Synthetic and Systems Biotechnology 1 (2016) 80–88

Contents lists available at ScienceDirect

Synthetic and Systems Biotechnology

journal homepage: keaipublishing.com/synbio

In silico methods for linking genes and secondary metabolites: The way forward

Shradha Khater, Swadha Anand [a], Debasisa Mohanty *

Bioinformatics Center, National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi 110067, India

ARTICLE INFO

Article history:
Received 20 November 2015
Received in revised form 18 February 2016
Accepted 1 March 2016
Available online

Keywords:
Secondary metabolite
Polyketides
Nonribosomal peptides
Genes to metabolites
Metabolites to genes
Genome mining
Retro-biosynthetic enumeration
Biosynthetic gene cluster

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. Most of the currently available bioinformatics tools for analysis of secondary metabolite biosynthetic gene clusters utilize the "Genes to Metabolites" approach. In contrast to the "Genes to Metabolites" approach, the "Metabolites to Genes" or retro-biosynthetic approach would involve enumerating the various biochemical transformations or enzymatic reactions which would generate the given chemical moiety starting from a set of precursor molecules and identifying enzymatic domains which can potentially catalyze the enumerated biochemical transformations. 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. We also discuss our preliminary work on development of algorithms for retro-biosynthetic enumeration of biochemical transformations to formulate a novel computational method for identifying genes associated with biosynthesis of a given polyketide or nonribosomal peptide.

1. Introduction

Polyketides and nonribosomal peptides are two major classes of secondary metabolite natural products with enormous diversity in chemical structures and bioactivities.[1] Examples of pharmaceutically important polyketides and nonribosomal peptides are lovastatin (a cholesterol-lowering agent),[2] erythromycin (an antibiotic), FK506 (an immunosuppressant) and epothilone (anticancer compound).[3] These secondary metabolites are biosynthesized by multifunctional megasynthases like polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) using a thiotemplate mechanism. The diverse and complex structures of polyketides and nonribosomal peptides arise from assembly line synthesis by these megasynthases. Details of the biosynthetic mechanism have been discussed in a number of earlier reviews.[4–8] Owing to their pharmaceutical and industrial importance, these natural products as well as their biosynthetic mechanisms have been subject of particular interest and extensive characterization.[9] Unraveling the "biosynthetic code" of these natural products has opened up the possibilities for identification of novel natural products in various bacterial and fungal organisms and also biosynthetic engineering of rationally designed secondary metabolites for their use as drug molecules.[10–13] The structural diversity arising from combinatorial complexity of their biosynthesis is the reason why these natural products are a great source of drugs. Understanding the mechanisms of their biosynthesis and devising clever strategies to tweak it can potentially yield fruitful results in the form of economically important products.[14] The extent of diversity of these natural products has been vastly underestimated and with new niches of microorganisms being explored, the number of novel bioactive metabolites is likely to increase many folds.[15,16] It has been anticipated that novel drugs can be discovered by cultivating and characterizing microorganisms like actinobacteria.[17] Therefore, these bacterial strains could be the new unexplored sources of natural products. In addition, the exponential growth of genome sequencing has unveiled many bacteria containing putative natural product biosynthetic gene clusters with unknown biosynthetic products.[18,19]

* Corresponding author. Bioinformatics Center, National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi 110067, India. Tel.: +91-11-26703749; fax: +91-11-26742125.
E-mail address: deb@nii.res.in, deb@nii.ac.in (D. Mohanty).
[a]Current Address: Bio-Sciences R&D Division, TCS Innovation Labs, Tata Consultancy Services Limited, Pune, India.
Peer review under responsibility of KeAi Communications Co., Ltd.

http://dx.doi.org/10.1016/j.synbio.2016.03.001

Linking biosynthetic genes to secondary metabolites and *vice versa* can potentially help not only in characterization of new secondary metabolites, but also in redesigning known biosynthetic pathways of secondary metabolites to produce novel compounds.[4,20] The problem can in principle be solved using two approaches: Forward (Genes to Metabolites) and Reverse/Retro-biosynthetic (Metabolites to Genes) Approach[21,22] (Fig. 1). In forward approach genomic sequence information is used to predict the chemical structure of the final metabolite. In contrast to forward approach which starts by considering the genes or gene clusters and attempts to predict its biosynthetic product, retro-biosynthetic approach starts from a known metabolite and attempts to identify which gene cluster might be biosynthesizing it.[23,24] Even though traditionally identification of natural products and their biosynthesis have been an area of interest for microbiologists, organic chemists and biochemists, elucidation of the catalytic machinery for biosynthesis of polyketides and nonribosomal peptides by genome encoded PKS and NRPS clusters has opened up the area of genomics driven discovery of new natural products' biosynthetic pathways.[13,25,26] Bioinformatics has played an important role in *in silico* identification of new secondary

