Effects of Sample Treatments on Genome Recovery via Single-Cell Genomics

Scott Clingenpeel[1], Patrick Schwientek[1], Philip Hugenholtz[2] and Tanja Woyke[1]*

[1]DOE Joint Genome Institute/LBNL, Walnut Creek, California, USA
[2]Australian Centre for Ecogenomics, University of Queensland, Brisbane, Australia

*Corresponding contributor: Tanja Woyke, DOE Joint Genome Institute, 2800 Mitchell Dr., Walnut Creek, CA 94598, 925-296-5839, twoyke@lbl.gov

30 April 2014

ACKNOWLEDGMENTS:

The work conducted by the US Department of Energy Joint Genome Institute is supported by the Office of Science of the US Department of Energy under Contract No. DE-AC02-05CH11231.

DISCLAIMER:

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United

Effects of Sample Treatments on Genome Recovery via Single-Cell Genomics

States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.

Effects of Sample Treatments on Genome Recovery via Single-Cell Genomics

Running title: Sample treatment effects on single-cell genomes

Scott Clingenpeel[1], Patrick Schwientek[1], Philip Hugenholtz[2] and Tanja Woyke[1]*

[1]DOE Joint Genome Institute, Walnut Creek, California, USA; [2]Australian Centre for Ecogenomics, University of Queensland, Brisbane, Australia

*Corresponding contributor: Tanja Woyke, DOE Joint Genome Institute, 2800 Mitchell Dr., Walnut Creek, CA 94598, 925-296-5839, twoyke@lbl.gov

ABSTRACT

Single-cell genomics is a powerful tool for accessing genetic information from uncultivated microorganisms. Methods of handling samples before single-cell genomic amplification may affect the quality of the genomes obtained. Using three bacterial strains we show that compared to cryopreservation, lower quality single-cell genomes are recovered when the sample is preserved in ethanol or if the sample undergoes fluorescence *in situ* hybridization, while sample preservation in paraformaldehyde renders it completely unsuitable for sequencing.

Keywords: single-cell genomics, sample preservation, FISH

Relatively little is known about the functioning of complex microbial communities largely due to the difficulty in culturing most microbes (Rappe & Giovannoni, 2003). Although metagenomics can provide information on the genetic capabilities of the entire community, it is difficult to connect predicted gene functions to specific organisms using metagenomics (Morales & Holben, 2011). One method to address these difficulties is single-cell genomics where a single microbial cell is isolated from a sample, lysed, and its genome amplified by multiple displacement amplification (MDA; Lasken, 2012, Blainey, 2013).

When working with environmental or clinical samples it is generally impractical to work with fresh samples. Most samples need to be preserved in some way before they are studied by single-cell genomics. A possible difficulty is that preservation methods could induce damage to the DNA that would negatively impact genome recovery. The effect of treatments on the cells is especially important in single-cell genomics because it is already established that the process produces incomplete genomes even with fresh cells (Marcy *et al.*, 2007). The average estimated genome completeness of 650 single-

33 cell genomes that are publically available in IMG (https://img.jgi.doe.gov) is 41%, as based on the
34 presence of conserved single copy genes by the method in Rinke *et al.* (2013).

35 The purpose of this study was to test the effects of common sample treatments on the recovery
36 of genomes from single cells. Three methods of sample preservation were tested: cryopreservation with
37 20% glycerol as a cryoprotectant, preservation in 70% ethanol, and preservation in 4%
38 paraformaldehyde. Because paraformaldehyde treatment causes crosslinks between nucleic acids and
39 proteins, we tested cells exposed to 4% paraformaldehyde with and without having their crosslinks
40 reversed by heat treatment. Additionally, there is demand for obtaining single-cell genomes from
41 particular groups within a microbial community. A common approach for this type of isolation makes
42 use of fluorescence *in situ* hybridization (FISH) to identify the target cells (Podar *et al.*, 2007; Haroon *et*
43 *al.*, 2013). Thus, cells that had undergone a typical FISH protocol without prior fixation (Yilmaz *et al.*,
44 2010) were included as an additional treatment to separate the FISH protocol from the fixation steps.

45 In order to determine if the GC content of a genome would affect the results, three bacterial
46 strains were chosen with differing genomic G+C contents: *Pedobacter heparinus* DSM 2366, 42% GC;
47 *Escherichia coli* K12-MG1655, 51% GC; and *Meiothermus ruber* DSM 1279, 63% GC. These have
48 complete genome sequences available and have the same cell structure (Gram negative) to control for
49 major differences in lysis efficiency. Forty cells of each of the three strains underwent each of the five
50 treatments outlined above (Figure 1).

