



United States Department of the Interior
U. S. GEOLOGICAL SURVEY

6480 Doubletree Avenue
Columbus, OH 43229-1111
614-430-7700
<http://oh.water.usgs.gov>

January 10, 2007

MEMORANDUM

To: Robert L. Joseph
Director, USGS Texas Water Science Center

From: Donald M. Stoeckel
Hydrologist, USGS Ohio Water Science Center

Subject: Texas Bacteria TMDL Task Force -- Report draft 12/4/2006 and meeting 12/18/2006

At your request, I've been a participant in the Texas Bacteria TMDL Task Force meetings that began in October 2006. The task force is closing in on a final product. I would like to offer the following comments related to the most recent report draft and the discussions at the meeting on 12/18/2006. As requested by the Task Force leaders, my comments are formulated as recommended text for the final report (plain text) with explanations (*italics*).

deleted USGS internal comments—DMS

Thank you for inviting me to participate in this process. The experience has been an education for me and I hope I've been able to provide useful information to the Task Force.

The statement made during the meeting of 12/18 that library-dependent MST is capable of providing quantitative allocation of fecal contamination to sources is debatable. My perspective is that quantitation by library-dependent MST is so uncertain as to be suitable for simple presence and absence categorization (or possibly major, minor, or absent categorization). The following hypothetical data set illustrates this point.

Taken at face value, the results in the following example indicate that each source contributes to each sample except, probably, wildlife in sample 4. The ARCC of 62% would be taken by many MST researchers as evidence that the library was capable of accurately classifying isolates. In the current state of the science, water-quality modelers would be tempted to take the data at face value and, for sample 1, allocate 10% of the fecal load to human sources, 52% to domestic animals, and 38% to wildlife.

Hypothetical sample data

# isolates	Total	Human	Domestic	Wildlife
Sample 1	100	10	52	38
Sample 2	100	35	46	19
Sample 3	100	15	72	13
Sample 4	100	42	56	2

Hypothetical quality-control data (compare to Wiggins et al., 2003, Stoeckel et al., 2004 and Moore et al., 2005)

# isolates	Total	Human	Domestic	Wildlife
Human (+)	100	62	12	26
Domestic (+)	100	23	58	19
Wildlife (+)	100	12	23	65

Rates of correct classification are **BOLD**
 The average rate of correct classification is 62%
 NOTE: higher accuracy than observed in studies cited

Minimum detectable percentage (MDP; calculated as in Whitlock et al., 2002 and Wiggins et al., 2003)

	Human	Domestic	Wildlife
Frequency of misclassification	18%	18%	23%
Average misclassification	19%		
Standard deviation:	3%		
MDP (Avg+4*SD):	31%		

Credible evidence of presence based on MDP

	Human	Domestic	Wildlife
Sample 1	No	Present	Present
Sample 2	Present	Present	No
Sample 3	No	Present	No
Sample 4	Present	Present	No

Proportion true identity in each class (P; bold is true positive)

# isolates	Human	Domestic	Wildlife
Human (test)	0.64	0.24	0.12
Domestic (test)	0.13	0.62	0.25
Wildlife (test)	0.24	0.17	0.59

Lower confidence limit per sample (True positive*test result)

	Human	Domestic	Wildlife
Sample 1	6	32	22
Sample 2	22	29	11
Sample 3	10	45	8
Sample 4	27	35	1

Upper confidence limit per sample (test + false negative)

	Human	Domestic	Wildlife
Sample 1	26	61	52
Sample 2	45	58	35
Sample 3	27	78	33
Sample 4	50	66	21

Credible evidence that one source contributes more than another

	Human	Domestic	Wildlife
Sample 1	<Domest	>Human	No
Sample 2	No	No	No
Sample 3	<Domest	>Others	<Domest
Sample 4	>Wildlife	>Wildlife	<Others

Refinement of interpretation to include a minimum detectable percentage (MDP) is recommended (USEPA MST guide document) to guard against false-positive results. In general, library-dependent methods have been shown to find all sources in all samples. (See the results of the Southern California Coastal Waters Research Program methods-comparison study, Journal of Water and Health, Harwood et al., Myoda et al., 2003.) The MDP calculated from the hypothetical quality-control data in this example is 31% -- in sample 1, there is not credible evidence that humans contribute at all. In fact, by this criterion, no more than two sources can be credibly depicted as “present” in any of the four hypothetical samples.

