New Developments in Microbial Source Tracking

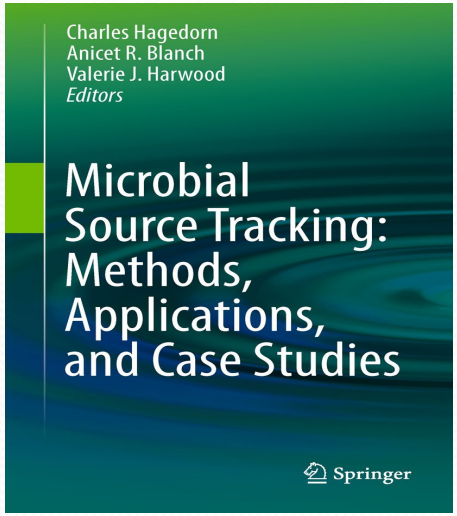
Dr. Charles Hagedorn and Dr. Brian Badgley

CH

1. Overview of MST methods and approaches
2. Advantages and disadvantages of library-based methods
3. The California source identification protocol project
4. Current method recommendations

BB

1. Interpreting new MST results
2. Integrating new MST data into TMDL requirements
3. Future directions and emerging tools



Pathogen Workshop Series Fact Sheet #4

Microbial Source Tracking

What is Microbial Source Tracking?

Microbial source tracking is a process of identifying a particular source (such as human, cattle, or bird) of fecal contamination in water, which is generally measured through fecal indicator bacteria, like *Escherichia coli* (*E.coli*) or Enterococci. The basic assumption of microbial source tracking is that there are characteristics unique to the fecal bacteria from a particular host and these characteristics allow researchers to identify the source of the contamination. Most of these target key genes that can be "fingerprinted" or tied to a type of mammal, human or bird.

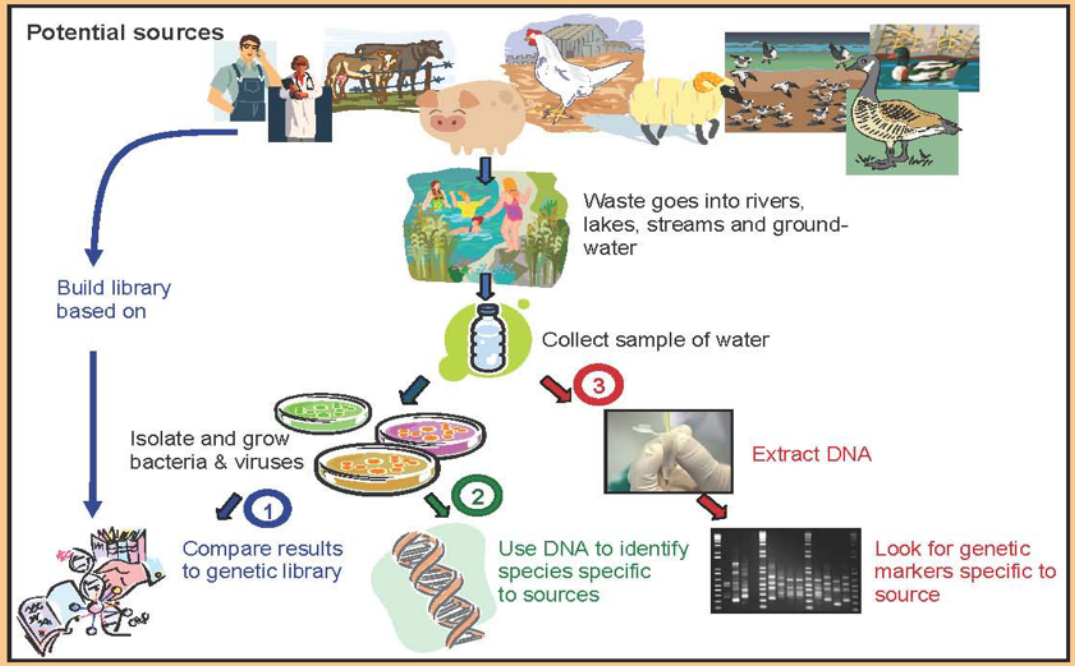
How is Microbial Source Tracking Done?

