



RESEARCH NOTE

# Metagenomic Chromosome Conformation Capture (3C): techniques, applications, and challenges [version 1; referees: 2 approved]

Michael Liu, Aaron Darling

ithree institute, University of Technology Sydney, Sydney, NSW 2007, Australia

**v1** First published: 30 Nov 2015, 4:1377 (doi: [10.12688/f1000research.7281.1](https://doi.org/10.12688/f1000research.7281.1))  
 Latest published: 30 Nov 2015, 4:1377 (doi: [10.12688/f1000research.7281.1](https://doi.org/10.12688/f1000research.7281.1))

**Abstract**

We review currently available technologies for deconvoluting metagenomic data into individual genomes that represent populations, strains, or genotypes present in the community. An evaluation of chromosome conformation capture (3C) and related techniques in the context of metagenomics is presented, using mock microbial communities as a reference. We provide the first independent reproduction of the metagenomic 3C technique described last year, propose some simple improvements to that protocol, and compare the quality of the data with that provided by the more complex Hi-C protocol.

**Open Peer Review**

Referee Status:

	Invited Referees	
	1	2
version 1 published 30 Nov 2015	 report	 report
1	Mick Watson, University of Edinburgh UK	
2	Titus Brown, University of California Davis USA	

**Discuss this article**

Comments (0)

**Corresponding author:** Aaron Darling ([Aaron.Darling@uts.edu.au](mailto:Aaron.Darling@uts.edu.au))

**How to cite this article:** Liu M and Darling A. **Metagenomic Chromosome Conformation Capture (3C): techniques, applications, and challenges [version 1; referees: 2 approved]** *F1000Research* 2015, 4:1377 (doi: [10.12688/f1000research.7281.1](https://doi.org/10.12688/f1000research.7281.1))

**Copyright:** © 2015 Liu M and Darling A. This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Grant information:** The author(s) declared that no grants were involved in supporting this work.

**Competing interests:** The authors declared no competing interests.

**First published:** 30 Nov 2015, 4:1377 (doi: [10.12688/f1000research.7281.1](https://doi.org/10.12688/f1000research.7281.1))

## Introduction

Metagenomics has been proposed as a means to characterize the microbial communities that are pervasive in our environment (Handelsman, 2004). Current metagenomic protocols, however, fail to capture critical information on the organisation of genetic material in microbial communities, as the fine-scale structure of the community and linkage among DNA sequences is intentionally destroyed by cell lysis and DNA shearing steps prior to sequencing. Computational methods of sequence binning attempt to assign sequences to the species or strains that were present in the sample, thereby inferring the linkage information destroyed by sample processing, but these methods have limited resolution despite many years of development (Lindgreen *et al.*, 2015; Peabody *et al.*, 2015).

Chromosome conformation capture (3C) and related approaches offer an alternative strategy that allows the spatial organization of genetic material in a microbial community to be preserved and measured, either via high throughput sequencing or other assays. In 3C, the fine-scale structure of the sample is preserved via reversible crosslinking, typically by soaking the sample in formaldehyde immediately after collection (Dekker *et al.*, 2002). The sample is then subjected to cell lysis and further steps are applied to interrogate the spatial structure in the sample.

Published protocols for coupling 3C with metagenomics involve restriction digestion, followed by a *proximity ligation*, followed by crosslink reversal, DNA collection, optional enrichment for ligation junctions, and sequencing library preparation (Beitel *et al.*, 2014; Burton *et al.*, 2014; Marbouty *et al.*, 2014). The proximity ligation is a key step wherein a DNA ligation reaction is carried out under highly dilute conditions. The low concentration of sample material favors ligation events among DNA strands which are crosslinked together in the same molecular complex. Crucially, this allows separate DNA macromolecules, e.g. a chromosome and a plasmid, or two chromosomes that were co-bound in a protein complex, to become ligated to each other (Beitel *et al.*, 2014). These ligation junctions can then be identified via high throughput sequencing. The rate at which such ligation events are observed in the data is highly correlated with the frequency at which the DNA was in close physical contact at the time of sample crosslinking (Lieberman-Aiden *et al.*, 2009).

