



Forensic drowning site inference employing mixed pyrosequencing profile of DNA barcode gene (*rbcL*)

Ting Fang¹ · Shiping Liao¹ · Xiaogang Chen¹ · Yuancun Zhao¹ · Qiang Zhu¹ · Yueyan Cao¹ · Qiuyue Wang¹ · Shu Zhang¹ · Zehua Gao¹ · Yiwen Yang¹ · Yufang Wang¹ · Ji Zhang¹ 

Received: 11 November 2018 / Accepted: 17 May 2019 / Published online: 12 June 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

The development of DNA barcoding method has given rise to a promising way of studying genetic taxonomy. Our previous study showed that pyrosequencing profile of 18S rDNA V7 hypervariable region can be used for identifying water sources without resolving the exact components of diatom colonies in water samples. In this continued study, we aimed to improve the established analysis method and to provide scientific evidence for forensic practices. A drowning animal model was set up by injecting mimic drowning fluid into the respiratory tract of the rabbit. In order to minimize the interference of animal DNA, the hypervariable region of chloroplast ribulose-1,5-bisphosphate carboxylase large unit gene (*rbcL*) was used as the pyrosequencing target region for the consistency analysis of plankton populations in tissues and water samples. After decoding the pyrosequencing profile of the targeted *rbcL* gene with the AdvISER-M-PYRO algorithm, the plankton colony that was inhaled into drowning animal lung tissue could be successfully traced back to the source of drowning fluid. Our data suggest that this method could be a reliable tool assisting forensic drowning site inference.

Keywords Forensic science · Drowning site inference · Diatom · Plankton · DNA barcode · *rbcL* · Pyrosequencing

Introduction

In forensic practice, accurate drowning site inference can shed light on criminal cases. With the development of molecular biology technology, DNA barcodes such as 18S rDNA and 16S rDNA have been used in diatom detection [1] and forensic drowning diagnosis [2, 3], as well as in inferring drowning site [4].

Ting Fang, Shiping Liao, and Xiaogang Chen contributed equally to this work.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00414-019-02075-4>) contains supplementary material, which is available to authorized users.

✉ Yufang Wang
wangyufang@scu.edu.cn

✉ Ji Zhang
zhangj@scu.edu.cn

¹ Present address: West China School of Basic Medical Sciences and Forensic Medicine, Sichuan University, Ren Min Nan Road 3-17, Chengdu 610041, China

Recently, we developed a pyrosequencing profile-based biogeographic analysis method. This novel method could be utilized to determine the consistency between diatom colonies without resolving the exact components of the colony [1]. Specifically, approximately 25 bp highly variable sequence of 18S rDNA V7 region was pyrosequenced, and the pyrosequencing profile was deciphered using the AdvISER-M-PYRO algorithm developed by Ambroise et al. [5]. The decoded profile showed high colonial specificity that could be used for consistency analysis between diatom colonies.

The 18S rDNA barcode gene is one of the basic constituents of all eukaryotic cells and can be detected in all mammalian cells. In order to apply diatom DNA barcode in forensic drowning site inference, the potential interfering effect of animal DNA should be minimized during the template preparation for pyrosequencing. The ribulose-1,5-bisphosphate carboxylase large-chain gene (*rbcL*), which exists in all phytoplankton (including diatom) but not in mammalian cells, has been widely used for molecular taxonomy of plants [6]. Therefore, *rbcL* can be used to produce the pyrosequencing profile of the plankton colonies in the tissues from drowning animal or human. In the current study, we optimized the plankton DNA extraction method for DNA barcode

analysis of animal tissue and studied the system effectiveness of this *rbcL* gene-based pyrosequencing profile decoding method.

Our results suggest that the *rbcL* gene can be used for drowning site inference with the minimal impact from animal DNA. The success rate of the drowning site inference is related to the amount of plankton contained in the drowning fluid.

Methods and materials

Diatom strains, mimic drowning fluids, and environmental water samples

Ten pure strains of different freshwater diatom species, FACHB-ZQ186 (*Gomphonema parvulum*), 163-ZLL (*Asterionella* sp.), 180-ZLL (*Melosira* sp.), FACHB-221 (*Hantzschia palea*), FACHB-1034 (*Melosira varians*), FACHB-1051 (*Nitzschia palea*), FACHB-1054 (*Navicula* sp.), FACHB-1631 (*Cyclotella*), FACHB-1665 (*Fragilaria* sp.), and FACHB-1682 (*Synedra* sp.), were obtained from the Freshwater Algae Culture Collection at the Institute of Hydrobiology (FACHB), Chinese Academy of Science. These diatoms were cultured with CSi medium [7] in a temperature-controlled illumination incubator at 25 ± 2 °C with a 12-h light/12-h dark light cycle, and the light intensity was maintained at 2000 lx. The cultures were refreshed by replacing half of the culture with fresh CSi medium every 3 weeks.

These cultured diatoms were then used to construct a series of mimic drowning fluids based on diatom density as defined by Williams [8]. Due to the exponential growth of diatom cells, Williams divided the diatom density into nine levels based on the amount of diatom individuals per genus per milliliter in rivers to assess water quality [8]. In this study, we chose level 6 to level 3 and redefined them as class 4 to class 1. In summary, classes 4, 3, 2, and 1 contain 7200, 3600, 1800, and 900 diatom individuals per genus per milliliter, respectively. Since the classes reflect the diatom amount in the mimic fluids, they were used to analyze the sensitivity of the system. To detect the distinguishing ability of the pyrosequencing target, which is the hypervariable region fragment of the *rbcL* gene, the ten strains of different diatom species were divided into three groups to mimic natural water samples. The G1 group contains five diatom species: *Synedra* sp., *Gomphonema parvulum*, *Nitzschia palea*, *Fragilaria* sp., and *Navicula* sp. The G2 group contains five other species: *Asterionella* sp., *Melosira varians*, *Melosira* sp., *Cyclotella*, and *Hantzschia palea*. The G3 group contains all ten of these species. Additionally, the amount of *Synedra* sp., *Nitzschia palea*, *Melosira* sp., and *Hantzschia palea* in the G3 group was 1800 diatom

individuals per genus per milliliter, while the amount of the other six diatoms was 900 diatom individuals per genus per milliliter (Fig. 1).

