

# Body fluids should be identified before estimating the time since deposition (TsD) in microbiome-based stain analyses for forensics

Jun Zhang,<sup>1,2</sup> Daijing Yu,<sup>1,2</sup> Tian Wang,<sup>1,2</sup> Niu Gao,<sup>1,2</sup> Linyu Shi,<sup>1,2</sup> Yaya Wang,<sup>1,2</sup> Yumei Huo,<sup>1,2</sup> Zhimin Ji,<sup>1,2</sup> Junli Li,<sup>1,2</sup> Xiaomeng Zhang,<sup>1,2</sup> Liwei Zhang,<sup>1,2</sup> Jiangwei Yan<sup>1,2</sup>

**AUTHOR AFFILIATIONS** See affiliation list on p. 13.

**ABSTRACT** Identification and the time since deposition (TsD) estimation of body fluid stains from a crime scene could provide valuable information for solving the cases and are always difficult for forensics. Microbial characteristics were considered as a promising biomarker to address the issues. However, changes in the microbiota may damage the specific characteristics of body fluids. Correspondingly, incorrect body fluid identification may result in inaccurate TsD estimation. The mutual influence is not well understood and limited the codetection. In the current study, saliva, semen, vaginal secretion, and menstrual blood samples were exposed to indoor conditions and collected at eight time points (from fresh to 30 days). High-throughput sequencing based on the 16S rRNA gene was performed to characterize the microbial communities. The results showed that a longer TsD could decrease the discrimination of different body fluid stains. However, the accuracies of identification still reached a quite high value even without knowing the TsD. Correspondingly, the mean absolute error (MAE) of TsD estimation significantly increased without distinguishing the types of body fluids. The predictive TsD of menstrual blood reached a quite low MAE ( $1.54 \pm 0.39$  d). In comparison, those of saliva ( $6.57 \pm 1.17$  d), semen ( $6.48 \pm 1.33$  d), and vaginal secretion ( $5.35 \pm 1.11$  d) needed to be further improved. The great effect of individual differences on these stains limited the TsD estimation accuracy. Overall, microbial characteristics allow for codetection of body fluid identification and TsD estimation, and body fluids should be identified before estimating TsD in microbiome-based stain analyses.

**IMPORTANCE** Emerged evidences suggest microbial characteristics could be considered a promising tool for identification and time since deposition (TsD) estimation of body fluid stains. However, the two issues should be studied together due to a potential mutual influence. The current study provides the first evidence to understand the mutual influence and determines an optimal process for codetection of identification and TsD estimation for unknown stains for forensics. In addition, we involved aged stains into our study for identification of body fluid stains, rather than only using fresh stains like previous studies. This increased the predictive accuracy. We have preliminary verified that individual differences in microbiotas limited the predictive accuracy of TsD estimation for saliva, semen, and vaginal secretion. Microbial characteristics could provide an accurate TsD estimation for menstrual blood. Our study benefits the comprehensive understanding of microbiome-based stain analyses as an essential addition to previous studies.

**KEYWORDS** microbial characteristics, body fluid identification, time since deposition estimation, mutual influence, codetection

**Editor** Jonathan L. Jacobs, American Type Culture Collection, Manassas, Virginia, USA

Address correspondence to Jiangwei Yan, [yanjw@sxmu.edu.cn](mailto:yanjw@sxmu.edu.cn).

Jun Zhang and Daijing Yu contributed equally to this article. Author order was determined by contributions toward study design and editing.

The authors declare no conflict of interest.

See the funding table on p. 13.