There are several different methods for microbial source tracking:

① **Library-dependent, culture based:** Samples are collected from all over a watershed and researchers grow bacteria in the lab to create a library from a variety of source organisms. Then, water samples are collected from rivers, lakes, or beaches and the bacteria in the samples are also grown in the lab. The results of the water sample are compared to the library to determine sources of contamination.

② **Library-independent, culture based:** Water samples are collected and the bacteria and viruses in the samples are grown or cultured in the lab. The bacteria and viruses grown are known to be from specific hosts or sources of fecal contamination so there is no need to compare results to a library.

③ **Library-independent, culture independent:** Water samples are collected and molecular techniques are used to isolate and identify DNA directly from the sample without first having to grow or culture the bacteria and viruses in the sample.



Library-Based Methodology - 1

Widely used in majority of MST projects from 1996 to 2005 – 2006
Been mostly replaced by newer no-library and no-culture methods

Library methods subject to both temporal and geographic limitations
- New strains appear, old strain disappear; libraries must be updated
- Geographic specificity much greater than first realized

Source ID results generated from libraries are based on statistical probability
- Works best when an unknown source category is included
- Unknown sources are typically 10% to 30% of all isolates tested

With no unknown source, library forces all isolates into a source category
- There is a bias towards fitting isolates into the largest categories

Library-Based Methodology - 2

Two large method comparison studies were undertaken in 2001 – 2004

First was by SCCWRP, based on *Enterococcus*

Second was by USGS, based on *E. coli*

These comparison studies accurately defined the problems with library approaches and both recommended that alternative methods were needed.

It was discovered that other bacterial genera (*Bacteroides*) were more source specific than the fecal indicator bacteria and the actual DNA sequences that provided such specificity could be identified.

By 2010, some 80+ source-specific DNA sequences had been published, many for the same sources, but which ones were the best?

Source Identification Method Protocol Project (SIPP) 2010-2012

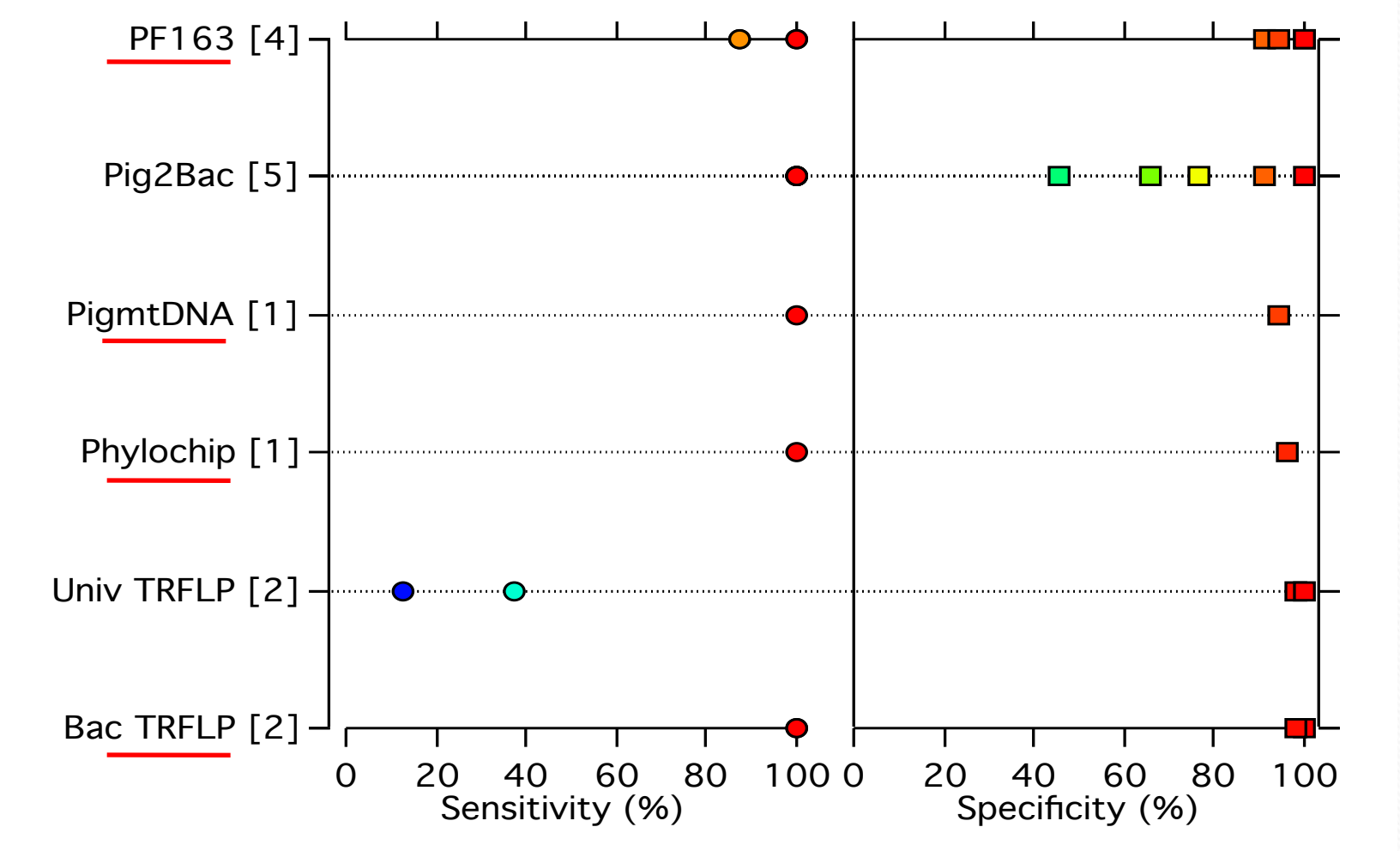
50 methods evaluated (PCR-based source-specific DNA sequences)
- 28 participating laboratories

