### Plasmid Detection and Assembly in Genomic and Metagenomic Datasets

Plasmids are extrachromosomal and independently replicating DNA molecules that provide their bacterial hosts with additional genetic material important for survival and adaptation. Before the sequencing era, plasmids were detected based on various phenotypic changes they provide, such as antibiotic resistance or ability to degrade recalcitrant organic compounds. However, sequencing efforts revealed many cryptic plasmids that do not contribute to the phenotype of the host cell in an obvious way. Although there are about 10,000 plasmids listed in the RefSeq database (Pruitt et al, 2006), many plasmids remain undetected since it is not trivial to assemble plasmids from genomic and metagenomic datasets (Antipov et al., 2016, Rozov et al., 2017). We thus conjecture that many classes of plasmids remain unknown, like many previously unknown classes of viruses that were found in recent studies (Paez-Espino et al, 2016, Roux et al., 2016).

Since plasmids exchange genetic material with the host chromosomes and vary in structure (circular or linear), size (from a thousand to millions of nucleotides), and gene content, it is not clear how to computationally define the concept of a plasmid in such a way that it would be possible to distinguish them from the chromosomes in draft assemblies. Also, plasmid assembly is complicated by various repeats that are difficult to resolve using short reads sequencing technologies.

Here we present a metaplasmidSPAdes algorithm that improves on existing tools(Antipov et al., 2016, Rozov et al., 2017) by (i) iteratively extracting subgraphs with gradually increasing read coverage from the metagenome assembly graph, (ii) finding putative plasmids as simple cycles in these subgraphs, and (iii) verifying the found putative plasmids using a new plasmidVerify tool. We applied plasmidSPAdes+ (plasmidSPAdes complemented by plasmidVerify) and metaplasmidSPAdes to diverse genomic and metagenomic samples and revealed 1000s of plasmids missed in previous studies, including many plasmids that share no significant similarities with known plasmids, and plasmids carrying antibiotic-resistance genes.