Five environmental water samples were collected from the same sites that we described previously [1]. All of these sites were highlighted on the map (Supplementary Fig. 1). At each site, 550 mL water sample was collected with a nuclease-free plastic bottle and was stored at 4 °C. All water samples were used within 24 h.

Primer design, specificity, and theoretical coverage analysis of the primers

In this study, two sets of primers were designed: PCR primers for obtaining sequencing templates and a sequencing primer for pyrosequencing reaction. To identify the hypervariable region and the relatively conserved sequence in *rbcL* gene, all *rbcL* gene sequences of diatom species were retrieved from NCBI (<https://www.ncbi.nlm.nih.gov/gene/?Ter=diatomea+rbcL>), then were aligned and compared using the MUSCLE program in the MEGA 6 software [9]. As shown in Supplementary Fig. 2, the PCR primers and the sequencing primer were designed in the relatively conserved region on both sides of hypervariable sequence. Primers were designed with the PyroMark Assay Design software (version 1.0.6, Biotage AB). The primer sequences were listed in Table 1, and the biotinylated primer was labeled. The nucleotide dispensation order of the pyrosequencing reaction (GAGCGACAGTCTGATCAGATAGTGACGTACATGTATACGT) was generated using the “senator” function of the AdvISER-M-PYRO algorithm.

To confirm that the designed primers could amplify the *rbcL* gene of phytoplankton including diatomea, we performed a specificity analysis by using the Nucleotide Blast algorithm (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) of NCBI website. Both forward and reverse PCR primers were used as the query sequence. To improve the accuracy of specificity analysis, the diatomea and other 12 common phytoplankton species, which include *Cyanobacteria* (blue-green algae), *Glaucophyta*, *Prasinophyta*, *Chlorophyta* (green algae), *Euglenophyta*, *Cryptophyta*, *Raphidophyta*, *Xanthophyta* (yellow-green algae), *Eustigmatophyta*, *Chrysophyta* (golden algae), *Haptophyta*, and *Dinophyta* [10], were used to define the “organism,” an optional parameter of the BLAST program. Then, all of the *rbcL* sequences of the phytoplankton species, of which the sequences could match with PCR primers, were downloaded from the NCBI database for theoretical coverage analysis. The downloaded sequences were aligned and analyzed with the MEGA 6 software. The theoretical coverage was calculated by counting the subspecies sequences that contain the primers. For each phytoplankton species, the coverage equals to the amount of

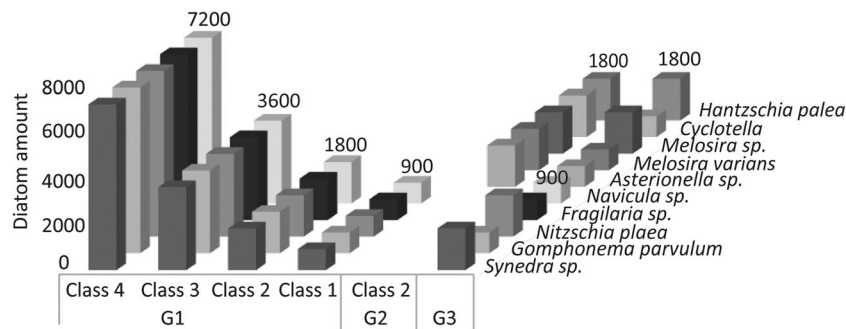


Fig. 1 Composition of the mimic drowning fluids. The number above the bar represents the diatom amount per genus per milliliter in each class. The z-axis is the diatom genus of each group. Groups consist of different diatom species. Classes are defined based on the amount of diatom used in groups. The G1 group contains five diatom species: *Synedra* sp., *Gomphonema parvulum*, *Nitzschia palea*, *Fragilaria* sp., and *Navicula* sp. The G2 group contains other five species: *Asterionella* sp., *Melosira*

variens, *Melosira* sp., *Cyclotella*, and *Hantzschia palea*. The G3 group contains all ten of these species. In classes 1, 2, 3, and 4, the diatom amounts are 900, 1800, 3600, and 7200 diatom individuals per genus per milliliter, respectively. In the G3 group, the amount of *Synedra* sp., *Nitzschia palea*, *Melosira* sp., and *Hantzschia palea* is 1800 diatom individuals per genus per milliliter, while the amount of the other six diatoms is 900 diatom individuals per genus per milliliter

sub-species containing the primers divided by the amount of all sub-species of the given phytoplankton species.

Establishment of an animal drowning model

To obtain adequate tissue and drowning fluids, New Zealand rabbits were used to establish the drowning model. The rabbits were deeply anesthetized by intravenous injections of 1.5% sodium pentobarbital (2 mL/kg). After tracheotomy, the “drowning fluid” was trickled into the rabbit trachea until respiration ceased [11]. In order to avoid loss of the fine froth produced during liquid inhalation, the trachea was ligated by a surgical suture before dissecting the thoracic cavity from the lower jaw to the diaphragm. The lung surface was rinsed with ultrapure water that was filtered with a 0.22-µm syringe filter. Then, the trachea was cut above the ligation point and lifted with hemostatic forceps, and the intact lung was dissociated. Approximately 200 mg of lung tissue near the hilum was cut and placed into a 2-mL microcentrifuge tube for further DNA extraction. All tissue samples were kept at -20 °C until use. This protocol was approved by the Ethics Committee of Sichuan University (No. K2016007, approved on March 1, 2016) and was in accordance with Guiding Recommendations on the

Treatment of Experimental Animals issued by the Ministry of Science and Technology of the People’s Republic of China (<http://www.most.gov.cn/fggw/zfwj/zfwj2006/200609/t2006093054389.htm>).