**Received** 14 June 2023

**Accepted** 20 February 2024

**Published** 12 March 2024

Copyright © 2024 Zhang et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Body fluid identification and determination of the time since deposition (TsD) are crucial complements for forensics to assess the relevance of stains from crime scenes and are helpful for crime scene investigation and reconstruction (1, 2). Body fluid identification is commonly performed in forensic labs with conventional approaches by testing a chemical reaction of specific enzyme catalysis (3). The tests are fast and visually detected but have a high error rate. Furthermore, the methods can not provide any information about TsD of body fluids. Spectroscopy, chromatography, and electron spin resonance have been explored for TsD determination of blood stains (4) but are useless for determining TsD of white or nearly colorless body fluids such as saliva, semen, and vaginal secretion. In recent years, several novel approaches by measuring gene expression of messenger RNA (mRNA) (5–8), microRNA (miRNA) (9, 10), and epigenetic markers (11, 12) have been reported to be used for body identification. The degradation patterns of mRNA examined through quantitative PCR could be used to predict the TsD of body fluids (13, 14).

In addition, microbiome-based stain analyses have emerged as a new research area to demonstrate microbial characteristics to identify body fluids and estimate the TsD of stains (15). The number of microbial cells is greater than that of human cells in the human body (16). These microbiotas contain abundant genomic information and even exceed the human genome (17). The results suggested that microbiotas from body fluids can still be detectable when human DNA or RNA are in low biomass or degraded.

Metagenomic information varies across body habitats (18). Therefore, the human microbiome is considered a potential tool to recognize the original body sites of biological materials. Hanssen et al. (19) collected large bacterial 16S sequencing data sets from the American Gut Project and the Human Microbiome Project (HMP) to verify the feasibility of using microbial composition data for body fluid identification. The results suggested that the optimal prediction accuracy was close to 98%. López et al. (20) trained deep learning networks containing sequencing data of 1,636 skin, oral, and vaginal samples from the HMP to classify these tissues. The values of the area under the curve were above 0.99. Most of the aged mock casework samples could still be correctly identified in the study of López et al. (20), although the prediction accuracy decreased in the aged mock casework samples compared with fresh samples (20). In another study, skin, saliva, vaginal fluid, menstrual blood, and semen samples were exposed to indoor conditions for 30 days, and the results showed that these samples were grouped by body site mainly based on the microbial community (21). Nevertheless, a few outliers that could not be correctly identified were observed in the study. These two studies suggested that specific microbial characteristics of tissue or fluid can be maintained for a period of time after tissue or fluid leaves the body. This provides us with a microbiome tool for recognizing body fluid stains left at crime scenes. The TsD of stains seems to have some effects on the microbial characteristics of specific tissues or fluids, causing a reduction in the accuracy of identification. However, previous studies were still insufficient to assess the exact effect of the TsD on body fluid identification based on microbial characteristics.

On the other hand, the microbiome was also considered a promising tool for predicting the TsD of body fluid stains. Environmental changes and the breeding of environmental microorganisms may alter the microbial community structure of the body fluid that leaves the original body site. Wang et al. (1) analyzed 16S rRNA gene sequencing data of saliva stains exposed to indoor conditions for up to 20 days, and the results suggested that the mean absolute deviation of TsD prediction was 1.41 days. Salzmann et al. (22) examined RNA sequencing data from blood, menstrual blood, saliva, semen, and vaginal secretion ranging from fresh to 1.5 years. They found that environmental bacteria altered the microbial composition of body fluid stains. The changing microbial community could be evaluated for TsD prediction. These studies predicted the TsD of stains under the condition that the type of body fluid was known. In fact, the type of body fluid at the crime scene is often unknown. Microbial successions of various body

fluids might be very different. The effect of body fluid identification on TsD prediction is still unclear.

In summary, body fluid identification and TsD estimation of stains based on microbial characteristics were investigated separately in previous studies. The feasibility of simultaneous body fluid identification and TsD estimation has not been well explored. In addition, changes in the microbial community of stains provide a chance to predict TsD but may damage the specific characteristics used for body fluid identification simultaneously. Incorrect body fluid identification may also result in inaccurate TsD estimation. Their mutual influence needs further exploration. This problem determines which process should take precedence in microbiome-based stain analyses for forensics.

