



Predicting fecal sources in waters with diverse pollution loads using general and molecular host-specific indicators and applying machine learning methods



Arnau Casanovas-Massana^{a,1,2}, Marta Gómez-Doñate^{a,1}, David Sánchez^b,
Lluís A. Belanche-Muñoz^b, Maite Muniesa^a, Anicet R. Blanch^{a,*}

^a Department of Microbiology, University of Barcelona, Av. Diagonal 643, Barcelona, Catalonia, Spain

^b Department of Software, Technical University of Catalonia, Jordi Girona 1-3, Barcelona, Catalonia, Spain

ARTICLE INFO

Article history:

Received 1 April 2014

Received in revised form

3 December 2014

Accepted 3 January 2015

Available online

Keywords:

Fecal pollution

Bacteria

Bacteriophages

Bifidobacterium

Bacteroides

Machine learning

Microbial source tracking

ABSTRACT

In this study we use a machine learning software (Ichnaea) to generate predictive models for water samples with different concentrations of fecal contamination (point source, moderate and low). We applied several MST methods (host-specific *Bacteroides* phages, mitochondrial DNA genetic markers, *Bifidobacterium adolescentis* and *Bifidobacterium dentium* markers, and bifidobacterial host-specific qPCR), and general indicators (*Escherichia coli*, enterococci and somatic coliphages) to evaluate the source of contamination in the samples. The results provided data to the Ichnaea software, that evaluated the performance of each method in the different scenarios and determined the source of the contamination. Almost all MST methods in this study determined correctly the origin of fecal contamination at point source and in moderate concentration samples. When the dilution of the fecal pollution increased (below 3 log₁₀ CFU *E. coli*/100 ml) some of these indicators (bifidobacterial host-specific qPCR, some mitochondrial markers or *B. dentium* marker) were not suitable because their concentrations decreased below the detection limit. Using the data from source point samples, the software Ichnaea produced models for waters with low levels of fecal pollution. These models included some MST methods, on the basis of their best performance, that were used to determine the source of pollution in this area. Regardless the methods selected, that could vary depending on the scenario, inductive machine learning methods are a promising tool in MST studies and may represent a leap forward in solving MST cases.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Fecal pollution of water poses public health risks and leads to economic losses and environmental deterioration throughout the world. Effluents from municipal and slaughterhouse wastewater treatment, combined storm-water and sewer overflows, leakage of septic systems, runoff from manure and fecal slurries deposited in fields and grazing pastures, uncontrolled discharge of fecal waste and droppings from wildlife may contaminate surface or ground

waters (Ritter et al., 2002; Simpson et al., 2002). Directly monitoring microbial pathogens is generally expensive and technically complex. In addition, pathogens are only present intermittently in water bodies and usually at low concentrations (Field and Samadpour, 2007; Savichtcheva and Okabe, 2006). Because of these limitations, water quality regulations are mainly based on the enumeration of indicator microorganisms such as total coliforms, fecal coliforms, *Escherichia coli* and enterococci. However, using these indicators does not provide information about the source of the fecal contamination. Determination of this source or sources is a key parameter to improve the management of fecal contamination at the origin, by increasing the efficiency of remediation efforts and resolving the legal responsibilities for remediation.

For this purpose, numerous microbial source tracking (MST) methods have been proposed in recent years, although various authors have reported that no single source-tracking method

* Corresponding author. Avinguda Diagonal, 643, 08028 Barcelona, Catalonia, Spain.

E-mail address: ablanch@ub.edu (A.R. Blanch).

¹ These authors contributed equally to this work.

² Present address: Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, Connecticut, USA.

correctly determines the source of fecal pollution in all scenarios (Blanch et al., 2006; Field et al., 2003; Griffith et al., 2003; Harwood et al., 2003; Moore et al., 2005; Myoda et al., 2003; Noble et al., 2003; Samadpour et al., 2005; Stoeckel et al., 2004). A particularly critical issue is the generally poor performance of most of these techniques in low fecal load matrices (Hagedorn et al., 2011). Therefore, simultaneously combining different methods might be a better approach to identifying the sources of fecal pollution (Blanch et al., 2006; Gourmelon et al., 2007; Griffith et al., 2003; Santo Domingo et al., 2007).

Machine learning is a branch of artificial intelligence concerned with the design and study of algorithms that enable machines, i.e. computers, to learn from data and, in particular, construct models based on empirical data collected from a modeled phenomenon. Thus, it could be used to develop predictive models based on a combination of several MST methods to determine the sources of fecal pollution in water (Belanche-Muñoz and Blanch, 2008; Brien and Lingireddy, 2003). To this end, an integrated open platform machine learning software program, called Ichnaea, was developed by Sánchez et al., 2011. This machine learning approach was originally designed to process a specific data matrix consisting of 103 fecal samples of human, cow, pig, poultry, horse, and sheep origin, described by 26 microbial and chemical fecal indicators obtained in a European study (Blanch et al., 2006). The machine learning method that was developed has been adapted to accept MST data matrices containing samples with fluctuating fecal concentration levels and fecal indicators with different environmental persistence. Moreover, it can be applied to different geographical and climatic areas.