The process for bracketing percent classifications with confidence intervals has not yet been proposed in the literature. In the example, I calculated a conservative minimum confidence limit by reducing the observed values by the misclassification rate for positive-control isolates in the quality-control data. I calculated a conservative maximum limit by increasing the observed values by the number of isolates that might have been misclassified to another source. Credible evidence that one source contributes more than another was indicated if the upper and lower confidence limits for two categories did not overlap. Continuing with the example of sample 1, there was no credible difference between domestic and wildlife inputs (human was previously categorized as “no credible evidence of presence”).

I hope this example is a convincing illustration of my perspective that, in most cases, library-dependent methods cannot provide quantitative allocation of fecal contamination loads to source with sufficient certainty to be incorporated into water-quality models.

This issue is raised, in part, in the summary passage for regulatory expectations on page 28 “Alternatively, a higher number of E. coli isolates (e.g. 50) can be analyzed from fewer water samples to identify statistically significant differences in pollution sources. However, this will only provide pollution source identification on very limited time scales, and not an overall assessment of the waterbody.” The implication is that statistically significant differences may be calculated by use of library-dependent methods. This implication was stated explicitly during the meeting of December 18th.

Consider adding the following:

Although quantitative allocation of fecal contamination to source categories is a goal of most TMDL projects, uncertainty in classification limits our capacity for absolute quantitation. In some cases, library-dependent methods may enable identification of a source that contributes more fecal contamination than other sources, or identification of sources for which there is no credible evidence of substantial contamination. The results of library-dependent classification are conservatively seen as semiquantitative and suitable for sample-level classification of sources as “contribution not detected” or “contribution detected” with possible refinement to “contribution detected greater than (alternative source).” This information may not be suitable for incorporation into quantitative water-quality models.

ERIC-PCR

Consider adding two aspects to this section:

ERIC primers are used much less commonly than the BOX AIR primer in the literature; many of the early MST studies and related studies used REP primers.

Repetitive DNA elements include BOX, ERIC, and REP (reference Versolavic). Each has been used for rep-PCR in microbial source tracking studies (reference Carson, Dombek, Hassan, Stoeckel, and/or Myoda, in addition to the others).

In general, rep-PCR fingerprints are not reproducible from lab-to lab. Though the major bands in a fingerprint are generally present, the calculated similarities can be quite low. At this point, the only way to share rep-PCR fingerprints among laboratories is to use Sadowsky's HFERP or Diversilab's reagent packs and/or the Bioanalyzer. Consider the following sentence to complement the mention of the Ribotyper in the next section.

Though the rep-PCR banding patterns for a primer tend to be generally stable, minor differences between laboratories result in low between-laboratory similarity and currently limit the ability to generate a composite library in multiple laboratories. Two analytical strategies that enhance data similarity between laboratories are the use of horizontal fluorophore-enhanced rep-PCR (HFERP; Johnson et al., 2004) or a commercially packaged product such as the DiversiLab system (<http://www.bacbarcodes.com/>).

Future directions

I have comments on two passages in this section:

“More importantly, these library-independent methods can only detect a limited range of pollution sources. For example, the Bacteroidales PCR (Bernhard and Field 2000; Dick, Bernhard et al. 2005) can detect fecal pollution from ruminants, humans, horses and pigs; but no further discrimination is possible.”

Further discrimination may be possible as the field progresses – the limitation is in the number of source-associated markers that have been developed and validated thus far.

“Identification libraries consisting of thousand of isolates from different geographical regions in Texas have already been established for ERIC-PCR, PFGE, RiboPrinting, CSU and KB-ARA patterns. In addition, several thousand more E. coli isolates from source samples have been archived and are available to researchers. Library development is one of the most costly components of BST studies. It would be most economical to build upon the libraries already established in Texas. It is recommended that agencies use contractors that use BST methods that will strengthen and expand the current Texas library.”

Questions raised related to the geographical and temporal stability of library-independent markers are also relevant to libraries (as mentioned in the next paragraph of the document). The apparent advantage of having existing libraries may be not be useful in all areas of Texas, and the investment represented by existing libraries will almost certainly diminish in value over decades. Maintenance and updating the existing library with additional isolates to keep it relevant is a heavy liability.

