Tracing Septic Pollution Sources Using Synthetic DNA Tracers: Proof of Concept

Authors: Georgakakos, Christine B, Richards, Paul L, and Walter, M Todd

Source: Air, Soil and Water Research, 12(1)

Published By: SAGE Publishing

URL: https://doi.org/10.1177/1178622119863794
Tracing Septic Pollution Sources Using Synthetic DNA Tracers: Proof of Concept

Christine B Georgakakos and Paul L. Richards and M Todd Walter

1Department of Biological and Environmental Engineering, Cornell University, Ithaca, NY, USA.
2Department of Earth Sciences, The College at Brockport, State University of New York, Brockport, NY, USA.

ABSTRACT: Contamination from septic systems is one of the most difficult sources of nonpoint source (NPS) pollution to quantify. Quantification is difficult in part because locating malfunctioning septic systems within a watershed is challenging. This study used synthetic-DNA-based tracers to track flows from 2 septic systems. Sample DNA was quantified using quantitative polymerase chain reaction (qPCR). This technology could be especially useful for simultaneously assessing multiple septic systems because there are essentially infinite unique combinations of DNA bases such that unique tracers could be engineered for each septi system. Two studies were conducted: the first, to determine whether the tracers move through septic systems (experiment 1), and the second, to determine whether the tracers were detectable at watershed scales (experiment 2). In both cases, clear, although complex, breakthrough curves were detected. Experiment 1 revealed possible preferential flow paths that might not have been otherwise obvious, indicative of short circuiting systems. This proof of concept suggests that these tracers could be applied to watersheds suspected of experiencing NPS septic system pollution.

KEYWORDS: nonpoint source pollution, septic system, DNA tracer, groundwater, quantitative polymerase chain reaction (qPCR)

Received: June 24, 2019. Accepted: June 24, 2019.

TYPE: Original Research

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by New York Great Lakes Research Consortium using funding provided by the New York Great Lakes Basin Advisory Council and was supported through funding authorized under Section 104b of the Water Resources Research Act, administered by the US Geological Survey.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Christine B Georgakakos, Department of Biological and Environmental Engineering, Cornell University, 111 Wing Drive, Ithaca, NY 14853, USA.

Email cbg46@cornell.edu

Introduction

Septic systems provide important, low-cost wastewater treatment, especially in rural areas. However, failing septic systems can have negative impacts on water quality and soil chemistry. Septic system leaks pose health hazards by potentially introducing viruses, fecal contamination, and nutrients to drinking water sources. Contaminant loads to streams can have especially negative impacts on headwater streams, potentially causing algal blooms and other ecological imbalances. Because approximately 70% of the landscape drains directly to first-order headwater streams, it is especially important to monitor pollution of these systems. Because septic systems are the predominant water treatment systems in rural areas, tracing septic pollution is critical in predicting wastewater impacts on surface waters.

Tracing septic system tracers

Detecting septic leaks has taken several paths in the last few decades. Traditional tracers such as dyes, salts, and ionic ratios have been used to detect septic system leakages by monitoring concentration fluctuations at differing points in septic systems and surrounding surface and groundwater. Ionic ratios are not optimal tracers because ionic signatures of waste sources can change between sources and degenerate in chemical reactions within the sewage system. Ionic ratios, dyes, and salts can all fail when concentrations drop below detection limits, often due to dilution by rainwater.6 Eiswirth and Hötzel detected leaks by identifying elevated groundwater concentrations of nitrate, chloride, and phosphate. Fluorescent dyes have also been used to determine whether a septic system is contaminating groundwater. Although Borchardt et al2 used fluorescent dyes to determine the point of leakage in a single system, this technology could be applied to various systems leaking into the same water body, assuming different dyes were used per system. The use of fluorescent dyes is limited due to photo decay by sunlight, adsorption to surfaces, background fluorescence values, and high detection limits.

Aside from traditional chemical and fluorescing tracers, there has been significant work tracing septic pollution using microorganisms and micropollutants. Fahrenfeld et al8 used fecal sterol analysis to determine the presence and absence of human fecal contamination. At the watershed scale, Bowah et al9 used bacterial genetic markers to identify watersheds with higher septic pollution. Source-tracking methods using the genetic idiosyncrasies among fecal bacteria and fecal indicator bacteria have been successfully used. However, this tracking requires extensive, sometimes prohibitive, effort (eg, developing genetic libraries) to isolate sources.