Animal tissue digestion and plankton DNA extraction

DNA of the drowning lung tissue was extracted by the proteinase K digestion method without pre-isolation of the plankton. Briefly, a 200-mg tissue was put into a 2-mL microcentrifuge tube with a pre-cooled stainless steel bead (5 mm diameter, Qiagen, Germany), and the tube was shaken at 25 Hz for 10–20 min in the TissueLyser II (QIAGEN, Germany) until all tissues were minced. To ensure all lung tissues were completely digested, proteinase K with a final concentration of 8 mg/mL was used. Tissue digestion was performed at 56 °C overnight. Then, DNA was separated with a commercially available DNA extraction kit (TIANamp Genomic DNA Kit, Tiangen Biotech, China) according to the manufacturer’s protocol. The DNA was quantified by a NANODROP 1000 spectrophotometer (Thermo Fisher Scientific, USA) and was then diluted with nuclease-free water to 100–500 ng/µL, kept at -20 °C until use. The DNA of the mimic drowning fluids and environmental water samples were extracted by the same method described above.

Table 1 The sequence of amplification primers and sequencing primer

Primer	Name	Sequence
Forward primer	rbcl-F	5'-GGTAAATTAGAAGGTGATCCTTTA
Reverse primer*	rbcl-R	5'-Biotin-GTRCCACCACCRAATTGTA
Sequencing primer	rbcl-S	5'-CCTTTAATGATTAAAGGTTTCT

*R in the midst of the reverse primer indicates that the nucleotide at this position is adenine or guanine

PCR amplification, pyrosequencing, and decoding the pyrosequencing profile

PCR was performed in a total reaction volume of 50 µL containing 2 µL DNA template (200–1000 ng), 25 µL GoTaq® Colorless Master Mix (Promega, USA), 0.2 µM primers, and 21 µL nuclease-free water. PCR was conducted with initial denaturation at 95 °C for 2 min, 40 cycles of

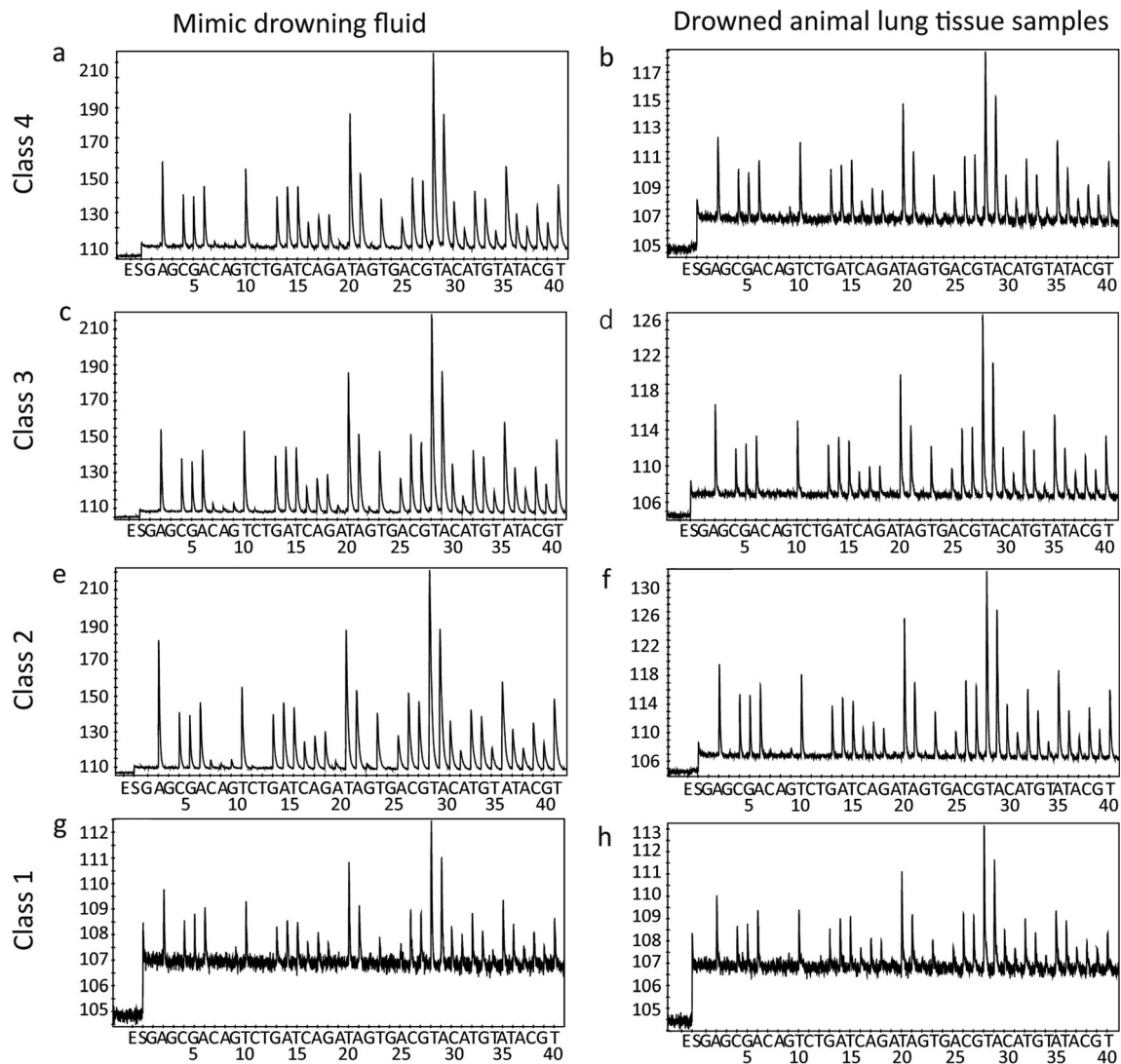


Fig. 2 Pyrosequencing profiles with different diatom amounts. Four classes with the same diatom constituents but in different amounts were used. **a, c, e,** and **g** are the pyrosequencing profiles of mimic drowning fluids in different classes. **b, d, f,** and **h** are the profiles of lung tissue samples extracted from animals drowned with the corresponding mimic drowning fluids. All of the profiles have similar peaks. Class 1, which has

main amplification (95 °C for 30 s, 50 °C for 30 s, 72 °C for 30 s), and final extension at 72 °C for 5 min.

