DNA Barcoding for Fractured Reservoir Analysis – An Initial Investigation

Yuran Zhang¹, Zhirui Zeng², Kewen Li¹ and Roland N. Horne¹

¹Dept. of Energy Resources Engineering, Stanford University, Stanford, CA, USA, 94305
²Dept. of Earth System Science, Stanford University, Stanford, CA, USA, 94305

yuranz4@stanford.edu

Keywords: smart tracer, fracture characterization, DNA, barcode, thermophilic microorganisms, archaea, nanoparticles

ABSTRACT

Tracer testing in fractured reservoirs has long been limited by the availability of unique and unambiguous tracers. In this study, we introduce the concept of using thermophilic microorganisms and DNA barcoding techniques for fractured reservoir characterization. On the one hand, thermophilic microorganisms have the potential of thriving in the extreme environment of geothermal and oil reservoirs. On the other hand, even if the microbes lose the basic functions of life after passing through the subsurface, their genetic information may still be retrieved and identified using modern sequencing technologies. In addition, because microorganisms are able to replicate themselves given appropriate nutrients and growing conditions, they are also a cost-effective solution for potential up-scaled field tracer testing applications. Heating experiments were conducted to examine the possibility of short-fragment identification of archaeal genomic DNA after prolonged heat treatment. The results show that it is feasible to still identify a short region of genomic DNA even after prolonged heat treatment, and a short region within a long genomic DNA appears to be more robust than a similar-length short-stranded DNA. We therefore conclude that the concept of using thermophilic microorganisms combined with DNA barcoding techniques to perform unambiguous reservoir tracer analysis shows good promise. To avoid confusion with the indigenous microbes in a reservoir, either exogenous microbes or engineered microbes containing artificial barcodes may be used.

1. INTRODUCTION

Geothermal energy extraction is largely dependent on the reservoir fracture network. In order to map fracture distribution in geothermal reservoirs, tracer testing has been widely used, mostly by injecting chemical tracers at the injection well and monitoring its breakthrough pattern at the production well, which gives information about inter-well connectivity. Because tracer transport is faster than the thermal front, tracer testing can also be used to infer in-situ reservoir temperature and hence forecast thermal breakthrough.

Conventional solute tracers, however, provide limited information about the reservoir due to their limited sensing mechanisms and limited varieties. In recent years, there has been increasing research interest in the development of “smart” tracers, including nanoparticle-based, that are able to carry much more information about the reservoir than conventional solute tracers. Alaskar et al. (2015) demonstrated the potential of irreversible thermochromic microspheres and dye-attached silica nanoparticles as temperature sensors to measure the temperature of subsurface reservoirs. Frane et al. (2014) investigated polymer microcapsules with encapsulated dye that could release dye at a certain temperature threshold. Rose et al. (2011) investigated Cadmium-based quantum dots with adjustable physical/chemical properties as both conservative tracer and sorbing tracer to characterize near-wellbore and interwell fracture surface area resulting from hydraulic stimulation processes. There has also been rising interest in cabogenic nanoparticles (CNPs) with various optical behaviors as potential reservoir tracers, and the feasibility has been demonstrated via field tests (Bourlinos et al. 2008, Kryssmann et al. 2012, Kanj et al. 2015, Hawkins et al. 2016).

