pathways of novel compounds inhibiting key cellular functions including translation.

**Keywords** — genome mining, LAPs, antibiotics, translation inhibitors

## Motivation and Aim

The spread of drug resistance to existing antimicrobials among pathogenic bacteria becomes a major threat to modern public healthcare. The urgent need for new antibiotics leads to the development of methods of drug discovery alternative to classical activity-based screenings. Genome mining strategy, which benefits from rapid accumulation of publically available genomic information, is becoming one of the most powerful tools, which has been already successfully applied to the search for novel compounds from diverse groups of microbial natural products.

Earlier we applied BLAST-based search for novel biosynthetic gene clusters (BGCs) of linear azol(in)e modified peptides (LAPs), which allowed the identification and subsequent structural and functional characterization of two compounds inhibiting protein biosynthesis on the bacterial ribosome – klebsazolicin [1] and phazolicin [2]. In this work we aimed to identify novel groups of BGCs of azol(in)e-containing peptide natural products through the comprehensive bioinformatic analysis of all genomic regions containing genes of YcaO-domain containing proteins, which are considered to be a hallmark of azol(in)e modification installation [3].

## Methods

First, we retrieved the sequences of YcaO-containing proteins from RefSeq databases using HMMs for YcaO from Pfam and TIGRFAMs databases. Subsequent filtration allowed to get rid of the closely related sequences from the genomes of overrepresented species. Second, genomic regions around the collected genes of YcaO-containing proteins were annotated and used to build a sequence similarity network of BGCs, which was then analyzed manually.

## Results

The approach described allowed the identification of approximately 15000 unique YcaO-domain containing BGCs in the publically available genomes. The analysis of the sequence similarity network of clusters revealed both BGCs directing the biosynthesis of already known and novel compounds. We performed manual in-depth bioinformatic analysis for three groups of previously uncharacterized clusters, which allowed to predict possible structure of the encoded bioactive peptides as well as shed the light on several intriguing aspects of the evolution of the modified peptide biosynthesis-encoding clusters.

## ACKNOWLEDGMENT

Supported by RSF grant (19-14-00266).

## REFERENCES

- [1] M. Metevlev *et al.*, “Klebsazolicin inhibits 70S ribosome by obstructing the peptide exit tunnel,” *Nat. Chem. Biol.*, vol. 13, no. 10, pp. 1129–1136, Oct. 2017.
- [2] D. Y. Travin *et al.*, “Structure of ribosome-bound azole-modified peptide phazolicin rationalizes its species-specific mode of bacterial translation inhibition,” *Nat. Commun.*, vol. 10, no. 1, p. 4563, 2019.
- [3] B. J. Burkhardt, C. J. Schwalen, G. Mann, J. H. Naismith, and D. A. Mitchell, “YcaO-Dependent Posttranslational Amide Activation: Biosynthesis, Structure, and Function,” *Chem. Rev.*, vol. 117, no. 8, pp. 5389–5456, Apr. 2017.