Several other methods can support direct measurement or inference of linkage among metagenomic DNA sequences. We describe these below. Metagenomic 3C has several advantages relative to these other methods, along with some disadvantages.

### Single cell sequencing

Single cell sequencing methods can capture data on a relatively large fraction of the genetic material in a cell (10–80% depending on the whole genome amplification conditions). However single cell techniques are vulnerable to reagent and equipment contamination and depend on cells being readily separable, making them difficult to deploy widely. Moreover, single cell techniques gather data on only a small fraction of the cells in a sample rather than the entire population.

### Long read single molecule sequencing

The Pacific Biosciences and Oxford Nanopore platforms implement sequencing technologies that can read DNA strands up to 100 kilobases (Laver *et al.*, 2015) and possibly more. Long sequence reads capture more information about the arrangement of genes into chromosomes than is available in short (<1000nt) reads typical of other sequencing technologies. Single molecule sequence reads currently have accuracy ranging from 80–90%, which is sufficient for detecting genes but offers only limited ability to identify single nucleotide variants and indels (Quick *et al.*, 2014). Consensus signal approaches such as Circular Consensus Sequencing can help to overcome the error in single molecule sequencing but do so at the expense of read length or throughput (Larsen *et al.*, 2014). These methods read single molecules and therefore they are unable to identify relationships between plasmids and host chromosomes without being coupled to a library preparation method like 3C or Hi-C.