51 All cells underwent an alkaline lysis and MDA (Woyke *et al.*, 2011). The time at which the
52 inflection point of the real-time amplification curves occur (Cp value) correlates with the completeness
53 of the recovered single-cell genomes (Figure S1). Cp order was ranked as follows: cryopreservation <
54 FISH < ethanol < paraformaldehyde with crosslinks reversed < paraformaldehyde with crosslinks intact =
55 negative controls (Figure 1). Although the real-time kinetics showed some amplification, in no cases did
56 the MDA products from paraformaldehyde with or without crosslink reversal treatments produce a 16S
57 rRNA gene amplicon from PCR (Figure 1); indicating that the MDA amplicons were likely due to
58 nonspecific amplification of primers. For each strain, eight cells from each of the three treatments that
59 produced a correct 16S rRNA gene sequence (cryopreservation, ethanol, and FISH) were selected for
60 shotgun sequencing. The cells chosen were ones that had a correct 16S rRNA gene sequence and had
61 the earliest Cp times in the treatment.

62 To eliminate any biases due to varying sequencing depth, the sequences were randomly
63 subsampled to a coverage depth of 315x for each single amplified genome (SAG; Figure S2). The reads
64 were mapped to the reference genomes, a *de novo* assembly was performed and the assembled data
65 was also mapped to the reference genomes.

66 Cryopreservation resulted in SAGs with the highest percentage of the genome recovered for all
67 three strains of bacteria (Figures 2, S3). FISH treatment produced reduced genome coverage, and
68 ethanol preservation resulted in the lowest amount of the genome recovered. These treatments are
69 significantly different from each other (Figures 2, S3). Despite the trend for cells with a higher GC

content to have higher MDA Cp values, there is no clear effect of GC content on the amount of genome recovered by the various treatments (Figures S1, S3).

The lack of amplicons from cells treated with paraformaldehyde is likely due to the crosslinks preventing the phi29 polymerase from accessing the DNA. The heat treatment to reverse the crosslinks was either insufficient to reverse them or it resulted in DNA strand breakage and depurination which damaged the DNA sufficiently that it was unable to be amplified.

The FISH process is intended to improve access of the oligo probe to its RNA or DNA target site in the cells. As this might result in greater access of the phi29 polymerase to the DNA, one would expect the MDA from this treatment to be the most efficient. However, the fact that the MDA kinetics were delayed and that the genome recovery was reduced compared to the cryopreserved samples indicates that other factors must be involved. As the washing steps should remove the components of the FISH buffers prior to MDA, they would not directly affect the MDA process. Thus the reduced genome recovery is likely derived from damage to the DNA during the FISH treatment. Since the FISH process can denature high AT regions of DNA at the temperatures used, the single stranded DNA that results could be more susceptible to damage (Blake & Delcourt, 1996). This interpretation is supported by the genome recovery in the low GC organism *P. heparinus* being lower than in the other two organisms (Figure S3). In addition, it has been shown that formamide can degrade purine nucleosides (Saladino *et al.*, 1996). Small amounts of damage could explain the reduction in the genome recovered by the MDA.

The significant reduction in genome recovery from ethanol preserved cells is challenging to explain. It is unlikely that ethanol could carry over and inhibit the MDA reaction because the cells were washed twice before sorting and the sorting process itself results in significant dilution of the sample (Rodrigue *et al.*, 2009). Since there is no known mechanism for ethanol to damage the DNA, the substantial reduction in the amount of the genome recovered must be due to the polymerase having restricted access to the DNA. Ethanol causes proteins to denature and precipitate (Yoshikawa *et al.*, 2012). These could aggregate around the DNA and prevent access by the polymerase.

Although single-cell genomics has great potential to provide insight into the vast number of microbes that have not been cultivated, sample handling can greatly impact the completeness of single-cell genomes. Our results suggest that samples that have been archived by preservation in paraformaldehyde will be unsuitable for the production of single-cell genomes and that ethanol-preserved samples are likely to produce single-cell genomes of reduced quality. Thus, we recommend use of cryopreserved specimens for best results and fixation-free FISH (Yilmaz *et al.*, 2010, Haroon *et al.*, 2013) if targeted flow sorting is to be employed.

