Computationally driven discovery and engineering of biosynthetic pathways

Marnix H. Medema, Microbial Genomics and Bioinformatics Research Group, Max Planck Institute for Marine Microbiology, Bremen, Germany

Plants and microorganisms are a rich source of bioactive compounds. In recent years, computational methods have become more and more important to identify these molecules. With the antiSMASH software, we constructed a key computational pipeline that identifies biosynthetic genomic loci covering the whole range of known secondary metabolite compound classes [1,2].

More recently, we combined antiSMASH with a new generic algorithm that probabilistically identifies both known and unknown types of biosynthetic gene clusters [3]. This allowed us to perform a global quantitative and comparative analysis of biosynthetic gene clusters throughout the microbial tree of life, which lead to the uncovering of a major new family of natural products, the aryl polyenes. In order to rapidly connect identified gene cluster families to experimentally observed molecules, we have developed a software package (Pep2Path) to automatically link peptide tandem MS-generated peak patterns to the gene clusters that encode the biosynthesis of the molecules they represent [4].

These and other developments in the field are rapidly increasing the speed with which novel gene clusters are being discovered. To fully exploit the riches of information that are being generated from this, it will be essential that contextual data on these gene clusters is stored in a consistent fashion. Therefore, we are launching a community data standard, the Minimum Information about a Biosynthetic Gene cluster (MIBiG), which allows effective integration of chemical, genomic and ecological data. This will help guide experimental researchers to the most promising targets and expedite advances in high-throughput synthetic biology methodologies for natural product discovery [5,6].

[1] Medema MH et al. (2011) Nucl. Acids Res. 39: W339-W346.
[2] Blin K, Medema MH et al. (2013) Nucl. Acids Res. 41: W204-W212.
[3] Cimermancic P, Medema MH, Claesen J et al. (2014) Cell 158: 412-21.
[4] Medema MH et al. (2014) PLoS Comp. Biol. 10: e1003822.
[5] Medema MH et al. (2011) Nature Rev. Microbiol. 9: 131-137.
[6] Medema MH et al. (2012) Nature Rev. Microbiol. 10: 191-202.