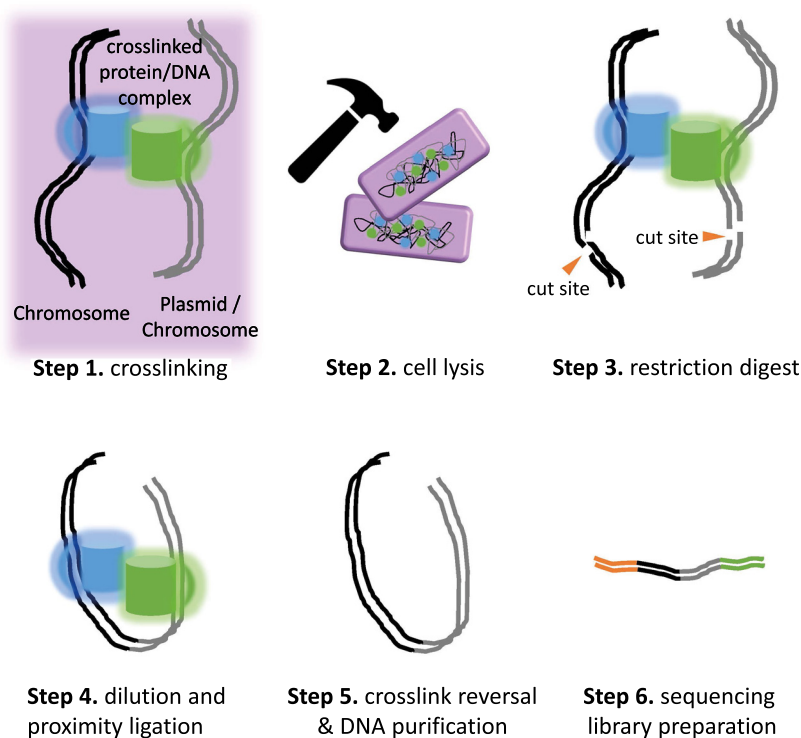
### Correlated coverage binning

This strategy leverages the observation that genetic material present in the same species or strain changes in abundance over time & space in a highly correlated manner. By generating metagenomic data on an environment across multiple time points, sampling sites, or even different cell lysis treatments, it becomes possible to reconstruct linkage information by identifying sequences whose abundances are highly correlated across samples (Albertsen *et al.*, 2013; Alneberg *et al.*, 2014; Imelfort *et al.*, 2014). The power to detect such associations grows with the number of samples and the extent of change across samples (Alneberg *et al.*, 2014). This approach has the advantage of being relatively simple to implement, only requiring the additional effort to collect and process a larger number of samples. A potential drawback is that in recombining populations, the abundance of a particular gene, plasmid, or polymorphism may not correlate strongly with one particular host species' abundance, leading to a failure to correctly identify the linkage relationship. Plasmids and bacteriophage may have copy number dynamics that are independent of host chromosomes, potentially making some associations difficult to detect. Finally, this approach does not provide direct information to order & orient assembly contigs into genome-scale scaffolds, however the inferred linkage information could in principle be used to eliminate ambiguity in assembly graphs and so yield more contiguous assemblies.

### Metagenomic 3C

Metagenomic 3C has thus far been implemented in two protocols. **Text box 1** gives an overview of these protocols and **Table 1** highlights the main differences in the quality of data generated by each protocol. The Hi-C approach was the first to be described in the context of metagenomics (Beitel *et al.*, 2014; Burton *et al.*, 2014), and involves steps that enrich the sample for proximity ligations. The basic metagenomic 3C approach has the advantage of being simpler to execute in the laboratory (Marbouty *et al.*, 2014).

We have succeeded in implementing and extending the protocol first described by Marbouty *et al.*, 2014 on a mock community to facilitate a detailed comparison of metagenomic 3C and Hi-C. Our



**Text box 1. 3C and proximity ligation methods.** Chromosome conformation capture (3C) was first developed as a means to determine the average three dimensional chromosome structure in a population of cells, for a single species (Dekker *et al.*, 2002). This general approach was later coupled with high throughput DNA sequencing (Lieberman-Aiden *et al.*, 2009), providing a means to generate detailed 3D structure models of chromosomes. Many extensions of the 3C technique have been developed (Dekker *et al.*, 2013).

The basic 3C protocol involves an initial step of reversible crosslinking, typically via formaldehyde at 1–3%. This step crosslinks proteins to each other and to DNA. The formaldehyde is then quenched and the cells are lysed either enzymatically or via physical disruption. Next, a restriction digestion is carried out using a 4- or 6-cutter that leaves a single-stranded overhang. Subsequently the sample is placed in a large volume DNA ligase reaction; yielding conditions that strongly favor the ligation of free ends that are co-bound in a protein complex. This step is referred to as *proximity ligation*. After proximity ligation, the crosslinks are reversed via heat incubation and the DNA is purified via proteinase K & RNase digestion and EtOH precipitation. Finally, the purified DNA is ready for standard high throughput sequencing library preparation, for example via adapter ligation and enrichment PCR.

Hi-C extends the protocol described above by incorporating steps that enrich the final sequencing library for proximity ligation events. In Hi-C, the single stranded overhangs left after the restriction digest are filled with biotinylated nucleotides. The proximity ligation which follows is thus a blunt-end ligation and the junctions contain biotinylated nucleotides. Biotinylated nucleotides must be removed from any remaining unligated free ends. In the final steps of sequencing library preparation, fragments containing the biotinylated ligation junctions can be captured on streptavidin-coated magnetic beads, yielding a library substantially enriched for proximity ligations (Lieberman-Aiden *et al.*, 2009).

extension of the original protocol adds a bead purification step following crosslink reversal and replaces the shearing & adapter ligation for sequencer library preparation with a tagmentation reaction. This in turn reduces input material requirements by several fold, enabling the reactions to be scaled down and reducing reagent cost. The details of the extended protocol and accession numbers for the associated data sets can be found in the [Supplementary material](#).

Several challenges emerge in applying 3C protocols to microbial communities. Samples often consist of heterogeneous cell types. The thick walls of some cells may affect the extent of crosslinking, causing some cells to crosslink more extensively than others.

