

IS2-1

Analyzing Microbes in Us and on Our Planet

Peer Bork

EMBL, Meyerhofstr. 1, 69117 Heidelberg, Germany

Environmental sequencing, that is metagenomics, has become a major driver for uncovering microbial biodiversity and increasingly also for cataloging molecular microbial functions in humans and on our planet. I will introduce in powers and limitations of metagenomics using examples from our work on the human gut, ocean and soil. In the human gut, a major line of focus is metagenome-wide associations studies (MWAS), which do not reveal causalities, but an increasing granularity in disease associations and thus can be a starting point for diagnostics, as I will illustrate using colorectal cancer. MWAS results lack directionality though and are often indirect or confounded, the latter we could demonstrate for type 2 diabetes, where the intake of the drug metformin rather than the disease itself leads to the association with the gut microbiome. This calls for complementary *in vitro* studies and indeed, in a screen of 1200 marketed drugs against each of 40 representative human gut bacteria we could show, in the context of a multi-group collaboration, that a quarter of all non-antibiotic drugs directly inhibit at least one gut microbial strain. In another screen, we could identify also many reciprocal microbe-drug interactions and show that human gut bacteria do not only biotransform, but also bioaccumulate therapeutic drugs. As all of our gut microbes are coming implicitly from the environment and our planet is one big ecosystem, it is crucial to study biodiversity and microbial communities at planetary scale. The feasibility of such a global approach I illustrate by (i) the TARA oceans project, surveying the microbial diversity of this vast ecosystem by studying plankton in all major ocean regions and (ii) topsoil metagenomics, revealing, for example, a global war between fungi and bacteria as well as regional antibiotic resistance gene reservoirs. To enable such data-driven findings, a considerable bioinformatics infrastructure is needed and tools as well as resources that are maintained and can quickly adapt to the vastly increasing amounts of data. I will illustrate how this can be achieved using a data science network setting.