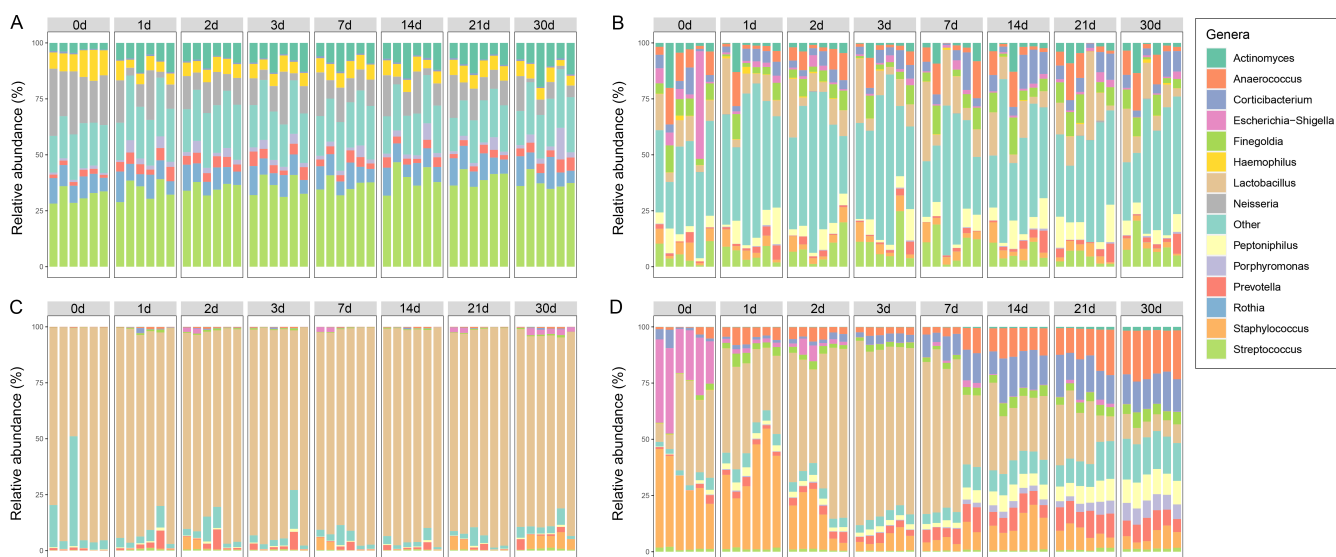
To investigate the problem, saliva, semen, vaginal secretion, and menstrual blood samples were exposed to indoor conditions ranging from fresh to 1 month. Microbial communities of body fluid stains were characterized by targeted 16S rRNA gene high-throughput sequencing at eight different time points. The mutual influences between body fluid identification and TsD prediction were measured to provide a reliability assessment and determine an optimal process for microbiome-based analyses for unknown stains at crime scenes.

## RESULTS

### Microbial community compositions of various body fluids at different TsD values

In the current study, a total of 192 samples of body fluids were sequenced and produced 14,990,142 high-quality (above Q20) sequences. We identified 12,347 amplicon sequence variants (ASVs) from these sequences. The sequence of each sample was rarefied to an even depth of 44,676 (using “single\_rarefaction.py” script in QIIME1 based on the minimum value of all samples) to normalize the ASV table.

We investigated the microbial compositions of four body fluids at eight time points at the phylum (Fig. S1) and genus (Fig. 1) levels. Firmicutes was the dominant taxon in the four body fluids at the phylum level. The highest relative abundance of Firmicutes was observed in vaginal secretion. Actinobacteria, Bacteroidetes, and Proteobacteria were important taxa in saliva, semen, and menstrual blood. In these three body fluids, a higher relative abundance of Proteobacteria and lower relative abundances of Actinobacteria and Bacteroidetes were observed in fresh body fluids than in aged stains (from Day 1 to Day 30).