ACKNOWLEDGEMENTS

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

S.C., P.H. and T.W. conceived of the study and designed the experiments. S.C. performed the experiments. S.C. and P.S. analyzed the data. S.C., P.S., P.H. and T.W. wrote the manuscript.

Supplementary information is available at ISMEJ's website.

REFERENCES

Blainey PC. (2013). The future is now: single-cell genomics of bacteria and archaea. FEMS Microbiol Rev 37:407-427.

Blake RD, Delcourt SG. (1996). Thermodynamic effects of formamide on DNA stability. Nucleic Acids Res 24:2095-2103.

Haroon MF, Skennerton CT, Steen JA, Lachner N, Hugenholtz P, Tyson GW. (2013). In-solution fluorescence *in situ* hybridization and fluorescence-activated cell sorting for single cell and population genome recovery. Method Enzymol 531:3-19.

Lasken RS. (2012). Genomic sequencing of uncultured microorganisms from single cells. Nat Rev Microbiol 10:631-640.

Marcy Y, Ishoey T, Lasken RS, Stockwell TB, Walenz BP, Halpern AL, *et al.*. (2007). Nanoliter reactors improve multiple displacement amplification of genomes from single cells. PLOS Genet 3:1702-1708.

Morales SE, Holben WE. (2011). Linking bacterial identities and ecosystem processes: can 'omic' analyses be more than the sum of their parts? FRMS Microbiol Ecol 75:2-16.

Podar M, Abulencia CB, Walcher M, Hutchison D, Zengler K, Garcia JA, *et al.*. (2007). Targeted access to the genomes of low-abundance organisms in complex microbial communities. Appl Environ Microbiol 73:3205-3214.

Rappe MS, Giovannoni SJ. (2003). The uncultured microbial majority. Annu Rev Microbiol 57:369-394.

134 Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, *et al.* (2013). Insights into the
135 phylogeny and coding potential of microbial dark matter. Nature 499:431-437.

136 Rodrigue S, Malmstrom RR, Berlin AM, Birren BW, Henn MR, Chisholm SW. (2009). Whole genome
137 amplification and *de novo* assembly of single bacterial cells. PLoS ONE 4:e6864.

138 Saladino R, Mincione E, Crestini C, Negri R, Di Mauro E, Costanzo G. (1996). Mechanism of degradation
139 of purine nucleosides by formamide. Implications for chemical DNA sequencing procedures. J Am Chem
140 Soc 118:5615-5619.

141 Woyke T, Sczyrba A, Lee J, Rinke C, Tighe D, Clingenpeel S, *et al.*. (2011). Decontamination of MDA
142 reagents for single cell whole genome amplification. PLoS ONE 6:e26161.

143 Yilmaz S, Haroon MF, Rabkin BA, Tyson GW, Hugenholtz P. (2010). Fixation-free fluorescence *in situ*
144 hybridization for targeted enrichment of microbial populations. ISME J 4:1352-1356.

145 Yoshikawa H, Hirano A, Arakawa T, Shiraki K. (2012). Mechanistic insights into protein precipitation by
146 alcohol. Int J Biol Macromol 50:865-871.

147

148

149 Titles and legends to figures

150 Figure 1. Schematic of experimental workflow. For clarity the treatments are color coded throughout
151 the figure and are: blue, Cryo – cryopreservation; green, FISH – FISH treatment; yellow, EtOH – ethanol
152 fixation; orange, PFA xlinks – paraformaldehyde fixation with crosslinks reversed; red, PFA –
153 paraformaldehyde fixation with crosslinks intact; brown, Pos – positive controls (wells each with 100
154 cryopreserved cells); purple, Neg – no cell negative controls. During single-cell isolation the layout used
155 for the single-cells on a 384-well plate is shown. Each strain/treatment combination was sorted
156 separately onto one plate per strain. The MDA step plots the Cp (inflection point) values for the real-
157 time MDA amplification curves for all three bacterial species combined. The MDA reaction is run for 16
158 hours so any wells that show no amplification by that time are indicated as 16+ on the graph. Each
159 column in the chart summarizes data from 120 separate MDA reactions except for the negative control
160 which is 60 reactions. 16S rRNA screening indicates the percentage of the wells for each treatment that
161 produced a 16S rRNA gene sequence from the proper organism.