Table 2 Contribution and correlation coefficients of drowned animal tissue samples from class 1 to class 4

Testing samples	Dictionary			<i>r</i>
	G1 (class 2)	G2 (class 2)	G3	
Class 4	16.24	0	0	0.994
Class 3	8.3	0	0	0.993
Class 2	9.56	0	2.08	0.994
Class 1	–	–	–	–

900 diatom individuals per genus per milliliter, showed a more significant baseline and lower peak height than classes 2–4 (which have 1800, 3600, and 7200 diatom individuals per genus per milliliter, respectively). The profiles of lung tissue (**b, d, f,** and **h**) showed more unstable baselines than the profiles of the homologous mimic drowning fluids (**a, c, e,** and **g**)

Pyrosequencing was conducted according to the manufacturer's instructions (PyroMark® Q96 ID, Biotage, Sweden).

To improve the sensitivity of the system, two-step PCR amplification was applied to increase the amount of pyrosequencing templates. For the first amplification, PCR was performed in a total reaction volume of 20 µL containing 1-µL DNA sample, 10-µL GoTaq® Colorless Master Mix (Promega, USA), 0.25-µM primers, and 8-µL nuclease-free water. PCR was conducted with initial denaturation at 95 °C for 2 min, 28 cycles of main amplification (95 °C for 30 s, 50 °C for 30 s, 72 °C for 30 s), and final extension at 72 °C for 5 min. The second amplification was in 50 µL; a 2-µL product of

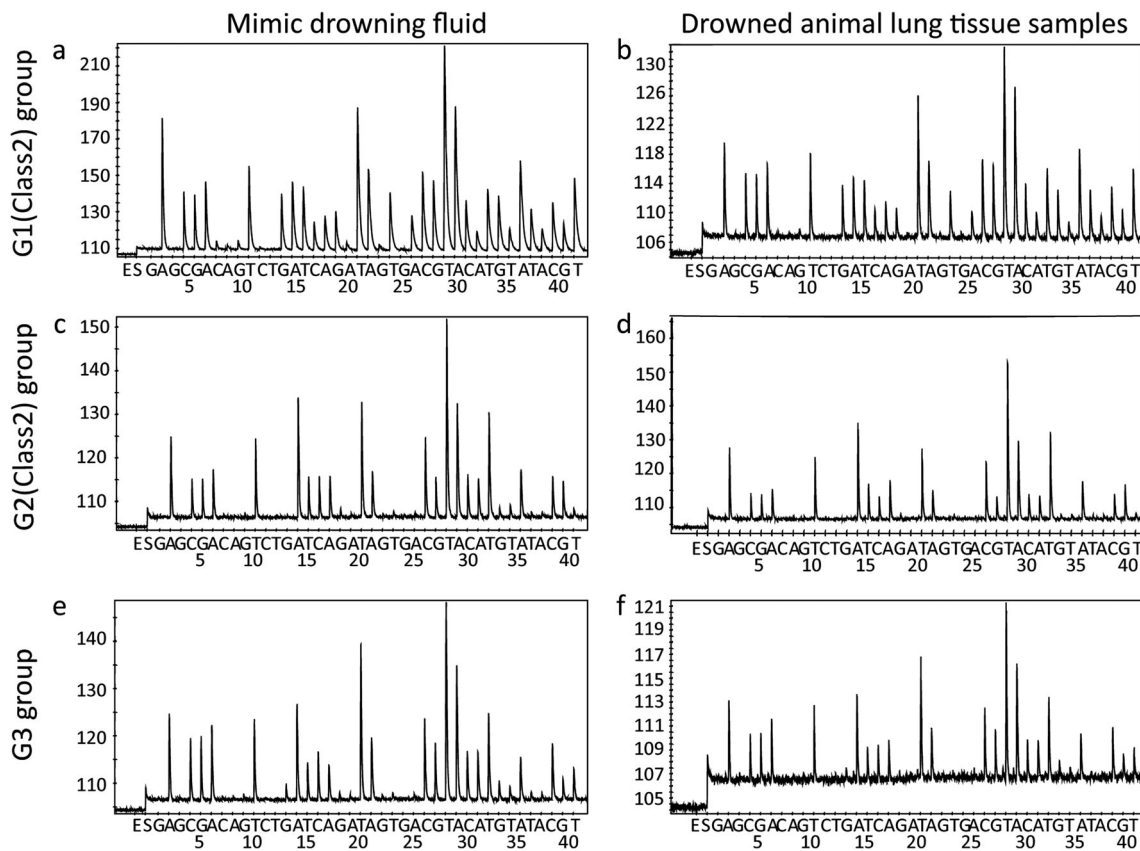


Fig. 3 Pyrosequencing profiles of different diatom components. Three groups showed different pyrosequencing profiles, but the lung tissue samples showed similar profiles with the corresponding mimic drowning fluid that was used to prepare the drowning animal model

the first step was used as the DNA template, and the same PCR conditions were used as the first step.

The decoding method was described in our preliminary study [1]. The first step was to create a standardized learning dictionary from the training dataset. For mimic drowning fluids in the present study, the dictionary was constructed with the peak height data of the G1 group in class 2, the G2 group, and the G3 group. For environmental water samples, the peak height data of each environment water sample was used as the data source for dictionary construction. The dictionary was then assigned to the “dictionary” parameter of AdvISER-M-PYRO, while the pyrosequencing peak

height data of drowning lung tissue samples was assigned to the “signal test” parameter. After running AdvISER-M-PYRO, the output was in the form of contribution coefficient and correlation coefficient.

