Use of bioinformatics to discover natural products

Álvaro Serrano Morrás, Grau en Biologia 2017

Introduction

The increasing demand for new drugs has become one of the major problems in health science nowadays. Since the discovery of penicillin in 1928, and after the great expansion of natural products discovery, the new drugs discovery ratio has diminished steadily every year. The efforts made in the synthetic design of drugs are not able to fulfill healthcare needs, so it is time to go back to researching bioactive compounds in nature. Recent advances in genome sequencing and increasing computational power open the avenue to bioinformatic approaches to speed up discovery and uncover the hidden natural products in bacteria, plants, and fungus. Additionally, metagenomics and the discovery of new patterns in gene clusters will unlock bioinformatics full potential in the near future (Milshteyn et al., 2015).

Objectives

- Define the methodologies used for the discovery of biosynthetic genes in bacteria, plants and fungi and establish their importance for the discovery of natural products
- Compare the applications of the computational methodologies, highlighting the various limitations and specific characteristics of different taxa.

The four approaches

Different theoretical approaches have been tried, each one exploiting distinct characteristic properties of the secondary metabolite genes. (Medema and Osbourn, 2016) In order to differentiate them in a comprehensible manner, I will classify the approaches in four groups: clustering (A), co-expression (B), epigenetics (C) and phylogeny (D) in order of relevance in the various tools (Fig. 1).

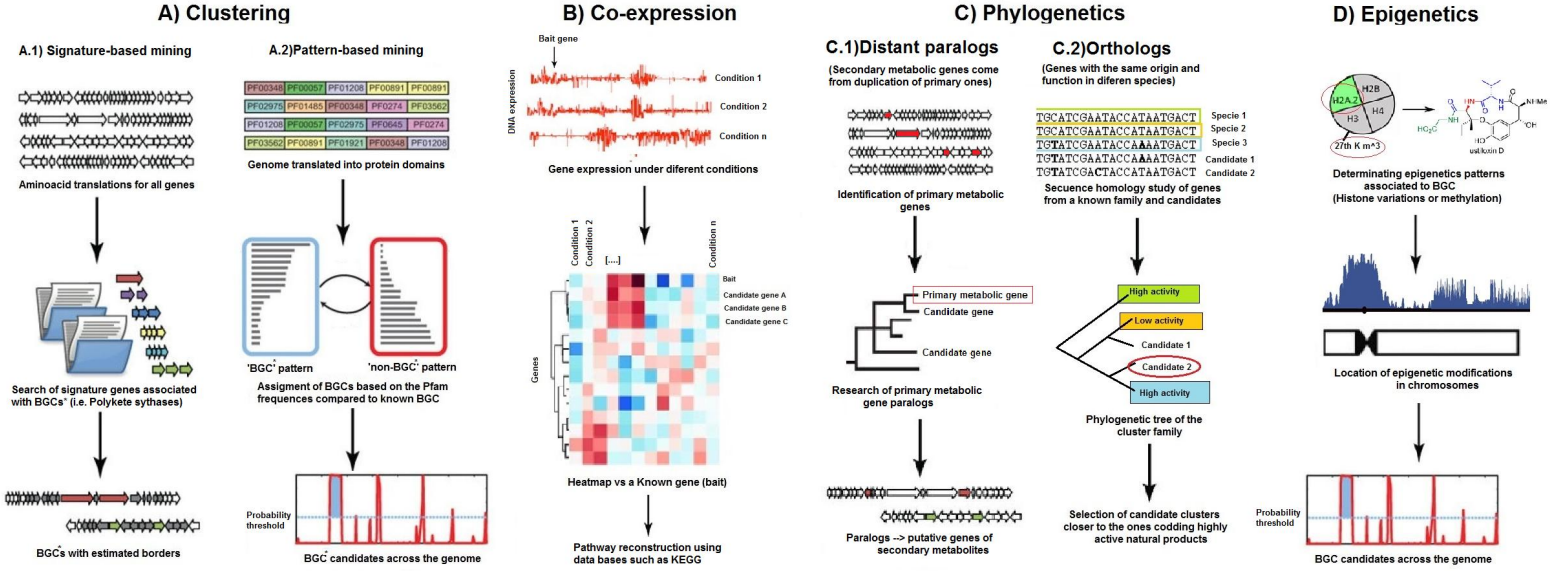

Figure 1. The four bioinformatic approaches Adapted from Medema and Fischbach, 2015 and Medema and Osbourn, 2016. * BGC refers to Biosynthetic Gene cluster

Aplications and limitations

The four approaches listed are not mutually exclusive and many tools are based on two or more of them. Also the implementation depends on the kind of organism its designed for.

· **Bacteria** represent the main source of natural products both from phenotypical screening and bioinformatic discovery (Newman and Cragg, 2016). Thus, most of the tools are designed for them. Their simple genome, with pathway genes are clustered, and their easy manipulation makes very accessible the usage of the clustering and co-expression approaches. Moreover, metagenomics have allowed the computational research on the uncultured bacteria, which are estimated to hold the 99% of the diversity.

· **Fungi** do not fall behind thanks to the discovery of a highly productive taxa in terms of natural products: the endophytic fungi. Also, fungi also have many of their biosynthetic genes clustered and have benefited very much with the implementation of pattern based cluster mining.

· **Plants** have a very complex genomic structure that hinder bioinformatic natural product discovery. New discoveries such as biosynthetic gene clusters in plants and the appearance of epigenetic imprints associated with natural product genes have improved the expectations and opportunities in plant natural product discovery.

Conclusions and future perspectives

These computational approaches establish the foundations to regain the discovery ratio achieved in the mid 1900's. Bacteria are leading the field, thanks to the metagenomics revolution and their compatibility with the most established approaches such as the clustering and co-expression. On the other hand, recent achievements and discoveries in plants and fungi have unlocked much of their potential, and they will soon rival bacteria as the main source for natural product discovery.

References

Medema M. H., Fischbach M. A. (2015). *Nature Chemical Biology, 11,9,, 639–648.*
Medema M. H., Osbourn A. (2016). *Natural Product Reports, 33,8, 951–962..*
Milshteyn A., Schneider J. S., Brady S. F. (2014) *Chemistry & Biology, 21,9, 1211–1223.*
Newman D.J., Cragg G.M. (2016). *Journal of Natural Products, 79, 629–661.*