A NETWORK APPROACH TO EFFICIENT DRUG DISCOVERY

SNORRE SULHEIM[1, 2, *], KRISTIN DEGNES[2], GIANG-SON NGUYEN[2], ALEXANDER WENTZEL[2] AND EIVIND ALMAAS[1]

[1] NTNU - Norwegian University of Science and Technology, [2] SINTEF Industry, Department of Biotechnology and Nanomedicine, [*] snorre.sulheim@sintef.no

CENTRE FOR DIGITAL LIFE NORWAY

The Research Council of Norway

ABSTRACT

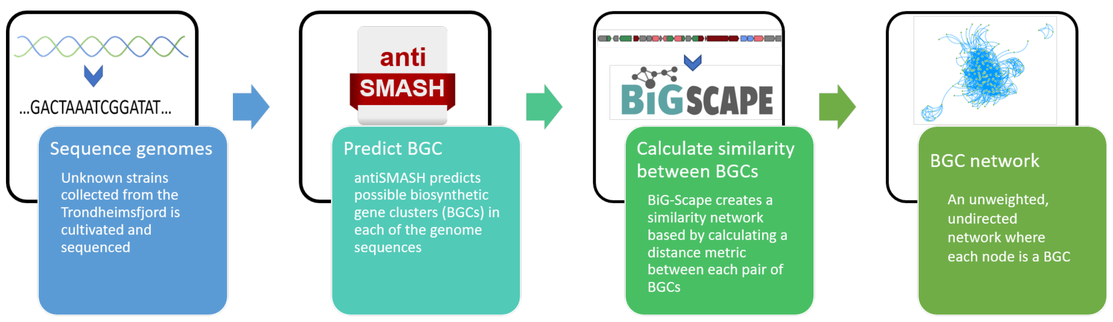
The increasing resistance of pathogenic bacteria to existing antibiotics is currently one of the biggest threats to global health. The occurrence of bacteria becoming resistance to antibiotics is a well-known and unavoidable process, but the rate is drastically increased by the global misuse of antibiotics. The second cause is the lack of new antibiotics. Bioactive compounds produced by microbes is the main source of antibiotics currently in use, but the traditional methods for exploiting this natural resource are no longer efficient. The genes encoding production of secondary metabolites may not be expressed when cultivated in the lab or the amount is too small to have a bioactive effect.

We are developing a workflow for identifying new, bioactive chemicals produced by microorganisms and we have access to a large strain collection of marine actinomycetes. We use mass spectrometry (MS1), tandem mass spectrometry (MS/MS) and UV absorption to characterize the products from the cultivated organisms. The products are clustered using GNPS (Global Natural Products Social Molecular Networking) [1] to create a network based on spectral alignment of the MS/MS data.

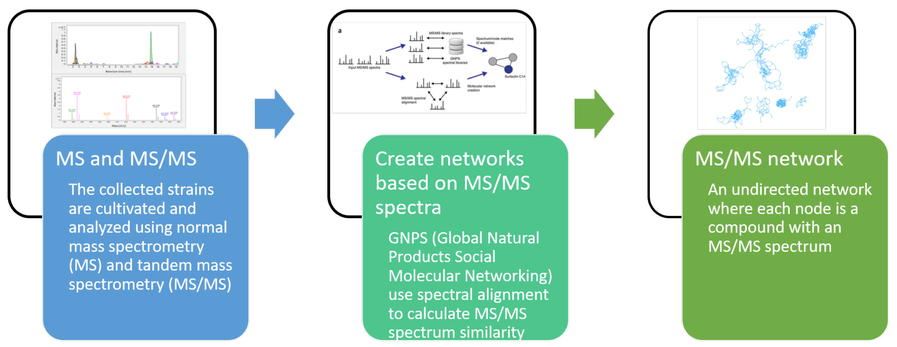
The same organisms are sequenced and we use antiSMASH [2] to predict the presence of biosynthetic gene clusters in their genomes. Biosynthetic gene clusters (BGCs) are genes spatially clustered on the genome which encodes a secondary metabolic pathway, and antiSMASH [2] also predict the most likely structure(SMILES [3]) for the product of this metabolic pathway. We use BiG-SCAPE [4] to calculated a distance metric between each pair of biosynthetic gene clusters to create the network.

We have combined these two networks by introducing an intermediate network with the SMILES, obtained either from the MS/MS data or from the antiSMASH predictions. By doing this we have created a structured, novel method for displaying and analyzing this complex data set. Further research is necessary to fully exploit this three-layered network.