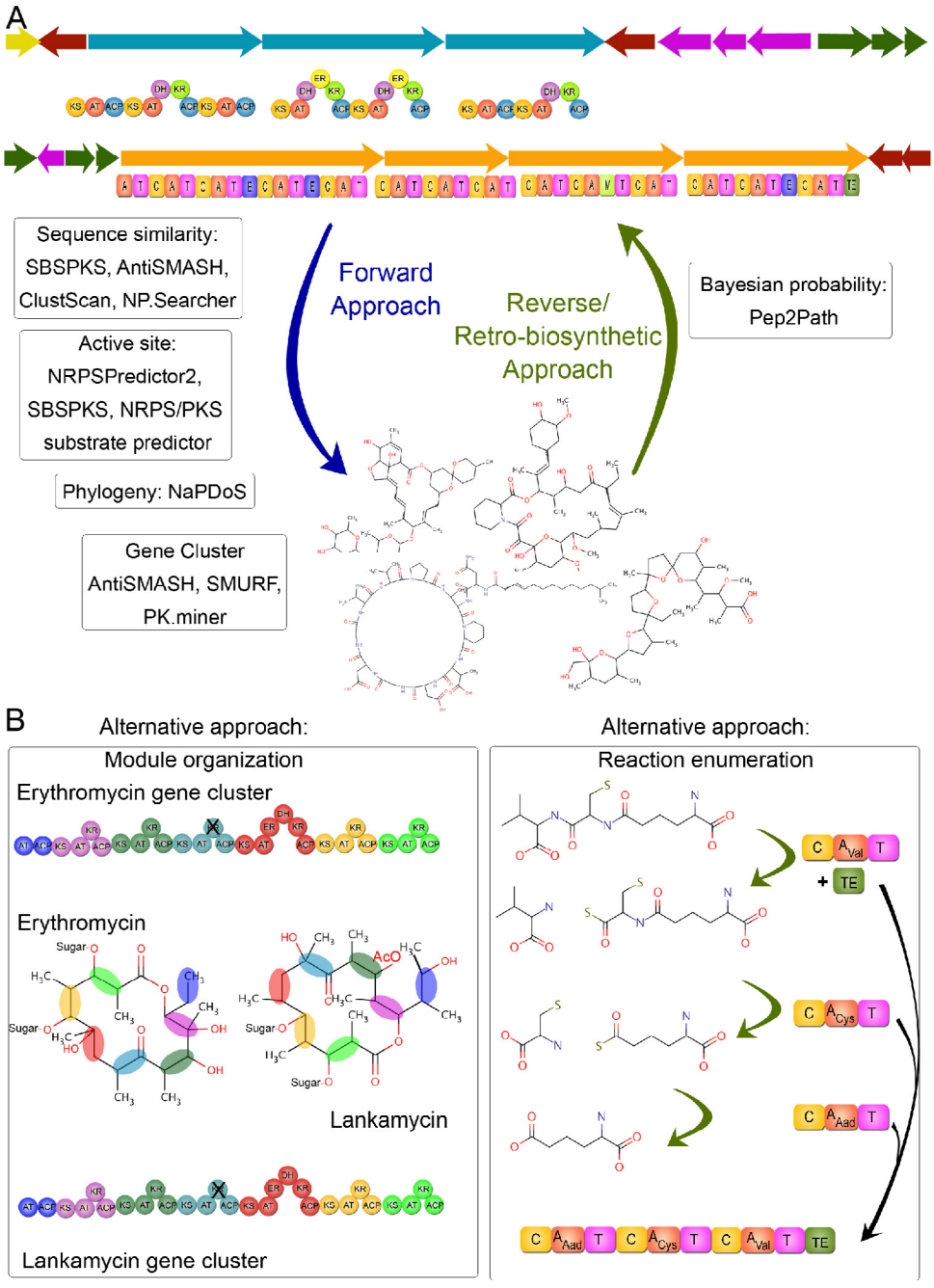

Fig. 1. Two approaches for deciphering new biosynthetic pathways. (A) "Forward approach", where information from genes is used to decipher the biological pathways. "Retro-biosynthetic approach" is where a known product is linked to the genes. Some of the available methods belonging to either approach have been mentioned in boxes. (B) Alternative approaches to connecting genes and metabolites. (Left Panel) Use of module organization in comparison of secondary metabolite gene clusters and prediction of the secondary metabolite synthesized. (Right Panel) Retro-biosynthetic approach for prediction of the gene cluster responsible for biosynthesis of a particular secondary metabolite.