Micropollutants have also recently been used as indicators of wastewater pollution. Carpenter and Helbling10 used micropollutants to identify differing pollution sources. Richards et al10 used micropollutants to trace human activity between watersheds of greater and lower septic system density. However, tracing micropollutants does not allow for the rapid determination of which septi systems among a network of multiple systems are contributing to water quality degradation without a detailed knowledge of the micropollutant signature entering...
each septic system of interest, which is difficult to obtain and may not be unique.

Despite the long history of using septic systems to treat wastewater, current leak assessment technologies lack an ability to make a rapid diagnosis to identify leaking septic systems among networks of septic systems. Relatively, recent synthetic DNA tracer technology overcomes this obstacle by identifying where pollutants originate and allowing direct management of those sources.

**Synthetic DNA-based tracers**

These tracers use short (~100 base pairs) single strands of synthetic DNA encapsulated in a biodegradable plastic sphere. The DNA sequences used are synthetically designed sequences that do not appear in the natural environment. Using synthetic sequences eliminates detecting naturally occurring environmental DNA that may introduce noise into the analysis. DNA is an ideal tracer because of the vast quantity of uniquely identifiable tracers that can be synthesized. Thus, even with short strands of ~100 bases, \(4^{100} = 1.6 \times 10^{60}\) unique strands can be theoretically created. Tracer concentrations are analyzed using quantitative polymerase chain reaction (qPCR). The qPCR technology allows for a theoretical concentration differentiation of DNA on a 0- to 10-strand scale, but a detection limit of ~100 DNA strands is more realistic according to this analysis. Thus, this technology is ideal for systems with large dilution challenges.

The synthetic DNA tracer technique, pioneered by Sharma et al.\(^{13}\) and fabricated here using the same protocol, provides a unique identifier to determine which system, among many, is contributing to water and soil contamination. This promising new technology, previously used for other water tracing applications,\(^{14}\) is applied here to the septic system problem, ie, identifying leaks and DNA tracers before running the experiment. This preliminary run allowed identification of 4 wells to be sampled less than 24 hours by testing 8 wells after introduction of chlorine.

**Objective**

The objective of this study was to ascertain the efficacy of using synthetic-DNA-based tracers in septic systems considering local and watershed scales. Two experiments were conducted in this study to test the tracers on differing scales. Experiment 1 tracks tracer movement on a local scale (~100 m), whereas experiment 2 tracks tracer movement at a small watershed scale (~1 km).

**Methods**

**Experimental designs**

In experiments 1 and 2, tracers were introduced to the septic system via a toilet. The septic systems are connected to leach fields where wastewater is dispersed via perforated pipes. All samples of both experiments were subsampled in duplicate and refrigerated until qPCR analysis. Both experiments were conducted in late fall when groundwater table was high and soils likely to be saturated.

**Experiment 1 design.** Experiment 1 tested the application of DNA tracers in a septic system using a single bedroom residence. The residence's septic tank was 1.9 m\(^3\) (500 gal). The toilet used in this experiment is the only toilet in the residence. This experiment was designed to assess tracer transport at a small scale through the septic system and leach field. The experimental site was located in Rush, New York, adjacent to Honeoye Creek, where the local depth to groundwater was on average 0.4 m from the surface during the experiment across all 4 monitoring wells. The septic tank and the leach field were located 12.5 and 11.3 m away from a wetland, respectively. This setup was designed to determine whether the tracers would be detectable near the wetland. This residence was chosen due to suspected short circuiting of the septic system as evidenced by surface erosion.

The residence time of the septic tank was determined to be less than 24 hours by testing 8 wells after introduction of chloride and DNA tracers before running the experiment. This preliminary run allowed identification of 4 wells to be sampled for the duration of the experiment. Samples were taken at 30-minute intervals for 10 hours for experiment 1. A polyvinyl chloride (PVC) ball bailer slurry sampler was used to sample wells, rinsing the sampler between wells and sample iterations. Samples were stored in 60 mL Nalgene bottles, subsampled, and refrigerated until analysis.