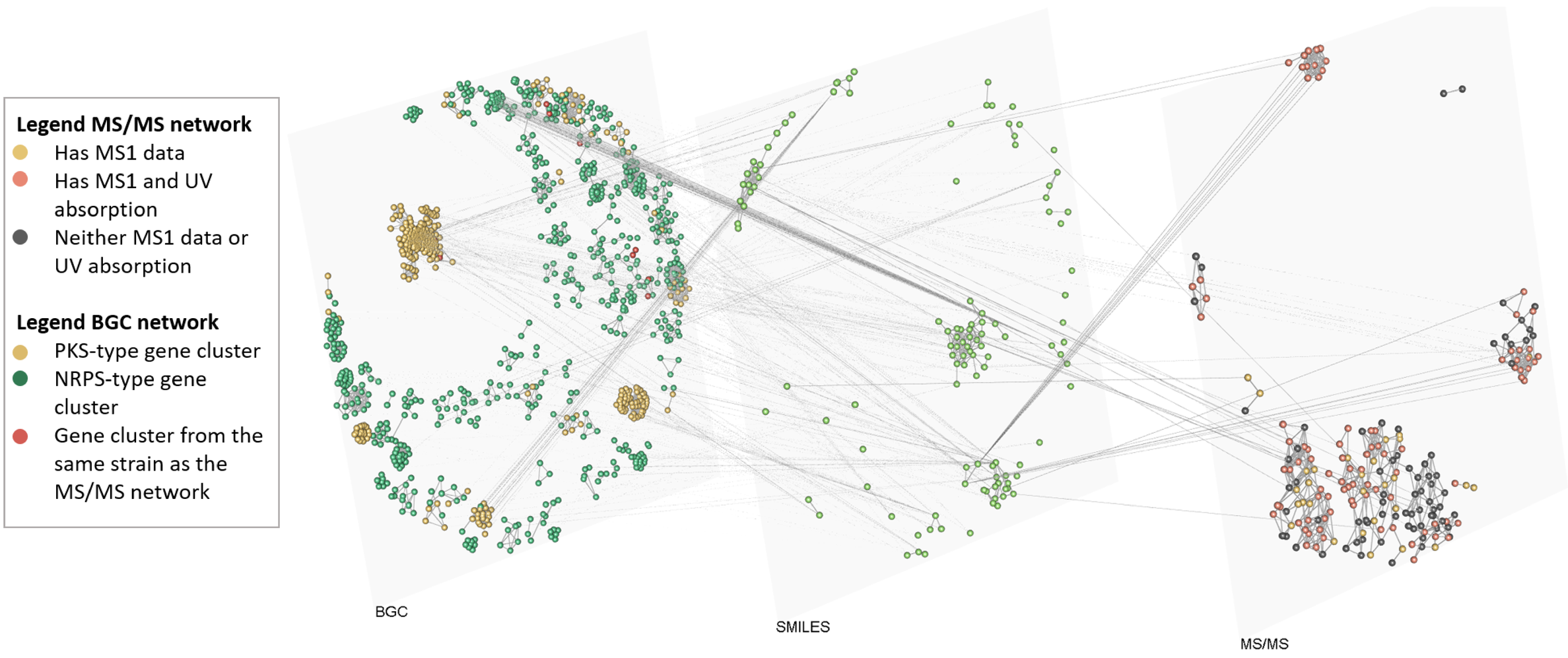
BGC NETWORK WORKFLOW



MS/MS NETWORK WORKFLOW



COMBINED NETWORK



This three-layer network display the MS/MS network of one bacterial strain combined with a network of biosynthetic gene clusters (BGCs) from a collection of 200 bacterial strains. All the strains have been collected from the marine environment in the vicinity of Trondheim. These two networks are connected by an intermediate layer of metabolite structures (SMILES). The intra-layer edges of the SMILES network is created by comparing fingerprints of the metabolite structures. The fingerprints are obtained from PubChem [5] and the network created by calculating the pairwise Tanimoto similarity [6] and 0.75 as the threshold.

ACKNOWLEDGEMENTS

This work is a part of the INBioPharm-project, funded by the Norwegian Research Council through the Centre for Digital Life Norway.

REFERENCES

1. Wang, M. *et al.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature biotechnology* **34,** 828 (2016).
2. Weber, T. *et al.* antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic acids research* **43,** W237–W243 (2015).
3. Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of chemical information and computer sciences* **28,** 31–36 (1988).
4. Kautsar, S. *et al. BiG-SCAPE: Biosynthetic Genes Similarity Clustering and Prospecting Engine* `https://git.wageningenur.nl/medema-group/BiG-SCAPE`. Accessed: 2018-05-15.
5. Kim, S. *et al.* PubChem substance and compound databases. *Nucleic acids research* **44,** D1202–D1213 (2015).
6. Tanimoto, T. T. Elementary mathematical theory of classification and prediction (1958).