High formalin concentrations lead to reduced DNA recovery in later stages of the protocol. Data from experiments using a range of formalin concentrations on the same sample suggest that concentrations between 2 and 3% provide an optimal trade-off between proximity ligation rate in gram positive cells and DNA yield (see [Supplementary material](#)). However, these data reflect only a small number of species relative to the currently described microbial diversity.

Microbial communities can consist of organisms with a wide range of genomic G+C composition, and this must be considered when selecting a restriction enzyme to use in 3C and related protocols.

**Table 1. Comparison of 3C and Hi-C for metagenomics.**

	3C	Hi-C
Proximity ligation read rate	Up to 6.5%	4% (Beitel <i>et al.</i> , 2014) or 12–51% (Burton <i>et al.</i> , 2014)
Resolution limit	1–2kbp	1–2kbp (4 cutter) or 15–30kbp (6 cutter)
Marked ligation junctions	No	Yes
Difficulty of library prep	hard	very hard
Erroneous association rate	<1%	<1%
Requires separate metagenomic library	No	Yes

Table 1. Differences in the features of metagenomic 3C and Hi-C are listed. The proximity ligation read rate indicates the fraction of all reads that contain proximity ligation events. For Hi-C the rate varies widely in published data. The resolution limit is dictated by the density of restriction cut sites in the chromosome, which are typically more dense when using a 4-cutter (3C or Hi-C), than with a 6-cutter (Hi-C only). Marked ligation junctions are created as a by-product of the end-filling in Hi-C and can be identified as a tandem duplication of the overhang sequence in the data. The erroneous association rate is defined as the fraction of read pairs found to associate two different species or strains in mock community experiments.

Data on synthetic communities shows that density of restriction sites is directly proportional to the rate of observed proximity ligation events in metagenomic 3C data. For example, a library created using the enzyme HpaII (recognition site C<sup>^</sup>CGG) yields very few reads with proximity ligation junctions for *S. aureus* (32% G+C) but for *P. aeruginosa* (67% G+C) up to 6.5% of reads contain proximity ligation junctions. Therefore it may be advantageous to process samples in parallel with two or more enzymes having diverse recognition sites.

## Applications of metagenomic 3C

### Reconstructing genomes from metagenomes

The data produced by metagenomic 3C or Hi-C can be used to address a range of questions in microbial community analysis. Chief among these is reconstruction of the so-called population genomes of each species present in a microbial community. A population genome does not reflect the genome of an individual cell in the community, but rather is a consensus genome sequence describing the genetic material present in a collection of closely related cells, e.g. a population or species. The population genome may represent an amalgamation of many closely related strains each with their own strain-specific gene content and mutations. The extent of such microdiversity among strains has a strong influence on the ability of current sequence assembly algorithms to reconstruct a metagenomic assembly. Once recovered, the population genomes can support a range of downstream analysis such as metabolic network reconstruction for individual community members. Predicted metabolic networks can in turn be used to inform analysis of species interactions and help guide strategies for identifying and cultivating microbes of interest (Imelfort *et al.*, 2014; Parks *et al.*, 2015).