Results

Specificity and theoretical coverage of the PCR primers

The PCR primer specificity analysis showed that the designed primers matched with the sequences of three phytoplanktons, including *Diatomea*, *Dinophyta*, and *Xanthophyta* (yellow-green algae). It was confirmed that our PCR primers were at least capable of amplifying the *rbcL* gene of the three phytoplankton species. The coverage analyzing results showed that 75.54% diatomea *rbcL* sequences (2844 matches out of eligible 3765 diatomea sequences documented in the NCBI database) were covered by our PCR primers. These primers also covered 8.62% (15 out of 174) *Dinophyta* and 4.87% (19 out of 390) *Xanthophyta* (yellow-green algae)

Table 3 Contribution and correlation coefficients of drowned animal tissue samples of groups G1, G2, and G3

Testing samples	Dictionary			<i>r</i>
	G1 (class 2)	G2 (class 2)	G3	
G1 (class 2)	9.56	0	2.08	0.994
G2 (class 2)	0	17.86	0	0.988
G3	0	1.76	4.49	0.997

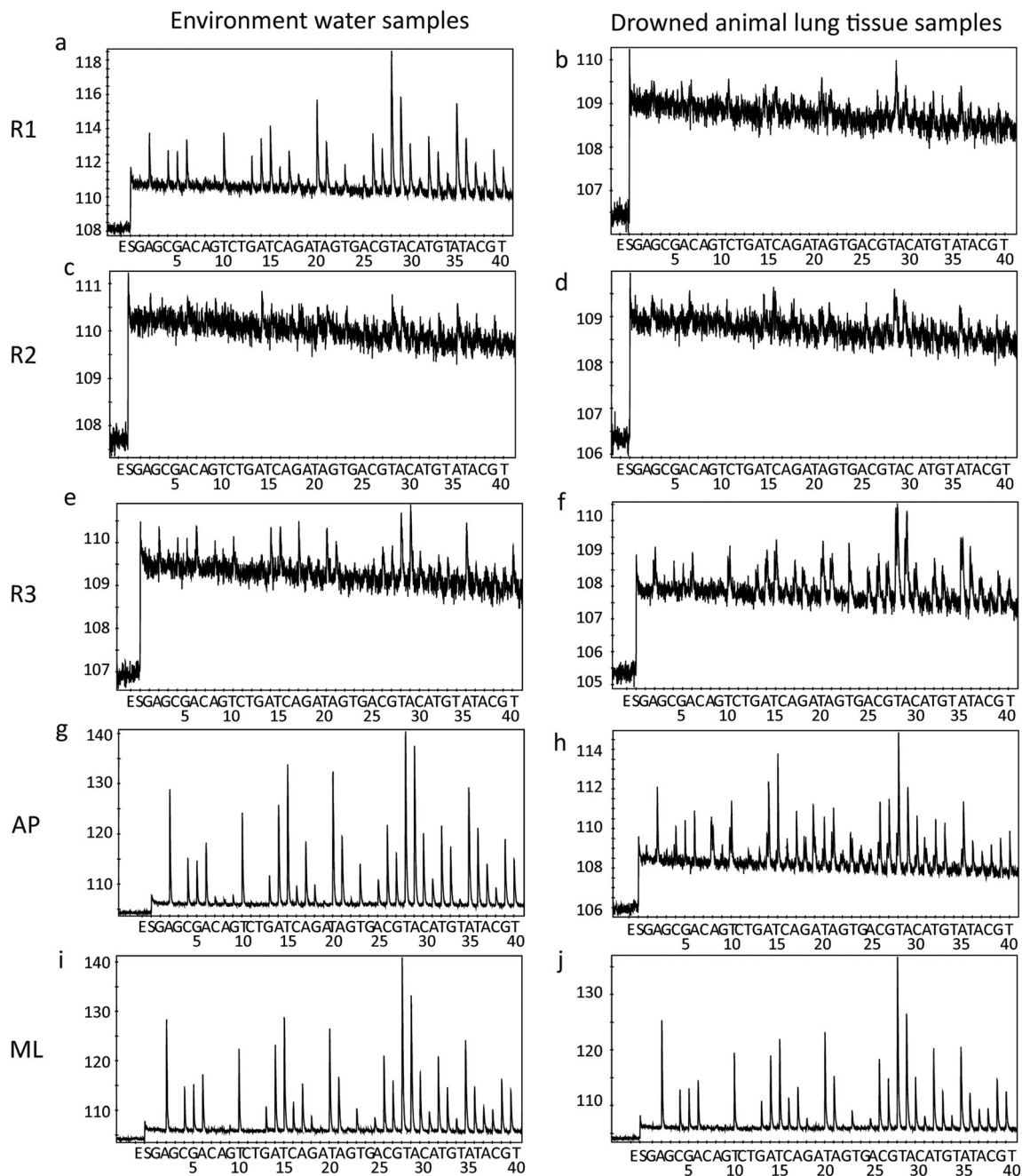


Fig. 4 Pyrosequencing profiles of environmental water samples with one-step PCR. R1-3 were collected from the Jiang'an River; ML were collected from Mingyuan Lake; AP were collected from an artificial pool closely connected with the Jiang'an River. **a**, **c**, **e**, **g**, and **i** were the

pyrosequencing profiles of the five water samples, and **b**, **d**, **f**, **h**, and **j** were the profiles of lung tissue samples extracted from animals drowned by the corresponding environmental water samples

species (Supplementary Fig. 3a). Among the 2878 sequences covered by the primers, diatom sequences accounted for 98.82%, which was far more than sequences of *Dinophyta* (0.52%) and *Xanthophyta* (yellow-green algae, 0.66%) (Supplementary Fig. 3b). Of note, all sequences of Chloroplastida and Opisthokonta clades, including Mammalia and Fungi, were not covered by these primers.

Sensitivity analysis with known diatom genera

To determine the detection limit of the system, a series of mimic drowning fluids at different classes constructed by combining different amounts of five diatom genera (Fig. 1) were analyzed. As shown in Fig. 2, the pyrosequencing profiles of the drowning lung tissue samples were consistent with the profiles of the corresponding mimic drowning fluids used

Table 4 Contribution and correlation coefficients of drowned animals with environmental water samples with one-step PCR

Testing samples	Dictionary					<i>r</i>
	R1	R2	R3	AP	ML	
R1	–	–	–	–	–	–
R2	–	–	–	–	–	–
R3	–	–	–	–	–	–
AP	–	–	–	–	–	–
ML	0	0	0	0	11.26	0.995

R1-3, Jiang'an River; ML, Mingyuan Lake; AP, artificial pond

to construct the drowning rabbit model at all classes. When the diatom amounts decreased from class 4 to class 1, the peak heights of both mimic drowning fluid and tissue samples showed a decreasing trend. In class 1, some peaks were difficult to distinguish from the baseline (Fig. 2g, h). After being decoded by AdvISER-M-PYRO (Table 2), the drowning tissue samples from class 4 to class 2 were matched with their corresponding mimic drowning fluids, as the diatom amounts were from 1800 to 7200 individuals per genus per milliliter. There was no valid result of the drowning animal lung tissue sample for class 1, which contains 900 individuals per genus per milliliter.