In the present study, Ichnaea was used to determine the main source of fecal pollution in two real scenarios in which the water presented low levels of fecal pollution. To obtain the data required by Ichnaea for the construction of the predictive models, three general fecal indicators (*E. coli*, enterococci and somatic coliphages) were selected, along with four library-independent MST methods, namely the detection of *Bacteroides* phages specific for human, cattle, pig and poultry fecal pollution (Gómez-Doñate et al., 2011); the PCR analysis of mitochondrial DNA genetic markers associated with cattle, pig and poultry (Kortbaoui et al., 2009; Martellini et al., 2005); a multiplex PCR with *Bifidobacterium adolescentis* and *Bifidobacterium dentium* markers specific for human pollution (Bonjoch et al., 2004); and four bifidobacterial qPCRs specific for human, cattle, pig and poultry pollution (Gómez-Doñate et al., 2012). All these methods were evaluated in unique point source

samples and the results used as training matrix for the Ichnaea software. Then, the results from the low fecal pollution scenarios were introduced in the software which generated thousands of predictive models using different combinations of indicators and proposed the potential main source of fecal pollution in each environmental sample. To our knowledge, this is the first attempt to use machine learning methods to determine the origin of fecal pollution in water in a real scenario, which could represent a step forward in the solution of the MST problem.

2. Materials and methods

2.1. Study sites and sample collection

The main site of study was a system of water irrigation channels potentially exposed to fecal contamination in the Ebre Delta (Fig 1). The Ebre Delta is a natural delta with an area of 320 km² located in Catalonia (Northeastern Spain), where the Ebre River spreads out and drains into the Mediterranean Sea. It is one of the largest wetlands in Western Europe, containing several fresh and saltwater lagoons with a high diversity of bird and fish species. More than 75% of the delta drainage area is composed of agricultural fields, most of which contain rice crops. The rice cultivation is cyclic: from April to early September, fresh water flows continuously from the river through different channels and floods the rice fields. The water then flows by gravity into collecting channels that finally discharge into the sea. After the harvest in early September, the water circulation is stopped and the fields dry up through evaporation during the winter. In addition to the agricultural fields, there are some scattered cattle-grazing patches in the area under study, and some poultry/livestock facilities nearby.

One sample was collected from each of the sampling sites in May, July, September and November 2010 (Fig. 1). The first and second sites (S1, S2) were located in the main channel immediately upstream and downstream of the town of Els Muntells (500–550 inhabitants), respectively. The third sampling site (S3) was downstream from Els Muntells sewage treatment plant. Site four (S4) was located near the Els Eucaliptus complex (250 vacation apartments) just upstream from where the main channel drains into a small lagoon. Samples were taken from the lagoon itself, which was site number five (S5). Samples were collected from two more secondary channels draining into the lagoon (sites S6 and S7). Finally, water was taken from the channel that carries the water from the lagoon to the sea (site S8). Samplings were performed bimonthly

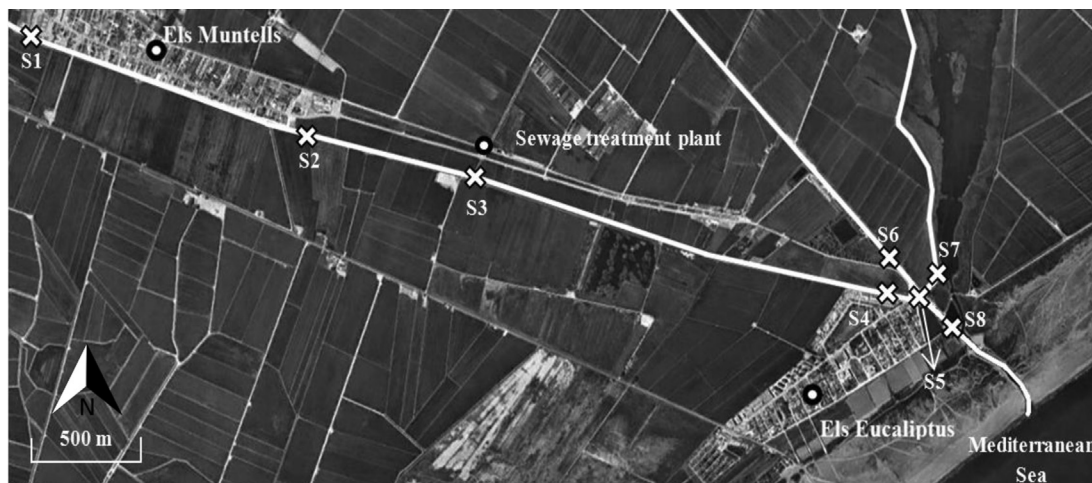


Fig. 1. Location of Ebre Delta sampling sites (S1–S8). Aerial photograph courtesy of <http://www.Earth.Google.com> and the Institut Cartogràfic de Catalunya.

Download English Version:

<https://daneshyari.com/en/article/7482944>

Download Persian Version:

<https://daneshyari.com/article/7482944>

[Daneshyari.com](https://daneshyari.com)