MetaPore-C: using chromatin conformation capture and PCR-free long reads for metagenomic analysis

Cross-linking can be used to fix all of the DNA in a cell, enabling more straightforward assembly of genomes from metagenomic mixtures and allowing plasmids to be associated with their host genomes

Contact: **publications@nanoporetech.com** More information at: **www.nanoporetech.com** and **publications.nanoporetech.com**

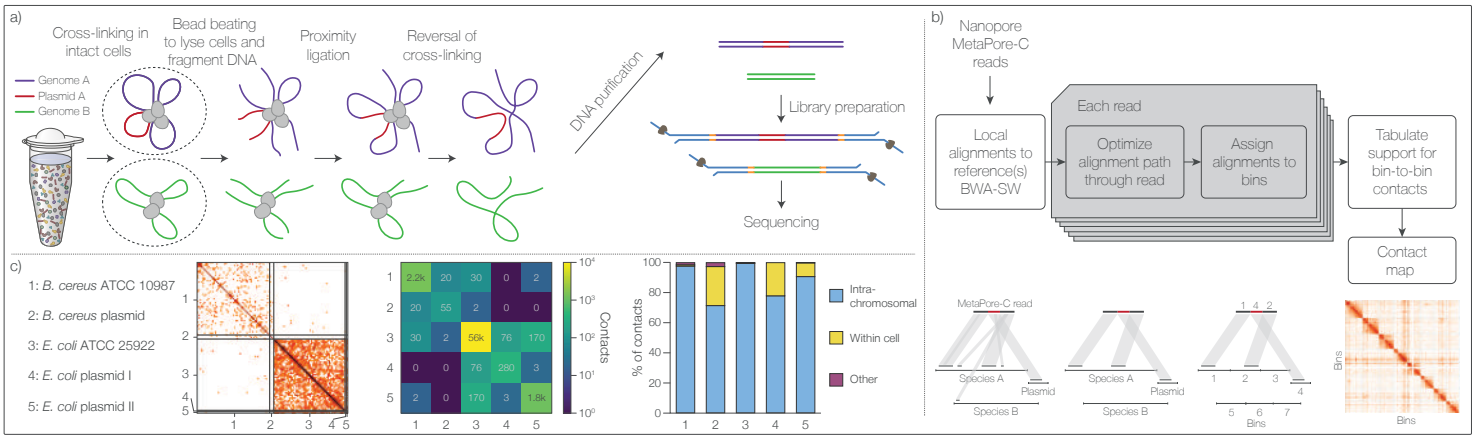

Fig. 1 MetaPore-C a) laboratory workflow b) overview of bioinformatics workflow c) correct association of plasmids with host genomes for a mixture of two bacterial species

Generating contact information from metagenomic mixtures, using chromatin conformation capture without restriction digestion or PCR, combined with nanopore sequencing

Chromatin conformation capture is a method used to investigate interactions between genomic loci that are not adjacent in the primary sequence. When applied to metagenomic mixtures, as well as simplifying assembly the technique provides a way to associate plasmids with their host genomes. It is typical to use restriction digestion to fragment cross-linked DNA before proximity ligation, but this step is influenced by the base composition of the genomes present, which is not always known in advance. Our protocol avoids digestion by using bead beating to simultaneously lyse the cells and to fragment the DNA (Fig. 1a). To produce contact information, MetaPore-C reads are first aligned to a collection of chromosomal and extra-chromosomal reference sequences using *BWA-SW*. Aligned reads are filtered to retain the minimal collection of alignments that traverse the majority of the read. The reference genomes are then divided into equally sized bins and each aligned segment of the MetaPore-C read is assigned a bin. Finally, the total number of bin-to-bin contacts is calculated from all reads and visualised in a contact map. Extra-chromosomal elements can be assigned to their host by determining which chromosome(s) share the most contacts with the element (Fig. 1b). This approach allowed us to assign plasmids to the correct host in a mixture of two bacteria (Fig. 1c).

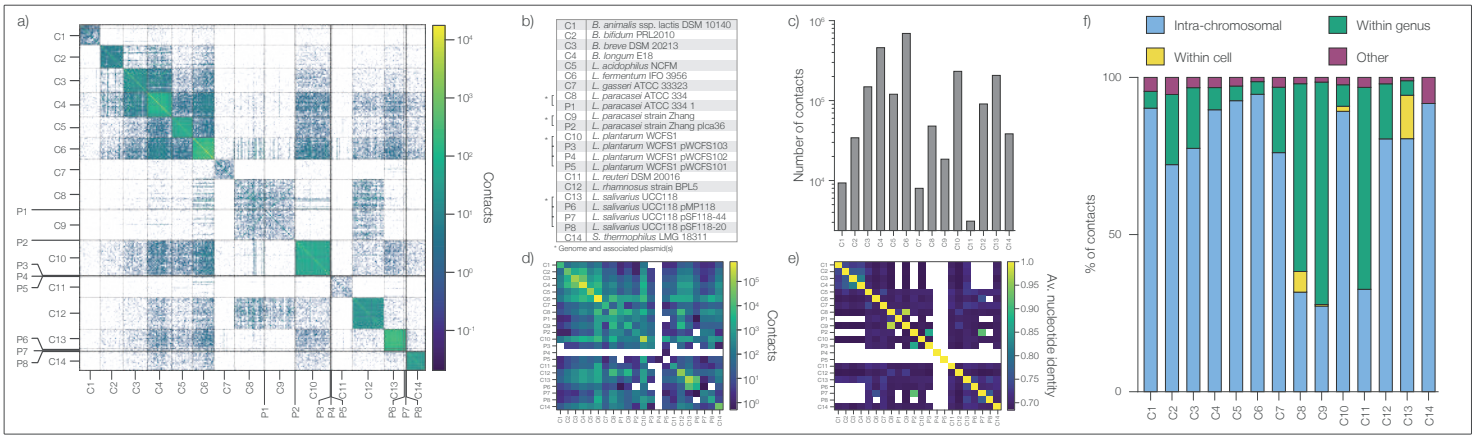

Fig. 2 MetaPore-C of a probiotic sample a) contact map b) bacteria in the sample and their associated plasmids c) number of contacts per genome d) scaled map e) background signal f) associations

Demonstrating MetaPore-C on a probiotic food supplement, to associate plasmids with their host genomes by identifying intra- and extra-chromosomal contacts

We generated MetaPore-C sequence data from a probiotic food supplement, without restriction digestion or PCR amplification, using the laboratory workflow outlined in Fig. 1. We then used the bioinformatics workflow to create a contact map for the bacterial chromosomes and plasmids within the sample (Fig. 2a). The probiotic supplement contained 15 different bacterial species and strains, many of which were closely related (Fig. 2b). The number of contacts identified per genome ranged from approximately 3,000 to 700,000 (Fig. 2c), reflecting the differences in relative abundance of each bacterial strain in the sample. Because plasmids are small compared to bacterial genomes, it is not possible to see the associations clearly from the contact map, so we plotted a version of the map using the total number of contacts between pairs of molecules, and in which all genomes and plasmids were given equal spacing. This allows contacts to be seen more clearly (Fig. 2d). The plot of average nucleotide identity (Fig. 2e) indicates a low level of spurious interaction between species, which is likely to be an artefact arising from mapping ambiguities. Future work will involve the use of more fine-grained approaches to reduce the background signal. In spite of this, we were able to associate plasmids correctly to the expected host genomes (Fig. 2f) and in addition we identified a high density of intra-chromosomal interactions, which are valuable for binning during metagenomic assembly.