The distinguishing ability of mimic drowning fluids containing different diatom constituents was also analyzed. As shown in Fig. 3, despite of slight differences, the drowning animal lung tissue of G1, G2, and G3 were consistent with the pyrosequencing profiles of the corresponding mimic drowning fluids. After decoding the peak height data, the drowning lung tissue samples were matched with the corresponding mimic drowning fluids (Table 3). The correlation coefficients (*r*) of G1, G2, and G3 were 0.994, 0.988, and 0.997, respectively.

Evaluation of the detection system with environmental water samples

To evaluate the feasibility to apply the method in a real drowning case, we examined the water samples from five sites (specified in the supplementary Fig. 1; namely R1-3, AP, and ML). We firstly compared the profiles obtained from the one-step PCR products. As shown in Fig. 4, for original environmental water samples (left column), AP (Fig. 4g) and ML (Fig. 4i) showed a stable baseline, but there were no analyzable peaks for sample R1 (Fig. 4a), R2 (Fig. 4c), and R3 (Fig. 4e). For samples extracted from drowning lung (right column), while there were no analyzable peaks for R1, R2, R3, and AP (Fig. 4b–h), ML showed with a stable baseline and consistent pyrosequencing profile in original environmental samples (Fig. 4j). After AdvISER-M-PYRO decoding, diatoms in the ML's

tissue sample were traced back to the corresponding water sample successfully, with the contribution coefficient value of 11.26 and *r* value of 0.995 (Table 4). It is suggested that robustness of profiles from one-step PCR products was limited and it could not generate valid results for samples from a variety of sources.

To address the unsatisfactory profiling results from the first-step PCR, we further examined the pyrosequencing profiles with the two-step amplification products. As shown in Fig. 5, the final profiles were significantly improved. The profiles of AP and ML samples showed consistent profiles between original environmental samples and drowning lung samples (Fig. 5g and h; i and j). Importantly, analyzable profiles were obtained for R1, R2, and R3 in both original environmental samples and drowning lung samples (Fig. 5a and b; c and d; e and f). After decoding the peak height data, the R1 and AP groups were identified with contribution coefficients of 17.75 and 36.25, respectively. The ML and R3 groups were correctly recognized with a dominant contribution coefficient of 19.81 and 11.2, respectively. However, the R2 group was wrongly identified as a combination of R1, AP, and ML (Table 5). In repeated tests (Supplementary Table 1), the AP and ML groups were identified correctly, R3 and R1 were correctly identified only once in a triplicated testing. It was suggested that pyrosequencing profiles with the two-step amplification products could successfully track the original sources of drowning lung extracts in most cases.

Discussion

In the previous study, we described a biogeographic analysis method using the pyrosequencing profile decoding system. With such method, we could successfully differentiate water samples by comparing diatom colonies in the water samples [1]. In actual drowning cases, forensic scientists have to enrich diatoms from the human tissue for morphology test and DNA analysis. During the process of enrichment, it is hard to avoid DNA loss, which in turn affects the consistency analysis of plankton groups based on DNA barcodes. In this study, one-step proteinase K digestion method was used to preserve the plankton DNA as much as possible, instead of using cetyltrimethylammonium bromide (CTAB) method in the previous study [1]. The DNA extracted by the proteinase K method is a mixture of planktonic DNA and mammalian DNA, and the latter will cause advance amplification if the targeted DNA barcode gene is not specific enough for plankton. To minimize the potential interfering impact of mammalian DNA on plankton DNA analysis, the *rbcL* gene was chosen as the target for pyrosequencing analysis in the drowning animal model.