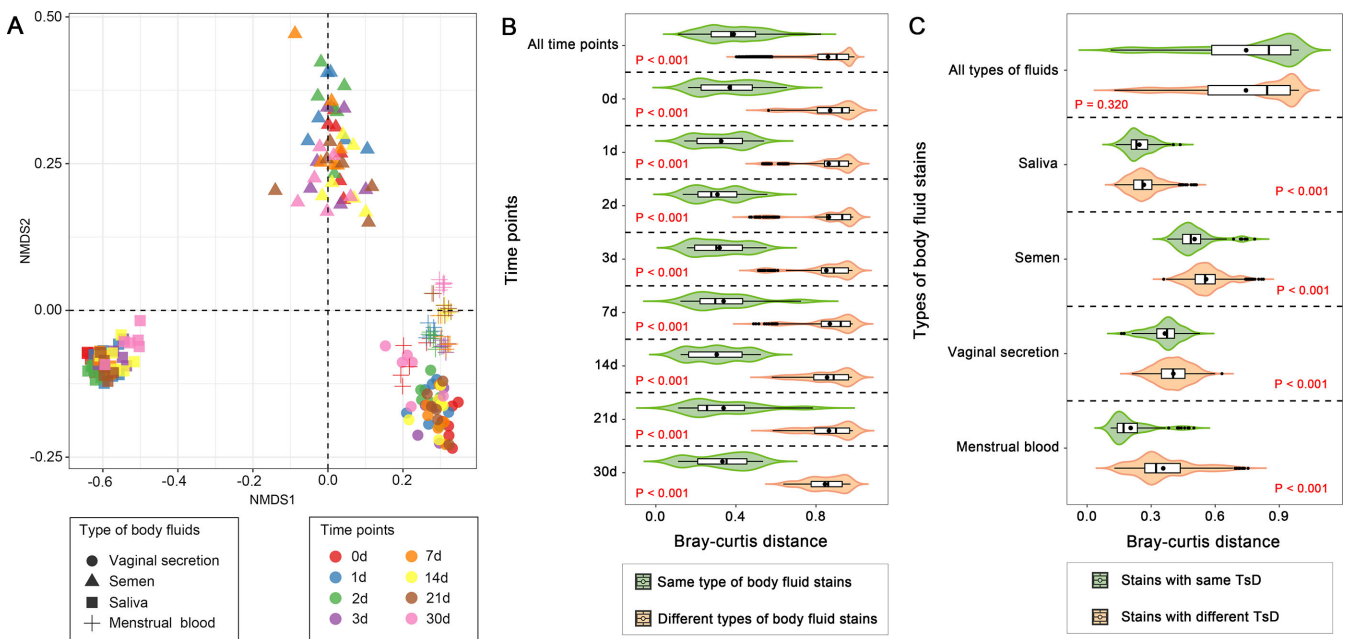
**FIG 1** Relative abundance of bacterial genera in saliva (A), semen (B), vaginal secretion (C), and menstrual blood (D) across 30 days of exposure.

*Streptococcus* was the dominant genus in saliva (Fig. 1A). Fresh saliva contained the lowest relative abundances of *Streptococcus*, *Actinomyces*, and *Prevotella* and the highest relative abundances of *Haemophilus* and *Neisseria*. Semen contained the most genera and lacked an obvious dominant genus in the microbial community composition (Fig. 1B). *Lactobacillus* was the obvious dominant genus in vaginal secretion, and its mean relative abundance reached 89.74%. A significant increase was observed in *Escherichia-Shigella* (Wilcoxon test,  $P = 0.041$ ) and *Staphylococcus* (Wilcoxon test,  $P = 0.004$ ) in vaginal secretion at Day 30 compared with the fresh fluids (Fig. 1C). The clearest trend of microbial succession was observed in menstrual blood (Fig. 1D). *Escherichia-Shigella* and *Staphylococcus* decreased (Kruskal–Wallis test,  $P < 0.001$ ), while *Anaerococcus*, *Corticibacterium*, *Finexgoldia*, *Peptoniphilus*, *Porphyromonas*, and *Prevotella* increased (Kruskal–Wallis test,  $P < 0.001$ ) in relative abundance from the fresh stage to aged stages. *Lactobacillus* first increased and then decreased, and its highest relative abundance was observed at Day 3.

