A novel roadmap through bacterial genomes leads the way to new drug discovery

October 6 2014, by Claudia Lutz

For millennia, bacteria and other microbes have engaged in intense battles of chemical warfare, attempting to edge each other out of comfortable ecological niches. Doctors fight pathogens with an arsenal of weapons—antibiotics—co-opted from these microbial wars, but their efforts are frustrated by the development of drug resistance that outpaces drug discovery. Researchers at the University of Illinois at Urbana-Champaign and Northwestern University have now innovated and demonstrated the value of an algorithm to analyze microbial genomic data and speed discovery of new therapeutic drugs.

A large proportion of the medications used today were discovered by screening bacteria and other organisms for their ability to produce natural products, biologically useful compounds. In recent years, pharmaceutical companies have largely abandoned this strategy in favor of screening synthetically created chemicals for useful properties, an area of research which has yielded a tiny number of new antibiotics.

Microbiologist and molecular and cellular biologist Bill Metcalf, a leading investigator in the new study, described the reason for pharmacological research's shift away from the exploration of natural products. "There was a reason why they gave up . . . they kept discovering the same things over and over and over again," he said. "They were getting very diminishing returns."

This type of problem will be familiar to anyone who has ever collected trading cards. It's easy to acquire a set of the common cards, but it can

be nearly impossible to find the rare ones scattered among them. A collector might wish that he or she could sneak a peek at all the cards hidden inside the wrappers, and only buy the new ones.

Genome sequence information, which is now available for an ever-increasing number of bacterial species, holds the promise to allow antibiotic hunters to do just that. Groups, or "clusters" of genes within each genome code for enzymes, proteins that work together to synthesize a natural product for that bacterium. Part of the vision of the Institute for Genomic Biology's Mining Microbial Genomes research group, led by Metcalf, is to use bacterial genome sequence data as an index of what products each one can produce.

If researchers could infer what type of product the bacterium is making by looking at its DNA, they wouldn't have to go through a lengthy screening process—they could just scan genomes for promising gene clusters. Unfortunately, this task is much harder than it sounds. Many clusters have some sequences or whole genes in common, making them indistinguishable by traditional comparative methods even though they enable the production of different compounds.

Metcalf, co-lead author and Institute for Genomic Biology Fellow James Doroghazi, and colleagues cleared this hurdle with a clever computational solution: they combined multiple comparative metrics, each with a carefully calibrated weight, to produce an algorithm that sorted 11,422 gene clusters from 830 bacterial genomes into an orderly, searchable reference. Their work was published this month in *Nature Chemical Biology*.

In the database created in the study, gene clusters predicted to make very similar products are linked with each other in networks referred to as families. These predictions mesh almost perfectly with prior knowledge; gene clusters that produce similar known compounds were sorted by the

new algorithm into the same family in every case but one.

The value of the new database was showcased by an experiment performed in collaboration with a group of chemists at Northwestern University, led by former University of Illinois faculty member Neil Kelleher. Kelleher's group used a high-precision analytical technique to infer the chemical composition of unknown compounds isolated from 178 different bacterial strains. The research team was then able to assign a biological function to gene cluster families by correlating their presence in the genome with the production of particular compounds.

In addition to this power to link gene cluster families to potential new antibiotics, the database is a huge step toward solving the "trading card" problem. By comparing the distribution of gene cluster families across bacterial species, researchers can now predict which species are most likely to contain novel antibiotics, and target the richest strains for study. "We've got the framework, we know the number of gene clusters, we know who has them and therefore we know where to look to find new drugs," said Metcalf. "It clearly leads to discovery."

 More information: *Nature Chemical Biology* paper - DOI: 10.1038/nchembio.1659

The database of gene cluster families is publically available: bit.ly/1rLR0O6.

Provided by University of Illinois at Urbana-Champaign