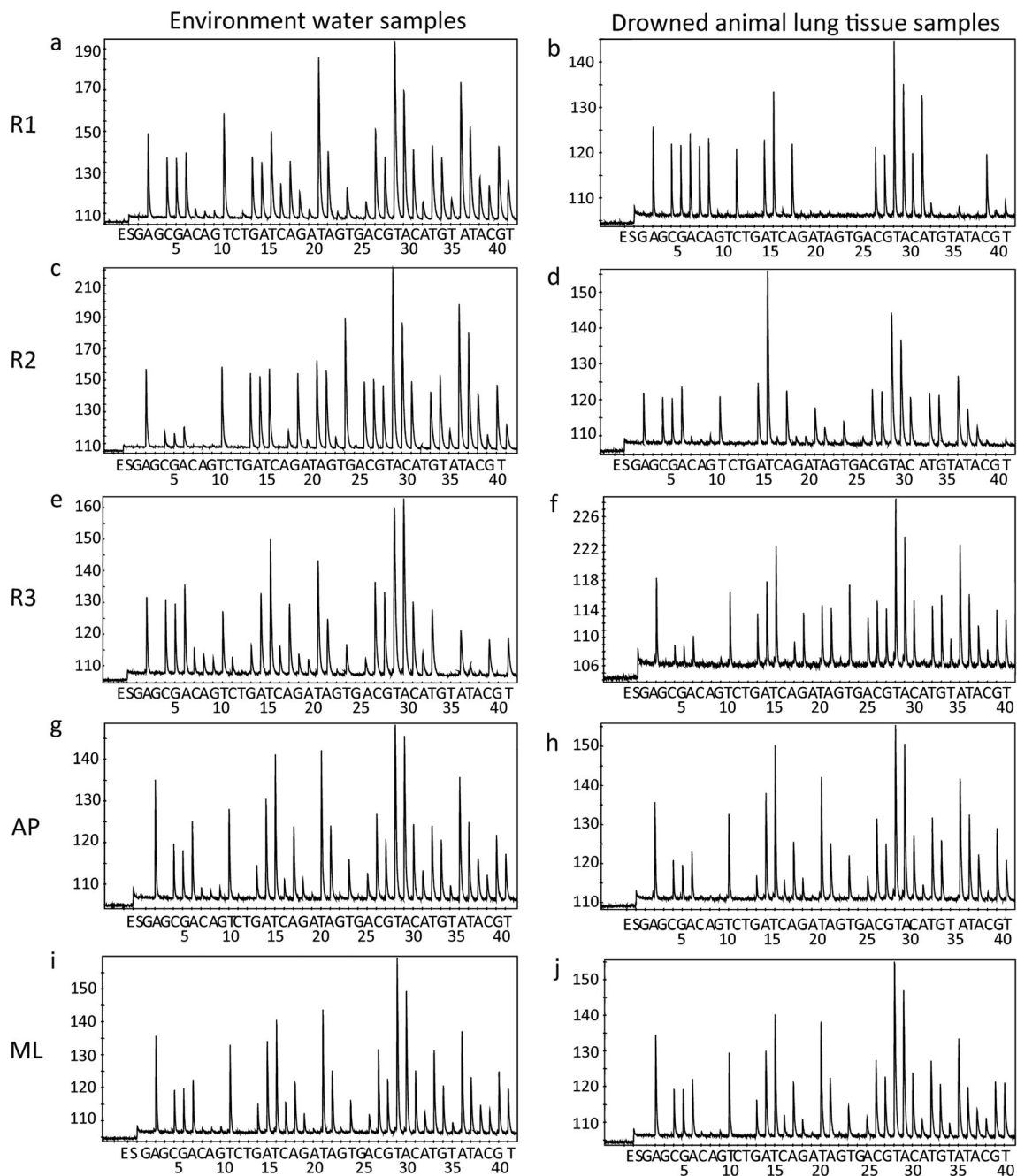


Fig. 5 Pyroseq, the feasibility of the method was verified from sequencing profiles of environmental water samples with two-step PCR. **a, c, e, g, and i** were the pyrosequencing profiles of environmental water samples of the five sites. **b, d, f, h, and j** were the profiles of lung

tissue samples extracted from animals drowned with the corresponding environmental water samples. All the samples were amplified with two-step PCR

Specificity of the *rbcl* gene

The *rbcl* gene is known to be unique in plants and other organisms that are capable of photosynthesis, including diatoms. To reduce the interference of rabbit DNA during amplification, the *rbcl* gene was selected to specifically amplify the diatom DNA from the rabbit-diatom DNA mixture.

As it is a chloroplast-specific gene, fungal and mammalian DNA cannot be amplified.

NCBI-based analysis suggested that the primer pair covered 75.54% of the diatom sequences, as well as few *Dinophyta* and *Xanthophyta*. Limited by the database size, the primer-specific analysis and coverage calculation might be biased. Because the aim of our study was to distinguish

Table 5 Contribution and correlation coefficients of drowned animals with environmental water samples with two-step PCR

Testing samples	Dictionary					<i>r</i>
	R1	R2	R3	AP	ML	
R1	17.75	0	0	0	0	0.603
R2	2.64	0	0	7.97	11.22	0.855
R3	0	2.98	11.2	0	0	0.816
AP	0	0	0	36.25	0	0.989
ML	0	0	2.69	8.15	19.81	0.996

R1-3, Jiang'an River; ML, Mingyuan Lake; AP, artificial pond

the drowning site rather than to identify the constituents of plankton species (including diatomea) in the water, such bias should not affect the system performance of our analysis method for the final goal. Moreover, the participation of the algae other than diatomea could expand the differences between drowning sites, since they can increase the diversity of the hypervariable sequences that were used as the core of this method.

Digestion method of drowning animal tissue

The strong acid digestion method is traditionally used to separate diatoms from animal tissue, but this method damages the organic matter, leaving the remaining diatom silicon frustules as the only feature for morphological observation [12–14]. For the DNA barcode analysis for the drowning animal tissues, the trypsin method was applied to remove tissue cell as much as possible, as described as Yu et al. [15]. Theoretically, the trypsin method could dissociate lung tissues and release the cells. This should be beneficial for diatom enrichment. We tried the method in the same way to digest the lung tissue mixed with nine diatom species (28,800 diatoms per genus per milliliter; 180-ZLL, 163-ZLL, FACHB-1682, FACHB-1631, FACHB-1034, FACHB-ZQ186, FACHB-1051, FACHB-1665, FACHB-1054), but only one diatom (FACHB-1034) DNA was amplified (Supplementary Fig. 4). This result suggests that the diatom cells might be damaged by trypsin due to improper digestion time or other uncontrollable factors, resulting in diatom DNA losses after discarding the supernatant. Conversely, the one-step proteinase K digestion method can maximize DNA retention. The consistent pyrosequencing profiles obtained between water samples and drowning lung tissue samples suggest that proteinase K digestion is an ideal tissue treatment protocol for our established analysis method. However, incomplete digestion would influence the yield of the final DNA concentration. In the present study, 200 mg of lung tissue was digested completely with the final concentration of proteinase K up to 8 mg/mL. Meanwhile, in this study, 200 mg of lung tissue near the hilus was enough for DNA

barcode analysis. However, in actual forensic practice, more tissue, if available, is recommended to guarantee satisfactory results.

System performance

In the current study, the feasibility of the method was verified from three aspects. First, the detection limit of the system was examined using diatom colonies with known density, the accurate source of diatoms in the drowning animal lung tissue could be identified correctly after one-step amplification, even when the diatom growth density was down to 1800 individuals per genus per milliliter (Table 2). Second, we determined the distinguishing ability to mimic drowning fluids with known but different diatom compositions. As shown in Table 3, after AdvISER-M-PYRO analyzing, the tissue samples of the animal model constructed with G1, G2, and G3 were correctly inferred to their corresponding mimic drowning fluids. Third, the method was tested in animal model with actual environmental samples mimicking forensic practice. Since it is difficult to obtain human tissue from real forensic drowning cases, five environmental water samples were used to mimic such situation. With one-step PCR, only one tissue sample (ML) was correctly traced to its original environmental water sample. Such low success rate was due to the low quantity of plankton, including diatoms in these water samples. As shown in Supplementary Fig. 5, both of the morphological detection and qPCR results suggest that the ML water sample has higher amount of diatoms and more DNA templates for PCR amplification of *rbcL* gene. ML is an artificial lake beside the Jiang'an River (the intake point of R1, R2, and R3 water samples), and such low flow water body would be beneficial for plankton accumulation, as Williams described [8].