Besides the physically/chemically adjustable nanoparticle tracers that provide in-situ reservoir information by undergoing physical/chemical changes after traveling through the reservoir, another category of DNA-based “smart” tracers has been proposed that may better “illuminate” the reservoir by the potential of resolving the ambiguity issue in tracer testing, as described by Zhang et al. (2015, 2016). Ambiguity refers to the situations where naturally occurring substances in the reservoir overlaps with injected tracers, resulting in the inability to identify the injected tracers in the produced fluids (Figure 1). Juliusson et al. (2015) reported a field tracer experiment in Krafla Reservoir, Iceland, where tracers were injected into three wells and collected in 20 production wells. However the partitioning tracer used in the test was detected in the production wells even before injection took place, rendering the partitioning tracer test of limited informativeness. Rose et al. (2011) also pointed out the complexity of recovered tracer analysis because the fluorescence wavelengths of many preexisting reservoir components overlap with that of the injected tracers. Brauchler et al. (2013) performed 72 consecutive gas-tracer (Helium) tests in a tracer tomography experiment by systematically changing the injection/monitor configuration after each test, in order to invert on the tracer data to allow the three-dimensional reconstruction of interstitial velocity distribution in the fractured sandstone block. Had there been more distinct tracers available, it could have been possible to obtain sufficient tracer data for the inversion after only one or a few tracer tests. In order to address the ambiguity problem, Zhang et al. (2015, 2016) investigated the synthesis and flow of DNA-embedded silica nanoparticle tracers with the potential of tagging flow paths or wellbores with infinite numbers of unique “signatures” that can be identified and quantified via qPCR (quantitative polymerase chain reaction), thus eliminating ambiguity.
In this work, further investigation of DNA-based tracers for unambiguous fluid identification was performed. The concept of DNA barcoding originated from biological classification was adapted for fractured reservoir characterization. The possibility of using naturally existing thermophilic microorganisms as tracers for flow path analysis was studied. A heating experiment was designed to examine the possibility for archaeal genomic DNA (gDNA) to still carry barcode information after prolonged heat exposure. The reservoir conditions (temperature, pressure, etc.) under which this concept is applicable was briefly discussed.

2. BACKGROUND

2.1 Thermophilic Microorganisms

Thermophilic microorganisms, or thermophiles, are a type of extremophiles that thrive at high temperatures up to ~120 °C (Takai et al. 2008), pH varying from highly acidic to strongly alkaline, and elevated concentrations of heavy metals and salts. Most hyperthermophiles (i.e. thermophiles with optimum growth temperature above 80 °C) are archaea. The mechanism behind the ability of hyperthermophiles to survive very high temperatures lies in their efficient thermoprotection of nucleic acids, including: high G+C content in RNA, presence of Na⁺, K⁺, Mg²⁺ that protect dsDNA against chemical thermodegradation of the phosphodiester bonds, DNA repair, etc. (Grosjean et al. 2007).

Geothermal reservoirs are known to harbor a vast variety of thermophilic microorganisms. The use of indigenous or exogenous microbes as reservoir tracers greatly utilizes biodiversity and could resolve one of the biggest challenges in geothermal reservoir engineering.

2.2 DNA Barcoding for Microorganisms

DNA barcoding refers to a taxonomic method that uses a short genetic marker within the DNA of an organism to identify it as belonging to a certain species (Hebert et al. 2003; Shokralla et al. 2012). While the cytochrome c oxidase subunit I (COI) gene region is generally used for discerning closely related species across Animalia, for the bacterial and archaeal communities, species identification are commonly conducted using 16S ribosomal RNA (16S rRNA) gene consisting of approximately 1.5k base pairs (bp) (Zoetendal et al. 2006; Flanagan et al. 2007; Kim et al. 2012; Shokralla et al. 2012; Yarza et al. 2014). Highly conserved regions of the 16S rRNA gene enable “universal” PCR primer to amplify this gene from most bacteria and archaea, whereas variable regions of the 16S rRNA gene allow the discrimination between bacterial or archaeal types (Flanagan et al. 2007; Mizrahi-Man, 2013). The ever-advancing sequencing technology has enabled the development of large DNA barcode reference libraries that can be easily accessed online. The complete genome information of most microbes of interest is therefore readily available for analysis, usually through fast and accessible algorithms such as BLAST, FASTA, etc.

Geothermal and oil reservoirs are known to harbor a large variety of microorganisms including both bacteria and archaea. There has been continuous research examining the potential for both indigenous and exogenous microbes to fulfill multiple applications such as enhanced oil recovery, CO₂ sequestration, etc., some of which have already been applied at field scale, contributing to great economic impact because microorganisms are cost effective and environmentally friendly (Lin et al. 2014; Liu et al. 2015). Although there has been hydrologic studies of aquifers that employed bacteria that form red or yellow pigments as tracers, those studies used conventional microscopy or culturing techniques to detect recovered microbes, and took little advantage of the inherent genetic diversity (Harvey 1997).