Challenge each method with 64 blind samples
- Singletons and doubletons of 12 fecal sources

Most methods run by multiple labs
- Want to understand method repeatability

Sources included humans (individuals, sewage, septage), dogs, gulls, cattle, pigs, horses, geese, deer, pigeons, and chickens

Pig Assays

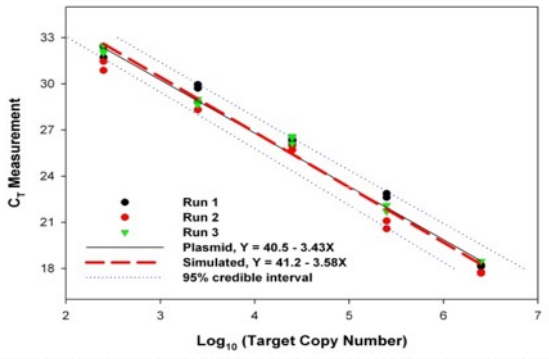

PF163 [4]
Pig2Bac [5]
PigmtDNA [1]
Phylochip [1]
Univ TRFLP [2]
Bac TRFLP [2]
0
20
40
60
80
100
Sensitivity (%)
0
20
40
60
80
100
Specificity (%)

Summary

	Human	Cow	Dog	Gull	Pig	Horse
Binary	HF183endpt, HF183SYBR	CF193 CowM2 CowM3 Rum2bac	BacCan	Gull2EndPt Gull2SYBR LeeSeaGull	PF163 mtPigDNA Phylochip Bac TRFLP	HoF597 Phylochip Bac TRFLP
Quant.	HF183Taqman BacH	BacR Rum2bac BacCow*	BacCan	LeeSeaGull	pig2bac	n.a.

A Fecal Source Identification Solution

DEFINITION… Process designed to collect, isolate, and characterize presence and/or concentration of a source identifier from an environmental sample.



