IN SILICO AND BIOINFORMATICS METHODS OF GENE IDENTIFICATION CARRIED OUT ABROAD

Khusanov Ozodbek Bakhtiyor ugli
Independant researcher

Abstract: In silico methods for linking genomic space to chemical space have played a crucial role in genomics driven discovery of new natural products as well as biosynthesis of altered natural products by engineering of biosynthetic pathways. Here we give an overview of available computational tools and then briefly describe a novel computational framework, namely retro-biosynthetic enumeration of biosynthetic reactions, which can add to the repertoire of computational tools available for connecting natural products to their biosynthetic gene clusters.
Keywords: Secondary metabolite, Polyketides, Nonribosomal peptides, Genes to metabolites, Metabolites to genes.

INTRODUCTION

Bioinformatics has played an important role in in silico identification of new secondary metabolites by genome mining and several pioneering studies have been successful in experimental characterization of new metabolites predicted by in silico analysis. However, majority of the available computational methods for analysis of secondary metabolite biosynthetic pathways utilize forward approach for linking Genes to Metabolites, while automated computational tools for linking secondary metabolites' chemical structures to their biosynthetic gene clusters are not available yet [1].

MATERIALS AND METHODS

In this article, we first give a brief overview of the presently available in silico tools and approaches for analysis of secondary metabolite biosynthetic pathways and identification of novel secondary metabolites by genome mining. Most of the in silico approaches use evolutionary information on sequence/structural features of individual catalytic domains of PKS or NRPS biosynthetic pathways for genome mining of secondary metabolites and for prediction of chemical structures of their putative products. We also discuss the feasibility of devising a retro-biosynthetic approach to link orphan secondary metabolites to their biosynthetic gene cluster. The retro-biosynthetic approach for linking "Metabolites to Genes" involves enumerating the various biochemical transformations or enzymatic reactions which would generate the given secondary metabolite starting from a set of precursor molecules and identifying enzymatic domains which can potentially catalyze the enumerated biochemical transformations [2].

RESULTS AND DISCUSSION

Bioinformatics tools for analysis of secondary metabolite biosynthetic genes have also been developed for analysis of metagenomic data. Metagenomic samples can be quickly scanned for novel natural products by using PCR primers specific for secondary metabolite biosynthetic gene clusters. This PCR-based sequence tag approach has been coupled with in silico phylogenomic tools to search for putative secondary metabolites. eSNaPD has been specifically developed to analyze large metagenomic sequence tag datasets and aid in the discovery of diverse secondary metabolite gene clusters. Another bioinformatics tool which accepts sequence tags from metagenomic datasets along with protein or genomic sequences is NaPDoS. It uses phylogenomic information to search and classify NRPS Adenylation and PKS Ketosynthase domains [3].

The benefits of the approach in reconstruction of pathways have been discussed earlier. This approach is beneficial in cases where the mass spectrometric or similar analysis has revealed the chemical structure of final metabolite but its biosynthetic gene has not been characterized. Retro-biosynthetic tools are available for predicting metabolic routes between two metabolites and predicting biosynthetic routes of plant secondary metabolites. Similar automated in silico tools have been also developed mainly for the prediction of biodegradation pathways [4]. These approaches are reaction rule based, where generalized reactions are applied to final metabolite to enumerate precursor metabolites. Application of all possible generalized reactions at each stage of precursor enumeration can lead to prediction of huge number of possible path- ways – combinatorial explosion [5].

The two major classes of natural products biosynthesized by various microbial, fungal and plant species are polyketides and nonribosomal peptides. Connecting these natural products and their gene clusters would not only broaden the understanding of their complex biosynthesis, but will also help in discovery of novel natural products and help in designing new natural product-based drugs. In silico tools for identification of new secondary metabolites have played an important role in successful experimental characterization of new polyketides and nonribosomal peptides. Most of these computational tools facilitate connecting “genes to metabolite”.

CONCLUSION

The best alignments can be picked and used to predict the probable metabolite synthesized by the biosynthetic cluster. It may be noted that such domain string approach is similar to the clusterBLAST method available in antiSMASH. However, domain string approach will be computationally faster in view of reduced representation of modules in terms of single identifiers. Hence, it can be used for quick comparison of newly identified clusters with experimentally characterized clusters present in various databases.

References:

1. Cane DE, Walsh CT. The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases. Chem Biol 2019;6:R319–25.
2. McKenney JM. Lovastatin: a new cholesterol-lowering agent. Clin Pharm 2018;7:21–36.
3. Cheng YQ, Tang GL, Shen B. Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. Proc Natl Acad Sci U S A 2013;100:3149–54.
4. Walsh CT, Fischbach MA. Natural products version 2.0: connecting genes to molecules. J Am Chem Soc 2010;132:2469–93.
5. Hertweck C. The biosynthetic logic of polyketide diversity. Angew Chem Int Ed Engl 2019;48:4688–716.