In this study, we propose the use of microorganisms, preferably thermophiles, as tracers for unambiguous fracture characterization in geothermal or oil reservoirs, facilitated by the DNA barcoding concept and modern sequencing technologies that has become much more efficient and cost effective within the past decade. Either exogenous or recombinant microorganisms may be applied to a geothermal or oil field to avoid confusion with the preexisting microbiota. In addition, microorganisms are able to replicate themselves
given appropriate nutrients and growing conditions, thus they are also a cost-effective solution for potential up-scaled reservoir characterization applications.

2.3 Heat Stability of Archaeal Genomic DNA

One of the challenges for a tracer candidate to function in geothermal reservoirs is its ability to survive high temperature. Although thermophiles or hyperthermophiles are able to thrive at temperatures from around boiling point of water to as high as 121 °C (Takai et al. 2008), geothermal reservoirs for electricity generation usually exceeds 150 °C. However, the objective of using the genetic information to trace flow paths does not require the viability of microorganisms. In other words, even if the high geothermal temperature caused the microbes to lose the basic functions of life, it is possible that the microbial genomic DNA still maintains sufficient integrity to be identified: Microbial genomic DNA generally contains millions of base pairs, however identifying a species usually only requires the sequencing of much shorter regions of the 16S rRNA gene (100~500 bp) (Liu et al. 2007; Bikel et al. 2015). Therefore it is likely that even if genomic DNA undergoes some degree of degradation after prolonged exposure to extreme temperature, the genetic information can still be extracted and identified based on the DNA fragments, as illustrated in Figure 2.

![Figure 2: Illustration of the possibility that target region is still identifiable even after prolonged heat treatment of genomic DNA (left). In comparison, shorter strands of DNA are much more exposed to degradation and become less identifiable after heat treatment (right). The dark red lines on the bottom right represent damaged target sequences that are no longer amplifiable.](image)

DNA detection after intensive heat exposure has been investigated by researchers from other fields including environmental engineering, food processing, etc. Fu et al. (2009, 2011, 2012) examined the possibility for thermally treated waste recombinant plasmid DNA to renature and potentially cause gene pollution when discharged into the environment. They found that after being thermally treated at 100 °C for as long as 2 hours, or being autoclaved at 120 °C, the 245 bp fragment inserted to the plasmid DNA was still detected and identified via qPCR. Yap et al. (2013) conducted autoclaving on bacterial genomic DNA followed by PCR, and found that although autoclaved bacteria were non-culturable, amplifiable PCR signal was observed with autoclaving duration as long as 40 min. Hird et al. (2006) observed PCR amplification using amplicons of as long as 351 bp for highly processed meat samples (boiled, baked, or autoclaved).

Therefore, thermophilic archaea possess good potential for geothermal reservoir tracers. On the one hand, their ideal culture temperatures are much higher than that of other microorganisms. On the other hand, in the case of inviability due to exposure to temperatures higher than their growing temperature, there is still good chance that their genomic DNA, even fragmented, are sufficient for them to be identified and correlated to the subsurface fracture network.

2.4 DNA Fragmentation Mechanisms and Stabilization

There are a number of mechanisms that lead to DNA strand breaks including oxidation, physical shearing, endonuclease activity, molecular crosslinks, etc. (Lindahl et al. 1972, 1993; Fu et al. 2009) A good understanding of DNA decay mechanisms provides insight into possible DNA stabilization in fracture characterization applications. NaCl is known to enhance DNA stability by increasing the ionic strength of solution. EDTA, a widely-used complexing agent, can bind metal ion co-factors of DNase and hence prevent DNase-induced DNA degradation. pH also affects the thermal stability of DNA because acidic conditions promote DNA degradation (Lindahl et al. 1972; Fu et al. 2009). Physical shearing is another important factor. Under unpressurized boiling conditions, microbial DNA samples contained more short fragments than those treated pressurized at 120 °C, which was likely to be caused by mechanical shearing in boiling water (Fu et al. 2012). Lindahl et al. (1993) also pointed out that high pressure provides some protection to the DNA double
Zhang et al.

helix, in that the melting temperature of DNA double helix is about 10 °C higher at 5000 atm. In addition to the effect of heating conditions, allowing DNA samples to cool and renature also has an impact on the detectability of DNA fragments. Fu et al. (2011) demonstrated that in the range of 4 °C to 37 °C, a higher temperature led to more efficient recovery of the thermally treated double-stranded plasmid DNA.