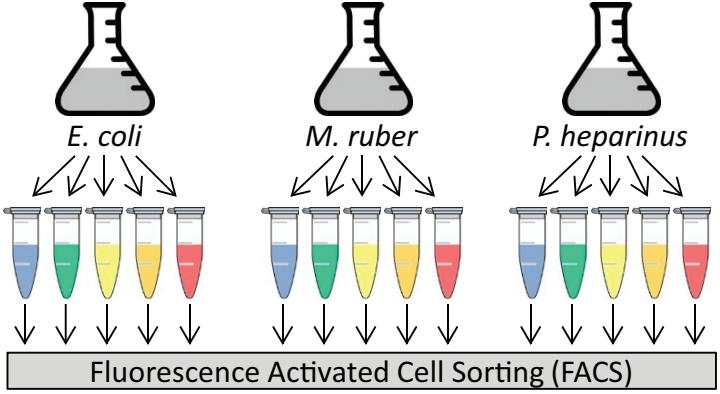
162

163 Figure 2. a - Coverage plots for each organism/treatment combination. Cryo – cryopreservation, EtOH –
164 ethanol preservation, FISH – FISH treatment. The horizontal black lines in the center of each plot
165 represent the complete reference genomes and the vertical colored lines indicate which parts of the
166 genome were recovered in the assemblies. Redder colors indicate that more single-cells contained that
167 region in their assembly. b - Percent of the genomes recovered for each treatmentwith the single cells
168 from all three organisms averaged together and the error bars indicating one standard deviation. Light
169 grey bars are the genome recovery by mapping the reads to the reference genomes. A base pair was
170 considered to be recovered if at least 10 reads were mapped to cover it. Dark grey bars are the genome
171 recovery by mapping contigs produced by *de novo* assembly to the reference genomes. The treatments
172 are significantly different from each other (ANOVA; $p < 0.0001$).

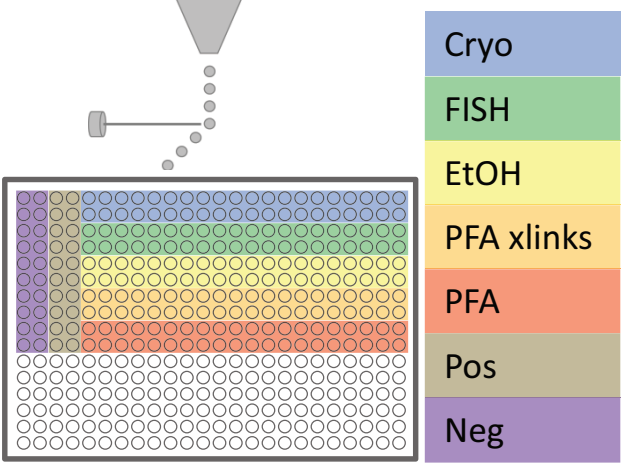
173

Culturing

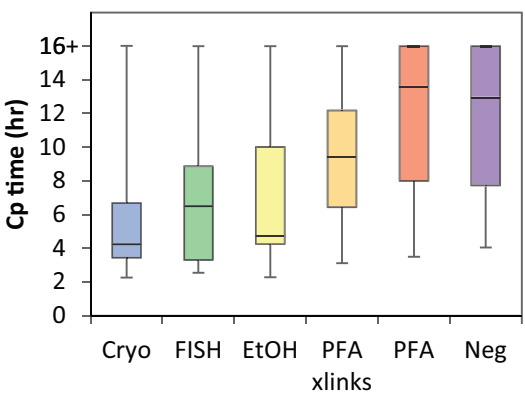
Treatments



Single Cell Isolation



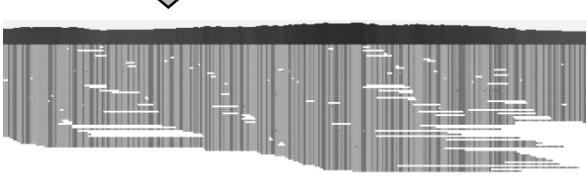
Multiple Displacement Amplification (MDA)