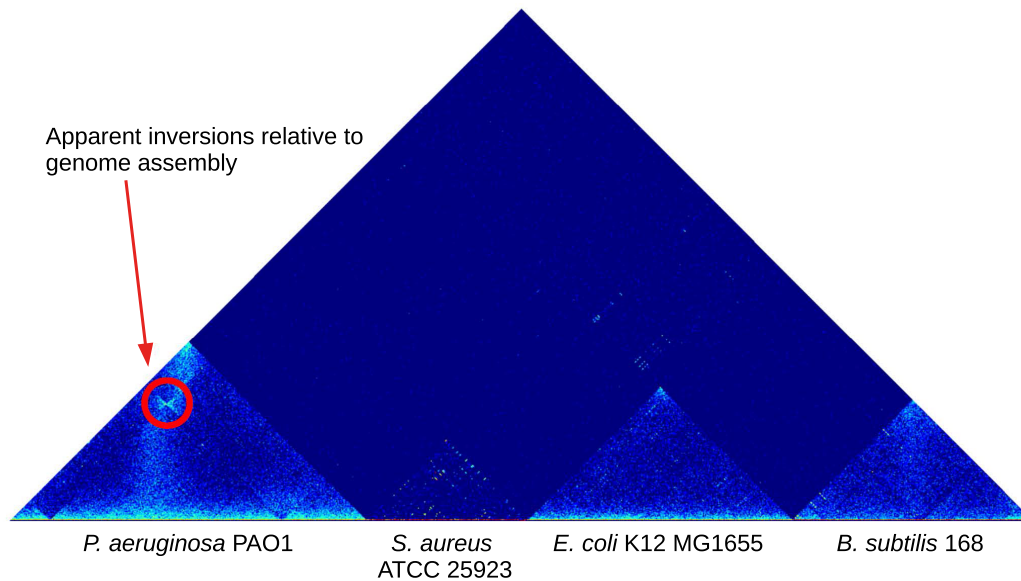
Current approaches for reconstructing population genomes are relatively simplistic and involve a first step of mapping the 3C read pairs to the metagenomic assembly, counting the number of links found among each contig in the read pair data, and then using a clustering algorithm to group contigs by population/species. Several clustering algorithms have been explored for this task.

Beitel *et al.*, 2014 applied Markov clustering and found that use of a low inflation parameter in the algorithm led to clusters that accurately reflect population genomes. Marbouty *et al.*, 2014 used Louvain clustering and were able to achieve similarly accurate results on simple test communities. Both of these algorithms have the advantage that prior knowledge of the number of population genomes is not required. Burton *et al.*, 2014 applied a custom algorithm that requires the number of population genomes in the sample to be known *a priori*. This requirement is likely to pose a difficulty in cases where independent lines of evidence are unable to yield a reliable estimate of the number of population genomes in a sample.

In addition to its use in reconstructing population genome content, metagenomic 3C can in principle be used to guide the scaffolding of metagenomic assembly contigs. Hi-C data has already been demonstrated to facilitate chromosome-scale scaffolding of large eukaryotic genomes (Burton *et al.*, 2013; Marie-Nelly *et al.*, 2014). When scaffolding microbial genomes, the much greater resolution afforded by 4-cutters (as used in the basic metagenomic 3C protocol) is likely to be essential for accurately ordering & orienting contigs in population genomes. The signal available for scaffolding can be visualized using the contact map concept, as shown in Figure 1. When the contigs are correctly ordered and oriented the majority of contacts occur locally, obeying a distance-decay relationship dictated by polymer physics (Marie-Nelly *et al.*, 2014). Figure 1 highlights an exception to this, where the strain used in the laboratory has undergone rearrangement relative to the finished reference genome.

### Tracking plasmids, bacteriophage, and mobile DNA

Metagenomic 3C offers the exciting possibility to quantify the frequency of association between mobile DNA such as plasmids and host chromosomes. In the simplest scenario, such data could be used in a purely descriptive capacity, to document the relationships between plasmids and hosts in various microbial ecosystems. Another possibility would be to characterise how the relationships between host chromosomes and plasmids change over time



**Figure 1. 3C/Hi-C heatmap.** Contact map of chromatin interactions identified by metagenomic 3C. A synthetic community of four bacterial isolates was subjected to metagenomic 3C and the resulting read data mapped back to reference chromosome assemblies. Heat intensity is proportional to the number of read pairs associating the two chromosome regions. In *P. aeruginosa* and *B. subtilis*, the two arms of the circular chromosome are colocalized, as reflected in the column of intense heat emanating from the middle of their chromosomes. Erroneous cross-species associations are seen to be rare (deep blue field).

in response to external stimuli, for example antibiotic exposure. 3C-based protocols that employ 4-cutter enzymes are likely to be essential for such applications, since the use of a 4-cutter increases the likelihood that suitable cut sites will exist in small plasmids.