I believe the economic tradeoff between developing and validating more source-associated markers and investing further in library development is not as clear as stated. The process of extracting the composite sample DNA and testing for markers is less costly than cultivating multiple fecal indicator isolates and typing them by molecular methods.

Also, the recommendation to use contractors and expand the current Texas library may be a reflection of opinion as much as it is a reflection of the state of the science. It will be very difficult to ensure comparability of data as multiple facilities add to the library database. Library expansion and application over larger areas and timeframes may not generate the anticipated high-quality data needed for application in TMDL efforts. Consider the following test for this paragraph:

If pursued, expansion of the current Texas library should incorporate accepted and consistent methods by experienced organizations, with substantial quality control, so data potentially can be combined into a statewide database.

References:

- Carson, C.A., Shear, B.L., Ellersieck, M.R., and Schnell, J.D., 2003, Comparison of ribotyping and repetitive extragenic palindromic-PCR for identification of fecal *Escherichia coli* from humans and animals: *Applied and Environmental Microbiology*, v. 69, no. 3, p. 1836-1839,
- Dombek, P.E., Johnson, L.K., Zimmerley, S.T., and Sadowsky, M.J., 2000, Use of repetitive DNA sequences and the PCR To differentiate *Escherichia coli* isolates from human and animal sources: *Applied and Environmental Microbiology*, v. 66, no. 6, p. 2572-2577,
- Harwood, V.J., Wiggins, B., Hagedorn, C., Ellender, R.D., Gooch, J., Kern, J., Samadpour, M., Chapman, A.C., Robinson, B.J., and Thompson, B.C., 2003, Phenotypic library-based microbial source tracking methods: efficacy in the California collaborative study. *J Water Health*, v. 1, no. 4, p. 153-166,
- Hassan, W.M., Wang, S.Y., and Ellender, R.D., 2005, Methods to increase fidelity of repetitive extragenic palindromic PCR fingerprint-based bacterial source tracking efforts: *Applied and Environmental Microbiology*, v. 71, no. 1, p. 512-518,
- Moore, D.F., Ferguson, D.M., Hannah, P., Getrich, M., Harwood, V.J., Brownell, M., and Lukasik, J., 2005, Evaluation of antibiotic resistance analysis and ribotyping for identification of faecal pollution sources in an urban watershed: *Journal of Applied Microbiology*, v. 99, no. 3, p. 618-628,
- Myoda, S.P., Carson, C.A., Fuhrmann, J.J., Hahm, B.K., Hartel, P.G., Yampara Iquise, H., Johnson, L.A., Kuntz, R.L., Nakatsu, C.H., Sadowsky, M.J., and Samadpour, M., 2003, Comparison of genotypic-based microbial source tracking methods requiring a host origin database. *J. Water Health*, v. 1, no. 4, p. 167-180,
- Stoeckel, D.M., Mathes, M.V., Hyer, K.E., Hagedorn, C., Kator, H., Lukasik, J., O'Brien, T.L., Fenger, T.W., Samadpour, M., Strickler, K.M., and Wiggins, B.A., 2004, Comparison of seven protocols to identify fecal contamination sources using *Escherichia coli*: *Environmental Science and Technology*, v. 38, no. 22, p. 6109-6117,
- Whitlock, J.E., Jones, D.T., and Harwood, V.J., 2002, Identification of the sources of fecal coliforms in an urban watershed using antibiotic resistance analysis: *Water Research*, v. 36, no. 17, p. 4273-4282,
- Wiggins, B.A., Cash, P.W., Creamer, W.S., Dart, S.E., Garcia, P.P., Gerecke, T.M., Han, J., Henry, B.L., Hoover, K.B., Johnson, E.L., Jones, K.C., McCarthy, J.G., McDonough, J.A., Mercer, S.A., Noto, M.J., Park, H., Phillips, M.S., Purner, S.M., Smith, B.M., Stevens, E.N., and Varner, A.K., 2003, Use of antibiotic resistance analysis for representativeness testing of multiwatershed libraries: *Applied and Environmental Microbiology*, v. 69, no. 6, p. 3399-3405.