The DNA tracer was injected into the system at time 0. After determining likely routes using electrical conductivity measurements (as in Alhajjar et al.\(^{23}\)), 4 wells were sampled...
every 30 minutes during the experiment (wells 0-3, Figure 1A). Wells 1, 2, and 3 were located 0.6 m upland from the wetland’s saturated area and well 0 was in the leach field (Figure 1A). Wells 1 and 2 were located 7.3 m apart, whereas wells 2 and 3 were located 4.4 m apart. Five additional wells (not shown in Figure 1A) were sampled initially to determine base electrical conductivity and ambient positive DNA of the groundwater samples. The 4 wells in this experiment were observed to be in the direct path of septic leachate as a result of that initial electrical conductivity experiment. Wells were installed 4 months prior to running the DNA tracer experiment.

In this experiment, the tracer travels through 5 to 13 m of silt loam underlain by clay before reaching the sampling wells. All soils in experiment 1 are located within the Wayland soil group in the 0% to 3% slope classification of frequently flooded areas.23 Due to the locations of wells 1 to 3 being 0.6 m from the saturated areas of the wetland, they experience more frequent flooding than well 0. Wells 1 to 3 all have high organic matter contents due to proximity to the wetland.

**Experiment 2 design.** In experiment 2, a Teledyne ISCO ©1996 water sampler was used to take 500 mL stream water samples in high-density polyethylene (HDPE) containers every 12 hours. The samples were collected every 2 weeks for about 1.5 months. The DNA tracer was injected into the system at time 0.

Experiment 2 was conducted using a 3-bedroom home in Webster, New York, located 100 m from an unnamed tributary to Four Mile Creek (Figure 1B). The residence is equipped with a 3.8-m³ (1000 gal) septic tank. The leach field was properly built to code and is approximately 250 m from the small stream. No issues have been reported for the septic system in the past and no evidence of surface erosion is present. The septic system treats wastewater from a 3-bedroom residence, inhabited by 3 people. The tracer in this experiment traveled through a minimum of 250 m of gravel loam and along 1 km of stream before sample collection. Samples were collected at the black triangular icon in Figure 1B.

This experiment is a small watershed-scale test of the methodology. Tracer solution was introduced to via a toilet in the residence. The residence contained one other toilet connected to the same septic system. The solution contained a total of 7.48E8 individual DNA tracer particles. The stream was sampled twice daily over a 32-day period.

**Tracer fabrication.** Synthetic DNA was fabricated by Integrated DNA Technologies (IDT; see Table 1 for exact sequences used). Poly(lactic-co-glycolic acid) plastic (3001D from NatureWorks LLC, Minnetonka, MN, USA) was dissolved in dichloromethane and used to encapsulate the DNA tracer using a double emulsion method,24–26 as completed by McNew et al,15 Dahlke et al,14 Sharma et al13 and Soil and Water Lab Protocol.27 The double emulsion method involves dissolving PLGA in dichloromethane, adding DNA solution, and sonicating the solution for 3 iterations of 15 seconds. Polyvinyl alcohol (PVA) is then
added to the solution, and sonicated for another 3 iterations of 30 seconds. This series of sonication steps emulsifies the tracers in PLGA spheres. Poly(lactic-co-glycolic acid) is allowed to harden, and tracers were concentrated using centrifugation and lyophilized 24 hours for storage until use.

**Analysis procedure**

Before qPCR analyses, tracer spheres were lysed using dichloromethane to dissolve the PLGA polymer. Samples were vortexed and centrifuged for 1 and 5 minutes, respectively, to separate PLGA from the DNA solution. Supernatant was extracted and used for analysis in qPCR.

In experiment 1, samples were analyzed 2 days after collection using qPCR with a BioRad CFX96 Real-Time System, C1000 Thermal Cycler, and iTaq Universal SYBR Green Supermix™ from BioRad (Hercules, California, USA). Each qPCR well contained 5 µL SYBR Green Supermix, 0.29 µL each of forward and reverse primer diluted to 1 µM concentration, 0.02 µL nuclease-free water, and 4.4 µL sample. Samples were run in a 3-step protocol cycling from 95°C to 58°C and finally to 68°C to determine effective annealing temperatures of the specific DNA strand tested. The mean difference of copies between duplicates was 16.9 copies; standard deviation of the mean difference was 23.3 copies for experiment 1.