In principle a similar strategy could be applied to characterise relationships between host chromosomes and bacteriophage or other types of mobile DNA. Previous work in mouse models has suggested that bacteriophage in the mouse gut selectively transduce antibiotic resistance genes and broaden their host range in response to antibiotic treatment (Modi *et al.*, 2013). Application of metagenomic 3C techniques in this context remains unexplored, although current protocols and computational techniques are adequate to support such applications.

### Future directions

Metagenomic 3C provides information on the spatial organisation of genetic material in microbial communities. This type of information is valuable and highly complementary to data generated by other strategies and technologies. In particular, the ability to link separate DNA polymers which are localized in the same cell creates opportunities for study that would be intractable with classic shotgun sequencing strategies, whether using long reads or not.

Several barriers currently prevent ready application of metagenomic 3C and related methods to microbial communities. Naturally occurring microbial communities can harbour a milieu of live and

dead cells, along with free DNA and protein. At the time of this writing, no application of the technique has yet been reported for a natural environmental sample. Marbouty *et al.*, 2014 described an application to a sample sourced from Seine river sediments, however, that sample was subjected to an enrichment culture prior to formalin fixation. The enrichment culture presumably created a population of intact cells and reduced the prevalence of free DNA in the sample.

Classic 3C and Hi-C protocols require large amounts of sample material, but microbial communities of interest can be of very limited biomass, for example subgingival dental plaques or individual soil particles. Improving the efficiency of the metagenomic 3C protocol will be essential before it can be applied to such sample types. Several possible avenues exist to improve the reaction efficiency, elements of which have already been described in the context of single-cell Hi-C experiments on mammalian cells (Nagano *et al.*, 2013).

A further major barrier to analysis of metagenomic 3C data is the presence of strain-level microdiversity in a sample. The existence of even just two strains with genomes around 98% average nucleotide identity is sufficient to cause extensive fragmentation in genome assemblies, depending on the assembly algorithm. The resulting assembly contigs can be too small to harbor restriction sites and therefore will fail to cluster into population genomes. In principle, advanced computational methods which operate directly

on genome assembly string graphs (Myers, 2005) instead of their contig-based representations could solve this problem. However, such computational tools do not currently exist for metagenomic 3C data analysis. It is worth noting that this problem also impacts the use of other strategies for generating population genomes such as correlated coverage binning.

Hi-C data has been demonstrated to facilitate phasing human chromosomes (Selvaraj *et al.*, 2013), and Beitel *et al.*, 2014 showed that metagenomic Hi-C data had characteristics that would support resolution of the genotypes of two *E. coli* strains in a synthetic mixture. Much work remains before 3C or Hi-C could actually be applied to strain resolution, however. The number of genotypes present in a microbial community is unknown *a priori*, and the degree of divergence among genotypes is also unknown but has a major influence on the technique's resolving power. Substantial investment will be required to develop tools for statistical inference on the genotypes present in samples characterized by metagenomic 3C sequencing. The fact that the number of genotypes and their divergences are unknown *a priori* will add significant complexity to the algorithms. It is likely the case that reconstructing the genotypes of individual cells in the sample will remain impossible, but inference algorithms may instead compute a probability distribution over cellular genotypes. Such a probability distribution could support testing & rejection of specific hypotheses, for example whether gene A and B are subject to an epistatic interaction, or whether population X is significantly more diverse than population Y. In the extreme case where strain genotypes are separated by just two variant sites in distant chromosomal locations, a very large

amount of 3C data would be required to generate enough read pairs covering the two sites to estimate their frequency of linkage. This is due to nature of 3C data, and reflects the fact that distantly located sites rarely interact in the cell in most cases (Beitel *et al.*, 2014; Marie-Nelly *et al.*, 2014). This represents a fundamental limitation of metagenomic 3C and highlights a need for complementary strategies such as the single cell or correlated coverage techniques.