### The effects of different factors on the microbial community compositions of body fluids

The result of nonmetric multidimensional scaling (NMDS) showed that samples clustered mainly based on their original body sites rather than the TsD of body fluids. Semen samples were more heterogeneous than other types of body fluids. Menstrual blood and vaginal secretion were very similar compared with the two other body fluids (Fig. 2A). The result of PERMANOVA showed that original body sites explained most of the variation ( $R^2 = 0.694$ ,  $P < 0.001$ ) in the microbial community of body fluids. TsD had a small but significant effect ( $R^2 = 0.012$ ,  $P < 0.001$ ) on the microbial community structure of body fluids. In addition, the interaction of original body sites and TsD explained 0.039 ( $P < 0.001$ ) of the variation (Table 1).

The Bray–Curtis distances between samples of the same body fluid were significantly ( $P < 0.001$ ) lower than those of different body fluids for each sampling time point. A similar result was observed in the merged data that included all samples from different



**FIG 2** The effects of body fluid origins and TsD on microbial community structures of body fluid stains. NMDS of Bray–Curtis distances between samples is shown in A. Bray–Curtis distances of samples from the same type and from different types of body fluids for all time points (without distinguishing TsD) and each time point are shown in B. Bray–Curtis distances of samples with the same TsD and with different TsD for all types of body fluids (without distinguishing types of body fluids) and each type of body fluid are shown in C. Statistical differences were analyzed using the Wilcoxon tests.

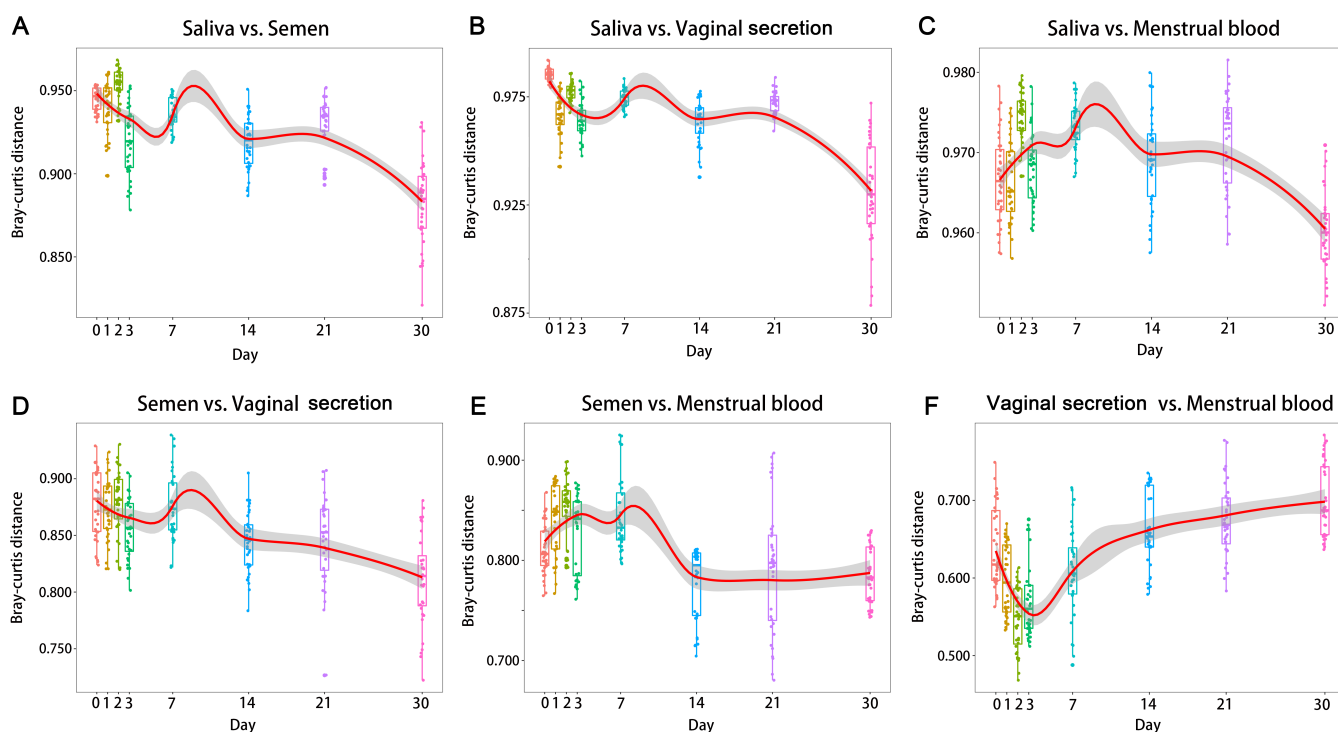
**TABLE 1** PERMANOVA tests of the effects of type of body fluid, TsD, and their interaction on microbial beta diversity between samples