The SCCWRP Guidance Manual

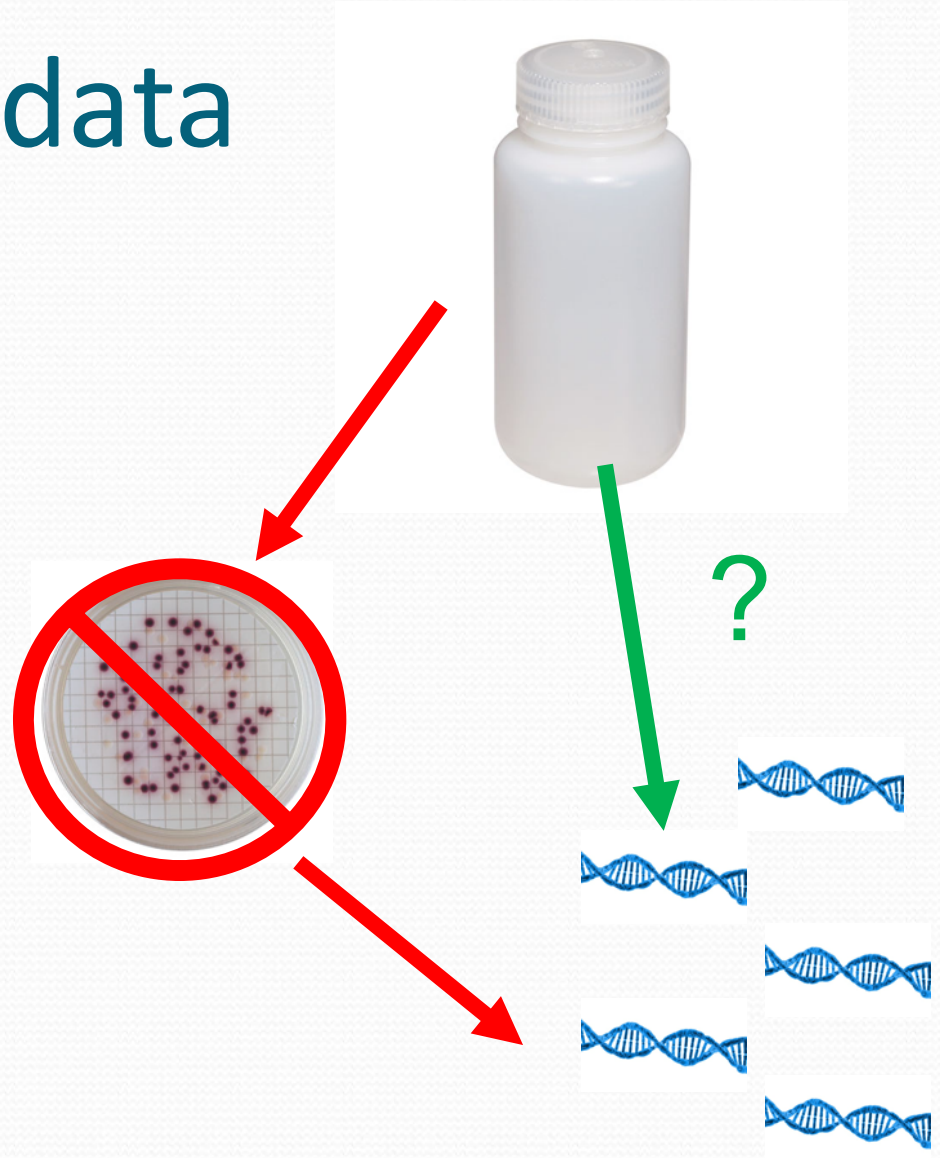
1. Conventional methods:
   - Sanitary surveys and intensive sampling
   - Dye and smoke testing
   - Robotics and CCTV cameras
   - Canine scent tracking

2. Chemical surrogates:
   - Optical brighteners in detergents
   - Fecal stanols and sterols
   - Pharmaceutical and household chemicals

3. DNA-based approaches:
   - PCR-based source-specific DNA sequences
   - Community analysis

Interpreting marker data

- So what, exactly do these data represent?
- qPCR results estimate the *number of copies of a particular genetic sequence* – a source-specific marker – in a volume of water
- It is an absolute measurement of a new variable – it is *not* based on isolation of fecal indicator bacteria



Interpreting marker data

- Presence/absence is meaningful, but the *amount* present is the most beneficial result
- On a broad scale, relative increases or decreases in marker concentration across sites or time can help establish the relative importance of different sources
- Regulatory or epidemiological standards have not been established: 3-4,000 copies/100 mL suggested recently

Interpreting marker data

- Two widely accepted uses:
  - Rule in/rule out: determine whether a source is regularly present in significant amounts
  - Relative or categorical comparisions:

<1,000 copies/100 mL
LOW

1-5,000 copies/100 mL
MODERATE

>5,000 copies/100 mL
HIGH

MST and TMDLs

- TMDLs require apportioning FIB loads to different sources
  - Library-based approaches fed into this very well

- But library-independent methods are decoupled from FIB…

500 + 2,000 + 750 ≠ 100%

Relating MST markers to FIB

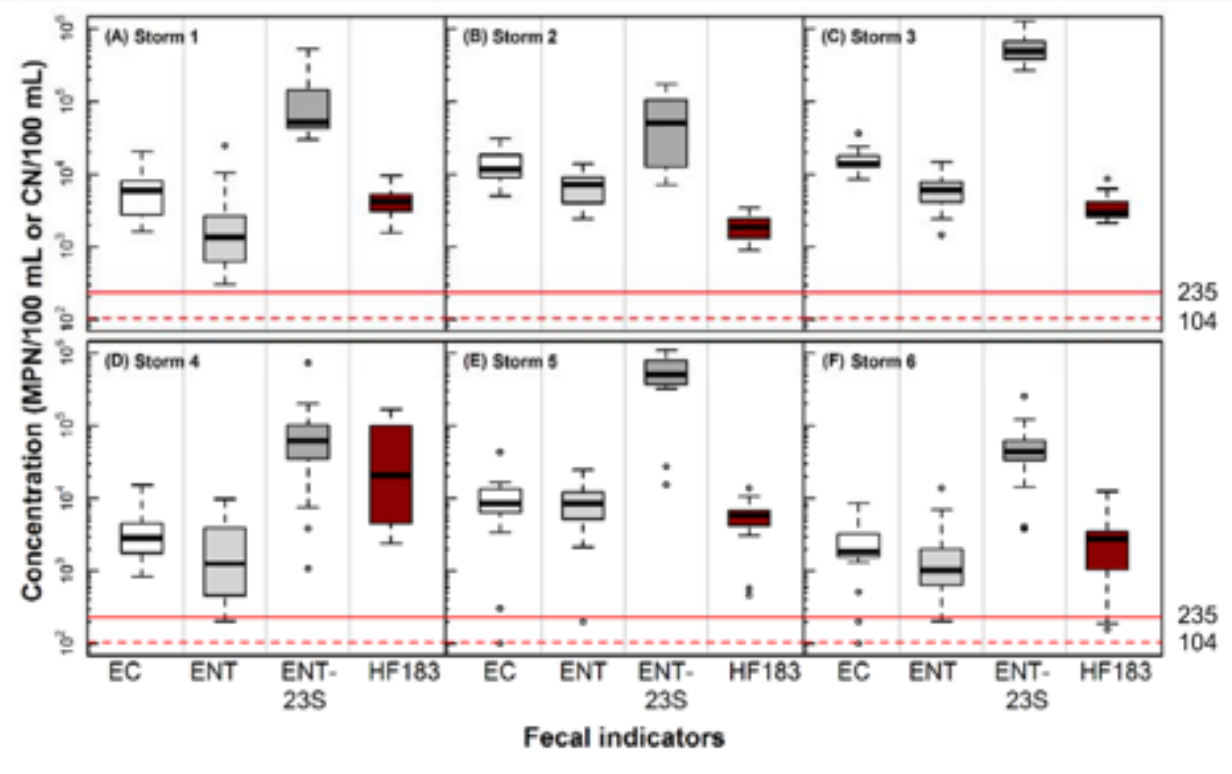
- One proposed approach is to establish ratios of MST markers to FIB in the source material

ENT : copy + *ENT* : copy + *ENT* : copy = known : total *ENT*

- Ratios could be used to predict FIB contributions of each source to establish a percentage of the total