---

#### Author contributions

Michael Liu carried out the experiments, generated the data, and wrote material for the manuscript. Aaron Darling analysed the data and wrote material for the manuscript. All authors have seen and agreed to the final content of the manuscript.

#### Competing interests

The authors declared no competing interests.

#### Grant information

The author(s) declared that no grants were involved in supporting this work.

#### Acknowledgements

We thank Christopher W. Beitel for assistance with outlining the topics discussed and constructive comments on a draft manuscript, Carly Rosewarne for suggestions on implementing the metagenomic 3C protocol, and Matthew Z DeMaere for providing a script to generate heatmaps from mapped read pairs.

## Supplementary materials

### Supplementary Figure S1

Raw, unnormalized rate of proximity ligation products in metagenomic 3C libraries, as a function of formalin concentration. A synthetic microbial community was subjected to metagenomic 3C library prep & sequencing at a range of formalin concentrations, and the fraction of read pairs mapping at distances >1000nt was taken as an estimate of the proximity ligation read rate.

### Supplementary information

Protocol for the construction of metagenomic 3C (Meta3C) libraries, Illumina sequencing, analysis of metagenomic 3C data and analysis of the Burton *et al.*, Hi-C 2014 data.

[Click here to access the data.](#)

## References

- Albertsen M, Hugenholtz P, Skarshewski A, *et al.*: **Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes.** *Nat Biotechnol.* 2013; **31**(6): 533–38.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Arneberg J, Bjarnason BS, de Bruijn I, *et al.*: **Binning metagenomic contigs by coverage and composition.** *Nat Methods.* 2014; **11**(11): 1144–46.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Beitel CW, Froenicke L, Lang JM, *et al.*: **Strain- and plasmid-level deconvolution of a synthetic metagenome by sequencing proximity ligation products.** *PeerJ.* 2014; **2**: e415.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Burton JN, Adey A, Patwardhan RP, *et al.*: **Chromosome-scale scaffolding of *de novo* genome assemblies based on chromatin interactions.** *Nat Biotechnol.* 2013; **31**(12): 1119–25.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Burton JN, Liachko I, Dunham MJ, *et al.*: **Species-level deconvolution of metagenome assemblies with Hi-C-based contact probability maps.** *G3 (Bethesda).* 2014; **4**(7): 1339–46.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Dekker J, Marti-Renom MA, Mirny LA: **Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data.** *Nat Rev Genet.* 2013; **14**(6): 390–403.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Dekker J, Rippe K, Dekker M, *et al.*: **Capturing chromosome conformation.** *Science.* 2002; **295**(5558): 1306–11.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Handelsman J: **Metagenomics: application of genomics to uncultured microorganisms.** *Microbiol Mol Biol Rev.* 2004; **68**(4): 669–85.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Imelfort M, Parks D, Woodcroft BJ, *et al.*: **GroopM: an automated tool for the recovery of population genomes from related metagenomes.** *PeerJ.* 2014; **2**: e603.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Larsen PA, Heilman AM, Yoder AD: **The utility of PacBio circular consensus sequencing for characterizing complex gene families in non-model organisms.** *BMC Genomics.* 2014; **15**(1): 720.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Laver T, Harrison J, O'Neill PA, *et al.*: **Assessing the performance of the Oxford Nanopore Technologies MinION.** *Biomolecular Detection and Quantification.* 2015; **3**: 1–8.  
[Publisher Full Text](#)
- Lieberman-Aiden E, van Berkum NL, Williams L, *et al.*: **Comprehensive mapping of long-range interactions reveals folding principles of the human genome.** *Science.* 2009; **326**(5950): 289–93.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lindgreen S, Adair KL, Gardner P: **An evaluation of the accuracy and speed of metagenome analysis tools.** *bioRxiv.* 2015.  
[Publisher Full Text](#)
- Marbouty M, Cournac A, Flot JF, *et al.*: **Metagenomic chromosome conformation capture (meta3C) unveils the diversity of chromosome organization in microorganisms.** *Elife.* 2014; **3**: e03318.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Marie-Nelly H, Marbouty M, Cournac A, *et al.*: **High-quality genome (re)assembly using chromosomal contact data.** *Nat Commun.* 2014; **5**: 5695.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Modi SR, Lee HH, Spina CS, *et al.*: **Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome.** *Nature.* 2013; **499**(7457): 219–22.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Myers EW: **The fragment assembly string graph.** *Bioinformatics.* 2005; **21**(Suppl 2): ii79–85.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Nagano T, Lubling Y, Stevens TJ, *et al.*: **Single-cell Hi-C reveals cell-to-cell variability in chromosome structure.** *Nature.* 2013; **502**(7469): 59–64.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Parks DH, Imelfort M, Skennerton CT, *et al.*: **CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes.** *Genome Res.* 2015; **25**(7): 1043–1055.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Peabody MA, Van Rossum T, Lo R, *et al.*: **Evaluation of shotgun metagenomics sequence classification methods using *in silico* and *in vitro* simulated communities.** *BMC Bioinformatics.* 2015; **16**: 363.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Quick J, Quinlan AR, Loman NJ: **A reference bacterial genome dataset generated on the MinION™ portable single-molecule nanopore sequencer.** *Gigascience.* 2014; **3**: 22.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Selvaraj S, R Dixon J, Bansal V, *et al.*: **Whole-genome haplotype reconstruction using proximity-ligation and shotgun sequencing.** *Nat Biotechnol.* 2013; **31**(12): 1111–18.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