Besides the fragmentation mechanisms of DNA in liquid solution available in the literature, Karni et al. (2013) pointed out that DNA may possess better heat stability at dry conditions with degradation starting at 130 °C. This may be an explanation of why DNA embedded within silica nanoparticles was able to withstand heat treatment better than unprotected DNA (Paunescu et al. 2013, Zhang et al. 2016), because separating DNA from water reduces the chance of hydrolysis.

These DNA stabilization mechanisms may provide valuable insights as to the future improvements to the DNA “tags”, selection of parameters for flow experiments, and ultimate field applications.

3. MATERIALS AND METHODS

In this study, heating experiments on the genomic DNA of the hyperthermophilic archaeon Sulfolobus Acidocaldarius (DSM 639) that grows optimally at ~80 °C were designed to examine the possibility of short fragment identification of archaeal gDNA after prolonged heat treatment. Synthetic double stranded DNA (dsDNA) with 113 bp was also heated under the same conditions to provide a comparison between the heat behavior of short dsDNA (113 bp) and a similar-length region on a long dsDNA (~2 million bp).

3.1 Genome DNA of Sulfolobus Acidocaldarius (DSM639) Extraction and Purification

The Sulfolobus Acidocaldarius (DSM639) cell culture used in this study was at log phase, with OD$_{600}$ = 0.6, which was equivalent to ~4.8×10$^9$ cells/ml. Because this initial study focuses on the genomic DNA of thermophilic archaea, the cell suspension was divided into 2-ml aliquots, and the archaea genome was extracted using PureLink™ Genomic DNA Mini Kit manufactured by Invitrogen according to the manual. All extracted genome solution was mixed to eliminate the possible concentration fluctuation resulting from the genome extraction process. The concentration of extracted genomic DNA was estimated to be 35.6 ng/µl via spectrophotometry (NanoDrop Lite, Thermo Scientific).

The genome of DSM639 contains 2,225,959 bp, and the complete sequence of the genome is readily available in the NCBI database. As explained in Section 2.2, the 16S rRNA gene of DSM639 (1480 bp) will be used as the target region for identifying and quantifying its gDNA.

3.2 Heating Tests

Heating tests of both short synthetic DNA and long genomic DNA were performed in stainless steel vessels that were able to pressurize and prevent boiling when temperature was raised to above 100 °C, one of the vessels is shown in Figure 3. 200 µl of the extracted gDNA (~35.6 ng/µl) was mixed with TE buffer ([1×]: 10 mM Tris, 1 mM EDTA) that eventually filled each vessel. 140 µl of synthetic DNA solution (~71.4 nM) was mixed with TE buffer ([1×]) that eventually filled each vessel. The weight of each vessel before and after fluid addition was recorded to obtain the exact volume of the fluid contained. The volume of each vessel was around 5 ml. The filled vessels were then put into an air bath and heated at various temperatures and durations, according to the schematic shown in Figure 4. A room temperature control was included as a comparison, and the control was also placed into the stainless steel vessel followed by subsequent solution transfer, in order to account for possible solution loss during the transfer. After heating, each vessel was placed in a room temperature water bath (~20 °C) and was allowed to cool for 40 min. After cooling, each sample was transferred to polypropylene centrifuge tubes and stored at -20 °C until subsequent analysis. No loss of fluid was observed after heating, indicating good sealing capability of the stainless steel vessels.

![Figure 3: Heating vessels were assembled using a 2-inch segments of stainless steel tubing with screw caps tightened on both ends to avoid solution loss caused by vaporization. 20 vessels were assembled for the 20 samples.](image-url)
3.3 Quantitative Polymerase Chain Reaction (qPCR) Analysis

All heated samples were analyzed via qPCR at Stanford Protein and Nucleic Acid (PAN) Facility using SybrGreen chemistry. DNA quantification was conducted by establishing standard curves using the expected PCR product with known concentrations for each assay (standard DNAs were synthesized also at Stanford PAN Facility).