Example from Blacksburg

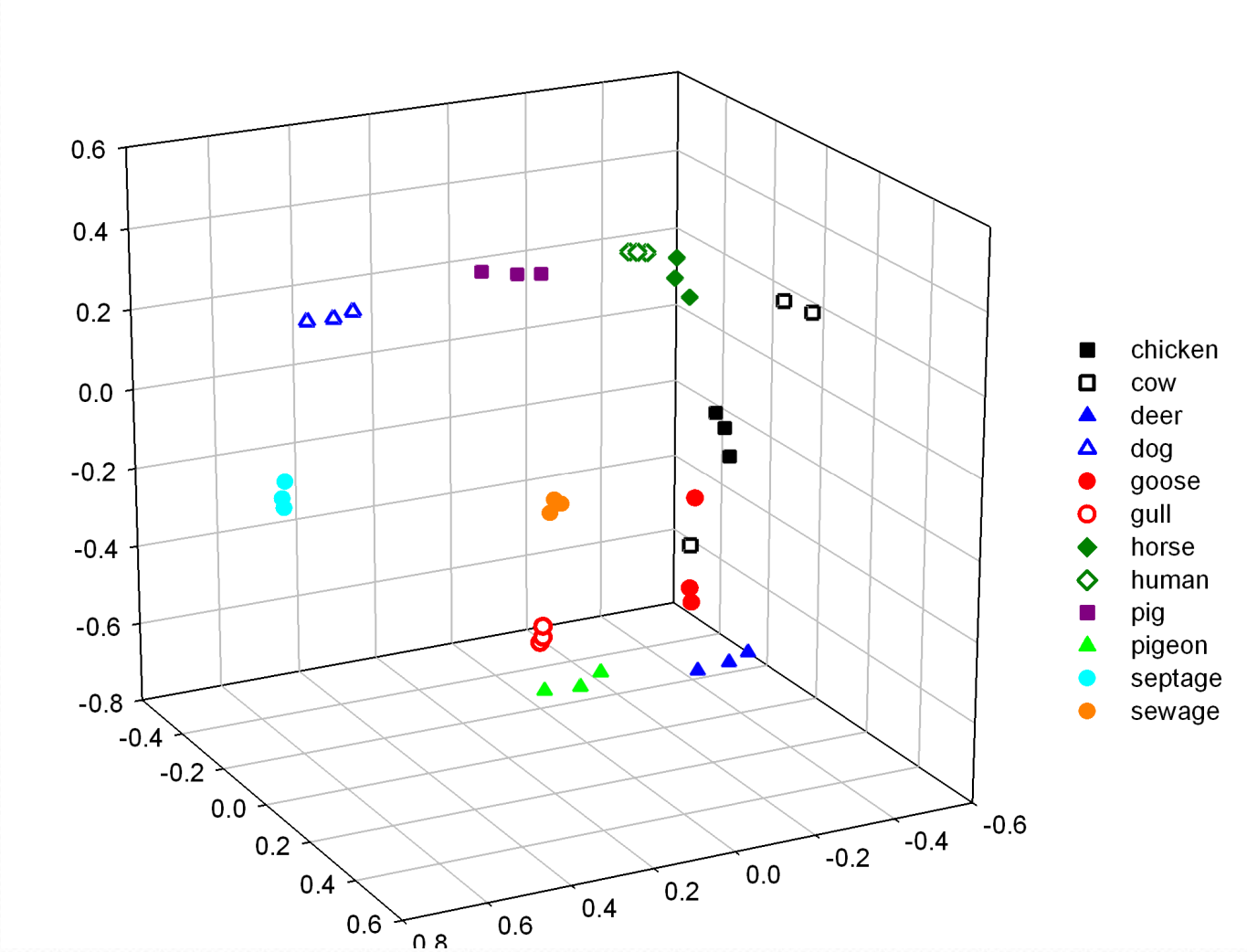
- Sewage:
  - 5 x 10^7 HF183
  - 5 x 10^6 *E. coli*
- Percent human estimated for each storm = 7%, 2%, 2%, 50%, 8%, 15%



Emerging genomic approaches

- Next-generation sequencing technology is creating a revolution in microbial ecology
- We can now identify thousands to millions of individuals, including unculturables, from environmental samples
- Unparalleled sensitivity in distinguishing similarity of microbial composition

Metagenomic approach to SIPP



Conclusions

- Library-based approaches couple with TMDL requirements, but more reliable methods are being developed.
- Library independent source-specific markers are the most promising avenue for continued improvement in MST
- Ideally the TMDL process needs to adapt to make the best use of new MST data
- Ratios have been suggested as one possibility for source allocation during the interim but have not been validated.