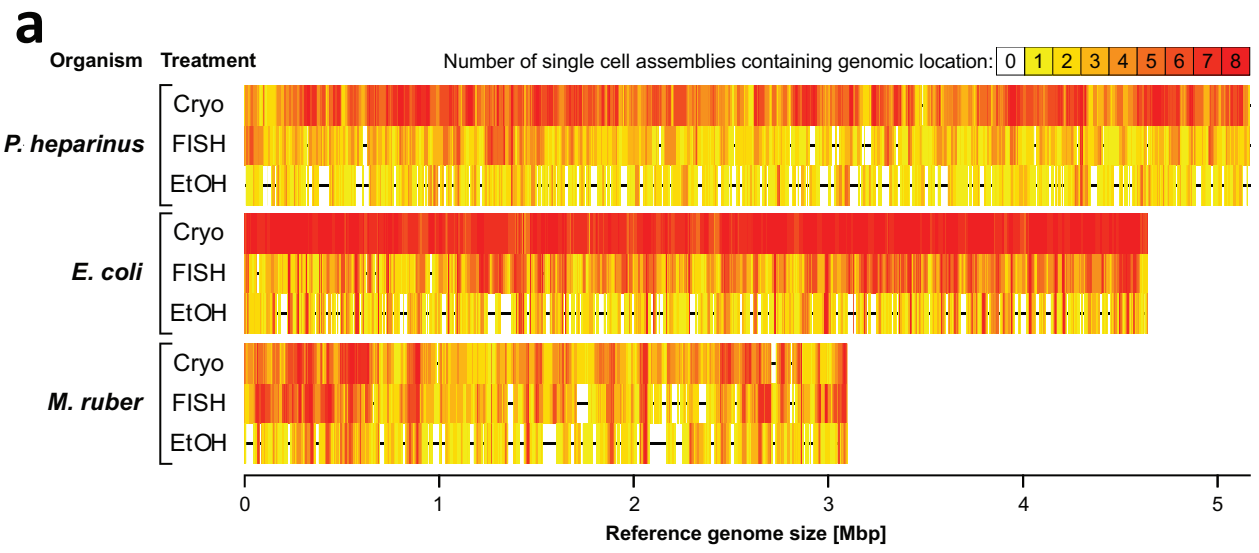


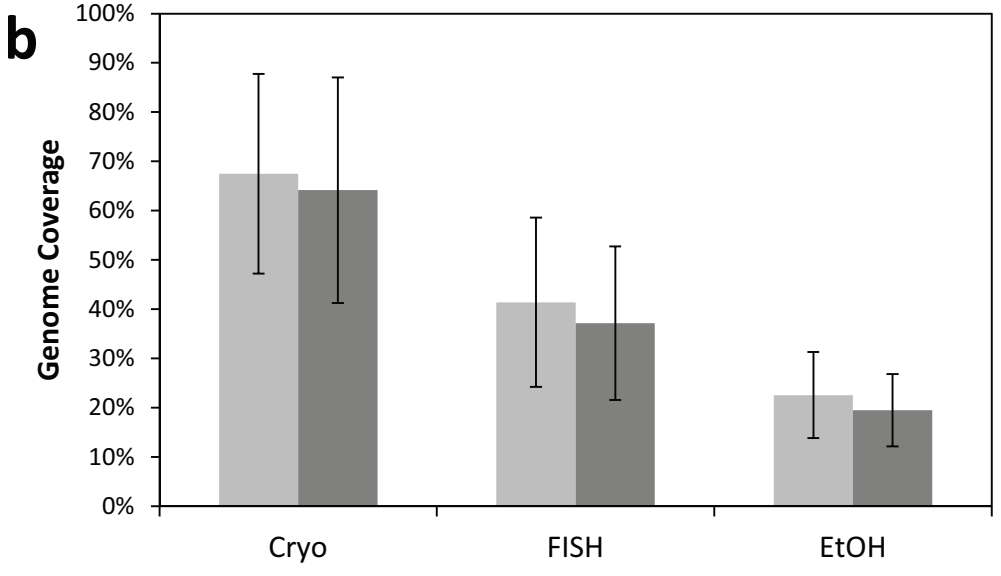
16S rRNA screening

Cryo	FISH	EtOH	PFA xlinks	PFA
69%	82%	50%	0%	0%

Sequencing & Data Analysis




a
Organism
Treatment
Number of single cell assemblies containing genomic location:
0
1
2
3
4
5
6
7
8
P. heparinus
Cryo
FISH
EtOH
E. coli
Cryo
FISH
EtOH
M. ruber
Cryo
FISH
EtOH
0
1
2
3
4
5
Reference genome size [Mbp]


b
Genome Coverage
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Cryo
FISH
EtOH