# Open Peer Review

Current Referee Status:



## Version 1

Referee Report 15 March 2016

doi:10.5256/f1000research.7846.r11377



### Titus Brown

Department of Population Health and Reproduction, University of California Davis, Davis, CA, USA

This is a thorough discussion of using 3C for metagenome reconstruction. I was a little surprised by the elaborate review-style article together with some experimental data but I think it works well.

I agree with the points made by the other reviewer (Mick Watson). It would be nice to see more complex metagenomes tackled!

One final suggestion - the reference to graph genomes is poor and rather insufficient - there's been a lot of recent work in this area. I refer you to Dilthey *et al*, 2015, [Improved genome inference in the MHC using a population reference graph](#) for one reference.

### References

1. Dilthey A, Cox C, Iqbal Z, Nelson MR, McVean G: Improved genome inference in the MHC using a population reference graph. *Nat Genet.* 2015; **47** (6): 682-8 [PubMed Abstract](#) | [Publisher Full Text](#)

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

**Competing Interests:** No competing interests were disclosed.

Referee Report 25 January 2016

doi:10.5256/f1000research.7846.r11375



### Mick Watson

The Roslin Institute, University of Edinburgh, Edinburgh, UK

Liu and Darling present an excellent review of the application of 3C capture techniques to metagenomic data analysis and I recommend indexing.

There are two issues I would have liked to have seen discussed/presented more:



1. Often metagenomic samples undergo bead bashing in order to disrupt the gram+ cell walls, and this often results in highly fragmented DNA. It would be interesting to hear the thoughts of the authors on how this might affect the results from 3C capture techniques.
2. Figure 1 shows a beautiful reconstruction of 4 genomes. However, many real environmental samples contain 1000+ genomes. The authors discuss this in great detail, and the problems that ensue; however, synthetic metagenomes exist consisting of more than 4 but less than 1000, and I wonder why nobody has applied 3C techniques to those synthetic communities?

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

***Competing Interests:*** No competing interests were disclosed.

---