For synthetic DNA, the same primers as illustrated in Zhang et al. (2015) were used to amplify the whole 113 bp region of the synthetic DNA. For DSM639 genome DNA, however, primers were designed using the online tool BLAST that amplifies a 141 bp region of its 16S rRNA gene. The sequence of the 141 bp region (i.e. expected PCR product) is:

\[(5'\rightarrow 3')\text{ AGTCA }GGCAA \text{ GGTG } TTAGC \text{ CGTCA } TACTG \text{ GGTCG } TCCCC \text{ CGGTA } AGATT \text{ CCAGG } CGTTG \text{ ACTCC } AATTG \text{ AGCCG } TGACG \text{ TAGT } TCCGT \text{ CACATG } CTTCT \text{ ACCGC } CGTCA \text{ TTTCG } GACCG \text{ C T}\]

The sequences of the primers for this 141 bp region are: (5'-3') ACGGT CGCAA GACTG AAACT and (5'-3') AGTCA GGCAA GGTG TTAGC CTTGC

qPCR was performed using a standard 40-cycle protocol in the StepOnePlus™ Real-Time PCR System.

4. RESULTS AND DISCUSSION

It is anticipated that the ~2 million bp genomic DNA would undergo less decay than the short 113 bp synthetic DNA. On the one hand, one of the mechanisms for hyperthermophiles like DSM639 to survive extreme environments may lie in their DNA structure (Grosjean et al. 2007). On the other hand, assuming DNA strand break occurs according to a random process long the DNA strand (Fu et al. 2009), then for short synthetic DNA, every strand break leads to the reduction of target PCR regions in the sample, whereas for gDNA, the likelihood for target PCR regions to become damaged is much lower because of the presence of the other ~2 million bp of non-target region (Figure 2).

The qPCR results for all samples are summarized in Figures 5-7. Due to the difference in DNA length and PCR amplicon length, it is infeasible to compare the heated sample concentrations directly. Therefore, all concentrations were normalized by dividing by the concentration of the room temperature control for each DNA type. In addition, the fluctuation in heating vessel volumes was also taken into account to adjust the concentration values. The final normalized concentrations of all heated samples were also organized into three plots by fixing one parameter at a time (Figure 6~7) for direct visualization of the DNA thermo behavior.

Note that strictly speaking, “the concentration of genomic DNA” refers to the concentration of the intact 141 bp region on the genome that was later amplified by qPCR.
Figure 5: Summary of normalized quantify of all heated samples. Blue boxes represent short (113 bp) synthetic DNA samples and orange boxes represent long (~2 million) genomic DNA samples. Blank (white) boxes are heating temperature and durations yet to be performed.

<table>
<thead>
<tr>
<th>Normalized Quantity of Heated Samples</th>
<th>Legend: 1-synthetic DNA</th>
<th>2-genomic DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>150 °C</td>
<td>0.81 1.09 0.87 1.55 0.78</td>
<td>0.98 0.68 1.21</td>
</tr>
<tr>
<td>120 °C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 °C</td>
<td>1.25 1.17</td>
<td>0.79 1.05</td>
</tr>
<tr>
<td>80 °C</td>
<td>0.93 1.18</td>
<td>0.69 1.46</td>
</tr>
<tr>
<td>RT</td>
<td>1.00 1.00</td>
<td></td>
</tr>
<tr>
<td>5 min 10 min 20 min 30 min 1 hr 2 hr 4 hr</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 6: DNA concentration (normalized) versus heating duration when treated at 150 °C.

Figure 7: DNA concentration (normalized) versus heating temperature for 5 min heating (left) and 30 min heating (right)

As shown in Figures 5~7, all heated samples remained above the detection limit of qPCR. From Figure 6, it can be seen that at 150 °C, heating time within 30 min did not have a significant impact on either synthetic DNA or genomic DNA. When heating duration was increased to 60 min, the degradation of both synthetic and genomic DNA became more prominent. However, the degree of degradation remains below the detection limit of qPCR.
in the 141 bp region of the hyperthermophilic DSM639 genome was less than that of the short synthetic DNA by approximately an order of magnitude. The difference was not much, but in accordance with our anticipation.

From Figure 7, it can be shown that heating durations of 5 min or 30 min at 80, 100 and 150 °C had little impact on the integrity of synthetic DNA or genomic DNA. It is hard to tell from Figure 7 which of the synthetic and genomic DNA performs better because their difference in normalized concentrations is very small, although it is worth noting that in both plots of Figure 7, genomic DNA outperforms synthetic DNA slightly.