In experiment 2, samples were analyzed between 2 and 14 days after collection (variation was due to sampling time relative to collection time). All samples were stored at air temperature while in the ISCO sampler, and frozen until analysis after collection. The samples were then analyzed using qPCR as discussed in experiment 1 above. The mean difference between duplicates of the same sample was 24.6 copies for experiment 2.

All samples were normalized to the base fluorescence value of groundwater for each experiment. Duplicates of all samples were analyzed and values averaged. Extreme differences in the duplicates were attributed to contamination and false positives, and duplicates differing by more than 50 to 100 copies of DNA were compared with surrounding samples and outliers removed; 4.4 µL of sample were tested with 5.6 µL of combined BioRad SYBR Green Supermix, forward and reverse primers, and nuclease-free water.

**Standard curve.** A standard curve (Figure 2) was generated using serial dilutions from tracer stock solution to determine the approximate detection range and relationship between qPCR machine cycles to threshold fluorescence and total number of DNA molecules. The standard curve was then used to relate experimental sample qPCR cycles to DNA molecules. Standard curve $R^2$ was 0.9987. The equation relating qPCR cycles to DNA molecules obtained from laboratory experiments is

$$\log(\text{No. of Molecules}) = 23.08 - 0.667 \times Ct$$

where $Ct$ is the cycles required to pass threshold detection.

**Results**

**Experiment 1 results**

Each of the 4 sampled wells in experiment 1 shows the tracer appearance in the system in the first 9 hours (Figure 3). Time 0 coincides with the flushing of the tracers down the residence toilet. The peak values are slightly shifted for each well, as would be expected for differing distances from the leach field, potentially varied flow paths, and dispersion.

Well 0, located within the leach field, showed the highest tracer concentration and displayed 3 local maxima. The first of these maxima is before any other peak in any well (at 4.5 hours). These local maxima may coincide with different

---

**Table 1. DNA sequences used for tracer, forward, and reverse primers.**

<table>
<thead>
<tr>
<th>SEQUENCE ID</th>
<th>SEQUENCE</th>
<th>NUMBER OF BASES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tracer</td>
<td>5′-AAAGTAAAGCAGCAGAGGGTGGCACAGAGGAGGAAAGGACAGAAGAAGGAAAGAATGCTGGGAAGAGGAAGAACGCAAGGCAAAGCGGAGGTA-3′</td>
<td>87</td>
</tr>
<tr>
<td>Forward primer</td>
<td>3′-TCATTTCGTCGTCCTCCACCTGT-5′</td>
<td>22</td>
</tr>
<tr>
<td>Reverse primer</td>
<td>3′-TTCTTGTCTCCTCCATCCCT-5′</td>
<td>20</td>
</tr>
</tbody>
</table>

**Figure 2.** Standard curve relating qPCR cycles to threshold florescence output to total copies of DNA in sample for target tracer strand. The horizontal axis ($Ct$ values) indicates the number of cycles the qPCR machine takes to exponentially replicate the DNA strands to reach the threshold fluorescence value that can be detected by the optical sensors. From this analysis, the ~100 copy detection limit was established. qPCR indicates quantitative polymerase chain reaction.
volumetric water additions to the septic system, eg, additional toilet flushing, showering, and sink use.

The breakthrough curve for well 1 started at 5.5 hours, 1 hour after the first peak in well 0. The initial breakthrough at well 2 occurs around 4 hours and peaks at 5 hours. Because this peak comes rapidly after first peak at well 0 (in the leach field), it is possible that the leach field is short circuiting toward well 2. Wells 2 and 3 are downslope of well 0 suggesting a peak detection after that observed in well 0 is expected. The breakthrough curve for well 3 starts at 5 hours and peaks at 8.5 hours. This later peak is expected for well 3 as it is farthest from the leach field. Overall, the tracers show a complex flow network that would be difficult to model using traditional convective-dispersion approaches.

Experiment 2 results

In experiment 2, the DNA tracer was introduced on November 6, 2015, at 1:13 p.m. (coinciding with day 0 in the experiment) via a toilet flush in the residence.