In order to improve the sensitivity of the system, we have adopted a two-step PCR amplification to increase the amount of DNA templates for pyrosequencing reaction. The number of successful detection was extended from one to four (Table 5). It is worth noting that the reproducibility of successful detection was also related with the amount of DNA templates for *rbcL* gene amplification. In all five freshwater samples, the ML has the highest reproducibility. As we mentioned above, this sample should have the highest plankton density among all five samples. In the real forensic practice, we recommend performing the profiling analysis with one-step PCR firstly, if the results are not informative, two-step PCR should follow in order to gain more sensitivity. However, two-step PCR may produce additional non-specific products and thus interfere with the pyrosequencing profile, which remains to be further investigated. Therefore, investigation of other methods will be beneficial to find alternative options in

certain special cases. In the AdvISER-M-PYRO analysis results, the contribution may be attributed to some library samples other than the right one. For example, in the sensitivity detection experiment, G3 was parsed into a combination of G2 and G3, but G3 was the main contributor with a contribution value of 4.49 (Table 3); G1 was detected as a mixture of G1 and G3, but G1 was the main contributor with a contribution value of 9.56. This phenomenon may come from diatom overlapped in the two groups. Similarly, in the environmental water sample groups, the contribution value from other library samples indicated that there were overlapping diatom constituents between water sites. Such extra contribution may reflect the similarity between environmental water samples, and may be helpful for forensic case investigations.

Conclusion

The pyrosequencing profile of the *rbcL* gene can be used for analyzing plankton populations in drowning animal tissues. The data decoded from the profile can provide information of plankton populations to infer the drowning site instead of analyzing exact components of colonies. This method could be a useful auxiliary approach to infer the drowning site in forensic cases. It is recommended to proceed firstly with the one-step protocol. If the results are not informative, the two-step protocol should be adopted to gain higher sensitivity.

Acknowledgements We thank Dr. John J Xin (Tufts University School of Medicine) for assistance in manuscript revision, as well as Miss Yue Zhang and Mr. Derek Xu (Brown University) for language editing.

Funding information This study was supported by grants from the National Natural Science Foundation of China (81630054, 81571861).

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interests.

References

- Zhao Y, Chen X, Yang Y, Zhao X, Zhang S, Gao Z, Fang T, Wang Y, Zhang J (2018) Potential forensic biogeographic application of diatom colony consistency analysis employing pyrosequencing profiles of the 18S rDNA V7 region. *Int J Legal Med* 132(6):1611–1620. <https://doi.org/10.1007/s00414-018-1849-x>
- Kakizaki E, Sonoda A, Sakai M, Yukawa N (2018) Simple detection of bacterioplankton using a loop-mediated isothermal amplification (LAMP) assay: first practical approach to 72 cases of suspected drowning. *Forensic Sci Int* 289:289–303. <https://doi.org/10.1016/j.forsciint.2018.05.035>
- Kakizaki E, Ogura Y, Kozawa S, Nishida S, Uchiyama T, Hayashi T, Yukawa N (2012) Detection of diverse aquatic microbes in blood and organs of drowning victims: first metagenomic approach using high-throughput 454-pyrosequencing. *Forensic Sci Int* 220(1–3):135–146. <https://doi.org/10.1016/j.forsciint.2012.02.010>
- He F, Huang D, Liu L, Shu X, Yin H, Li X (2008) A novel PCR-DGGE-based method for identifying plankton 16S rDNA for the diagnosis of drowning. *Forensic Sci Int* 176(2–3):152–156. <https://doi.org/10.1016/j.forsciint.2007.08.005>
- Ambroise J, Deccache Y, Irengel L, Savov E, Robert A, Gala JL (2014) Amplicon identification using SparsE representation of multiplex PYROsequencing signal (AdvISER-M-PYRO): application to bacterial resistance genotyping. *Bioinformatics* 30(24):3590–3597. <https://doi.org/10.1093/bioinformatics/btu516>
- Gamache J, Sun G (2015) Phylogenetic analysis of the genus *Pseudoroegneria* and the Triticeae tribe using the *rbcL* gene. *Biochem Syst Ecol* 62:73–81. <https://doi.org/10.1016/j.bse.2015.07.038>
- Kawachi M, Ishimoto M, Mori F, Yumoto K, Sato M, Noël M-H (2013) MCC-NIES list of strains, 9th Edition. Tsukuba, Japan
- Williams LG (1964) Possible relationships between plankton-diatom species numbers and water-quality estimates. *Ecology* 45(4):809–823. <https://doi.org/10.2307/1934927>
- Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30(12):2725–2729. <https://doi.org/10.1093/molbev/mst197>
- Reynolds CS (2006) Ecology of phytoplankton. Cambridge University Press, Cambridge
- Terazawa K, Takatori T (1980) Isolation of intact plankton from drowning lung tissue by centrifugation in a colloidal silica gradient. *Forensic Sci Int* 16(1):63–66. [https://doi.org/10.1016/0379-0738\(80\)90180-2](https://doi.org/10.1016/0379-0738(80)90180-2)
- Carballeira R, Vieira DN, Febrero-Bande M, Muñoz Barús JJ (2018) A valid method to determine the site of drowning. *Int J Legal Med* 132(2):487–497. <https://doi.org/10.1007/s00414-017-1708-1>
- Ago K, Hayashi T, Ago M, Ogata M (2011) The number of diatoms recovered from the lungs and other organs in drowning deaths in bathwater. *Leg Med (Tokyo)* 13(4):186–190. <https://doi.org/10.1016/j.legalmed.2011.04.002>
- Yen LY, Jayaprakash PT (2007) Prevalence of diatom frustules in non-vegetarian foodstuffs and its implications in interpreting identification of diatom frustules in drowning cases. *Forensic Sci Int* 170(1):1–7. <https://doi.org/10.1016/j.forsciint.2006.08.020>
- Yu Z, Liu C, Wang H, Chen L, Hu S, Zhao J (2014) The effect of enzyme digestion time on the detection of diatom species. *Pak J Pharm Sci* 27(3 Suppl):691–694

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.