Note that the phrase “no significant degradation of genomic DNA” does not necessarily mean the whole ~2 million bp genome is intact. Instead, it indicates the integrity of the 141 bp target region of the ~2 million bp genomic DNA, which is sufficient for the genome to be identified.

Also note that due to the limitations of the genome extraction kit as described in Section 3.1, the initial concentration of heated genomic DNA in terms of copy number per volume was much lower than that of the heated synthetic DNA. In addition, the primers that are able to identify the 16S rRNA gene of DSM639 are not unique. The primer set used in this study gives a PCR product of 141 bp, which is 25% longer than the synthetic DNA comparison. Had the primers been chosen to give a shorter PCR product, the probability that the shorter amplicon remains intact after heat treatment may rise, resulting in better thermo stability of genomic DNA as reflected by qPCR analysis. Primer specificity, on the other hand, will decrease in this case.

Nevertheless, the drop in concentration after being heated at 150 °C for 60 min is also an indicator that the structure of genomic DNA itself may only play a partial role in the capability of hyperthermophiles to survive high temperatures. It is therefore worthwhile to examine the possibility of using living hyperthermophilic microorganisms as reservoir tracers so that other survival mechanisms can come into play, as mentioned in Section 2.1. To avoid ambiguity with the indigenous microbes in a reservoir, either exogenous microbes or artificially engineered microbes that do not naturally exist may be used.

5. CONCLUSIONS

In this study, the concept of using thermophilic microorganisms and DNA barcoding techniques for fractured reservoir characterization was introduced. On the one hand, thermophilic microorganisms have the potential of thriving in the extreme environment of geothermal and oil reservoirs. On the other hand, even if the microbes lose the basic functions of life after passing through the subsurface, it is still likely that their genetic information can be retrieved and identified using modern sequencing technologies that have become much more affordable in the past decade. In addition, because microorganisms are able to replicate themselves given appropriate nutrients and growing conditions, they are also a cost-effective solution for potential up-scaled tracer testing applications.

A heating experiment was designed and partially fulfilled that examined the possibility of short-fragment identification of archaeal genomic DNA after prolonged heat treatment. Synthetic double stranded DNA (dsDNA) with 113 bp was also heated under the same conditions to provide a comparison between the heat behavior of short dsDNA (113 bp) and a similar-length region on a long dsDNA (~2 million bp). The results showed that heating at 150 °C within 30 min did not have a significant impact on either synthetic DNA or genomic DNA. When heating duration was increased to 60 min at 150 °C, the degree of degradation in the 141 bp region of the hyperthermophilic DSM639 genome was less than that of the short synthetic DNA by around an order of magnitude. Therefore, it is demonstrated by the heating experiments that: 1. DNA itself is not too fragile under certain conditions; 2. It is feasible to still identify a short region of genomic DNA even after prolonged heat treatment, and a short region within a genomic DNA appears to be more robust than a similar-length short stranded DNA.

Therefore, the concept of using thermophilic microorganisms combined with DNA barcoding techniques to perform unambiguous reservoir tracer analysis shows good promise. To avoid ambiguity with the indigenous microbes in a reservoir, either exogenous microbes or engineered microbes containing artificial barcodes may be used.

6. FUTURE WORK

Future work will focus on fulfilling the rest of the heating experiments as illustrated in Figure 4 in order to obtain more thorough understanding of the thermal stability of hyperthermophilic archaea genome under high temperatures. Other DNA analyzing techniques (gel electrophoresis, Sanger sequencing, etc.) will be incorporated to provide more information on PCR efficiency, primer specificity, etc. The feasibility of using living hyperthermophilic microbes as tracers will be investigated. The flow properties of genomic DNA and hyperthermophiles will be examined via flow experiments.

ACKNOWLEDGEMENTS

The authors would like to thank the Office of the Dean of the School of Earth, Energy & Environmental Sciences at Stanford University for supporting this work, Karrie Weaver at Stanford ICP-MS/TIMS Facility for providing resources for DNA extraction experiments, Jeremy Wei and Paula Welander from Welander’s Lab at the School of Earth, Energy & Environmental Sciences at Stanford University for providing valuable resources and thoughts, and Alberto Lovell at Stanford PAN Facility for his time and insight regarding qPCR analysis.

REFERENCES

Zhang et al.