Figure 4 shows the largest breakthrough curve occurred between 30 and 32, while a smaller peak suggests some tracers may have been detected between 10 and 20 days. Tracers numbering below 2 on the $C/C_{\text{min}}$ scale are considered instrument noise as they are indistinguishable from samples with no tracer present. It took approximately 32 days to positively detect the tracer at the sampler ~1.25 km downstream from release point. The smaller peak between days 10 and 20 may suggest that some tracers experienced faster preferential flow, though this was not consistent across all samples taken during that time.

The travel path includes movement through an in-stream pond, which increased the residence time of the stream water, as well as the dilution effects of groundwater recharge.
**Discussion**

Septic systems are designed to eventually discharge treated wastewater into the groundwater. It is therefore important to consider the time required for chemical and bacterial breakdown of contaminants in a septic system. Any conservative tracer, including these DNA-based tracers, does not necessarily fully mimic the transport and transformation of pollutants. But, tracers do provide valuable information about expected residence times, which, in the case of this study, would not be easily modeled. The premise for potentially using synthetic DNA tracers is the opportunity to simultaneously assess multiple pollutant sources (e.g., septic systems). While this study did not evaluate this specifically, it demonstrated that these tracers will move through septic systems at detectable levels at both the local (Experiment 1) and watershed (Experiment 2) scales.

Increased concentrations of previously unregulated wastewater contaminants, e.g., pharmaceutical and personal care products (PPCPs) such as over-the-counter drugs, prescription medications, or cleaning/sanitizing chemicals, are a rising concern. Yang et al. show septic systems to be a major source of these contaminants. Pharmaceutical and personal care products can pose great threat due to their continuous introduction, despite sometimes rapid degradation. Tracing septic system flow paths will be important in future assessment of these contaminants’ transport at watershed scales, which may help differentiate pollution between septic, agricultural, and urban sources.

One potential limitation of the tracers used here is that they are colloidal sized and likely move preferentially through porous media. Preferential flow paths could account for some of the variability in arrival of tracers to each of the wells. These flow paths could effectively short-circuit the leach field, allowing a shorter residence time than required for water treatment. Experiment 1 suggests differential transport of tracers and leachate can vary over small areas around the leach field, while experiment 2 suggests 2 pulses of tracers. The differences between the experiments and between wells in experiment 1 make modeling these systems more complex than traditional plug flow or continuously mixed model systems. Aside from preferential flow paths, tracers would be expected to move differentially through differing soil types. Because the study site of this experiment 1 is located exclusively within the Wayland soil group classification as outlined by the Web Soil Survey, the authors do not believe that the complexity of the results is explained by this soil characteristic.

Another challenge is establishing a reliable protocol for filtering out low-count samples. Samples in this study are normalized to the lowest concentration detected in each run. This practice identifies reliable peaks and removes background noise. This is often a challenge in environmental sampling, but it is especially challenging here because qPCR is exceptionally sensitive. There has been discussion on concentrating samples using magnetic particles in the tracers before analysis to amplify signals, though this practice has not been tested.

**Conclusions**

These 2 experiments test the applicability of using synthetic DNA tracers to infer potential septic system derived pollution transport at local levels (immediately downslope of leach fields; experiment 1) and at watershed scales (experiment 2). Experiment 1 garnered results that were more or less what one would expect although it also suggested preferential flow paths that were not anticipated. Most importantly, experiment 1 confirmed that the tracers were not critically impeded by soil filtration and, thus, this technology appears applicable to septic systems. Experiment 2 verified that the technology was applicable to larger scales, ostensibly, whole watersheds. The next generation of experiments will involve introducing multiple tracers into multiple septic systems in a densely populated watershed in Georgia.

**Acknowledgements**

Appreciation is extended to Steven Noble and the Baldwin Family for the uses of their residences as experimental sites and Katie Henderson for her assistance in sample collection.

**Author Contributions**

MTW & PLR generated the research questions studied and secured funding. PLR installed instrumentation for sample collection. PLR & CBG conducted the experiments. CBG analyzed all samples. CBG drafted manuscript, PLR and MTW provided edits.

**Data Availability Statement**

The raw data that led to the conclusions of this study will be made available, without undue reservations, to any qualified researcher upon request.

**ORCID iD**

Christine B Georgakakos https://orcid.org/0000-0003-4843-7087

**REFERENCES**