| Factors   | $R^2$ | $P$    |
|-----------|-------|--------|
| Type      | 0.689 | <0.001 |
| TSD       | 0.024 | <0.001 |
| Type: TsD | 0.068 | <0.001 |
| Residual  | 0.217 |        |

time points (Fig. 2B). This suggested that we could still identify the body fluids effectively without distinguishing time points. Significant differences ( $P < 0.001$ ) in Bray–Curtis distance were observed between samples from the same day and different days for each type of body fluid. However, the Bray–Curtis distances of samples from the same day and different days almost overlapped without distinguishing the types of body fluids (Fig. 2C). This suggested that we could not distinguish whether the samples were from the same day without knowing the type of body fluids.

We also investigated the Bray–Curtis distance between pairwise types of body fluids varying in TsD (Fig. 3). The results showed that the distance of the microbial community between vaginal secretion and menstrual blood decreased from fresh to Day 2 and then increased from Day 2 to Day 30. The greatest value was observed at Day 30 (Fig. 3F). The changing pattern was very different from other combinations of pairwise comparisons. The smallest value was observed at Day 30 for those comparisons (Fig. 3A through E). This suggested that exposure for 30 days decreased (Wilcoxon tests,  $P < 0.001$ ) the discrimination between different body fluids except for between vaginal secretion and menstrual blood. In addition, the Bray–Curtis distances between vaginal secretion and menstrual blood were smaller (Wilcoxon tests,  $P < 0.001$ ) than those of other comparisons.

The effect of TsD on the microbial community structure for each body fluid was investigated in our study. In addition, the effect of individual differences (ID) was also measured. The results showed that individual differences explained a greater variation in microbial variation than TsD in saliva, semen, and vaginal secretion. In contrast, TsD



**FIG 3** Pairwise comparisons (Bray–Curtis distance) between different types of body fluids across 30 days of exposure.

TABLE 2 PERMANOVA test of the effects of type of TsD, ID, and their interaction on microbial beta diversity between samples for saliva, semen, vaginal secretion, and menstrual blood, respectively

| Factors  | Saliva |        | Semen |        | Vaginal secretion |        | Menstrual blood |        |
|----------|--------|--------|-------|--------|-------------------|--------|-----------------|--------|
|          | $R^2$  | $P$    | $R^2$ | $P$    | $R^2$             | $P$    | $R^2$           | $P$    |
| TsD      | 0.077  | <0.001 | 0.036 | 0.026  | 0.046             | 0.034  | 0.427           | <0.001 |
| ID       | 0.471  | <0.001 | 0.212 | <0.001 | 0.519             | <0.001 | 0.021           | 1.000  |
| TsD:ID   | 0.058  | 0.380  | 0.095 | 0.400  | 0.042             | 0.642  | 0.013           | 1.000  |
| Residual | 0.395  |        | 0.657 |        | 0.393             |        | 0.539           |        |

explained more variation than individual differences in the microbial community of menstrual blood (Table 2; Fig. S2).

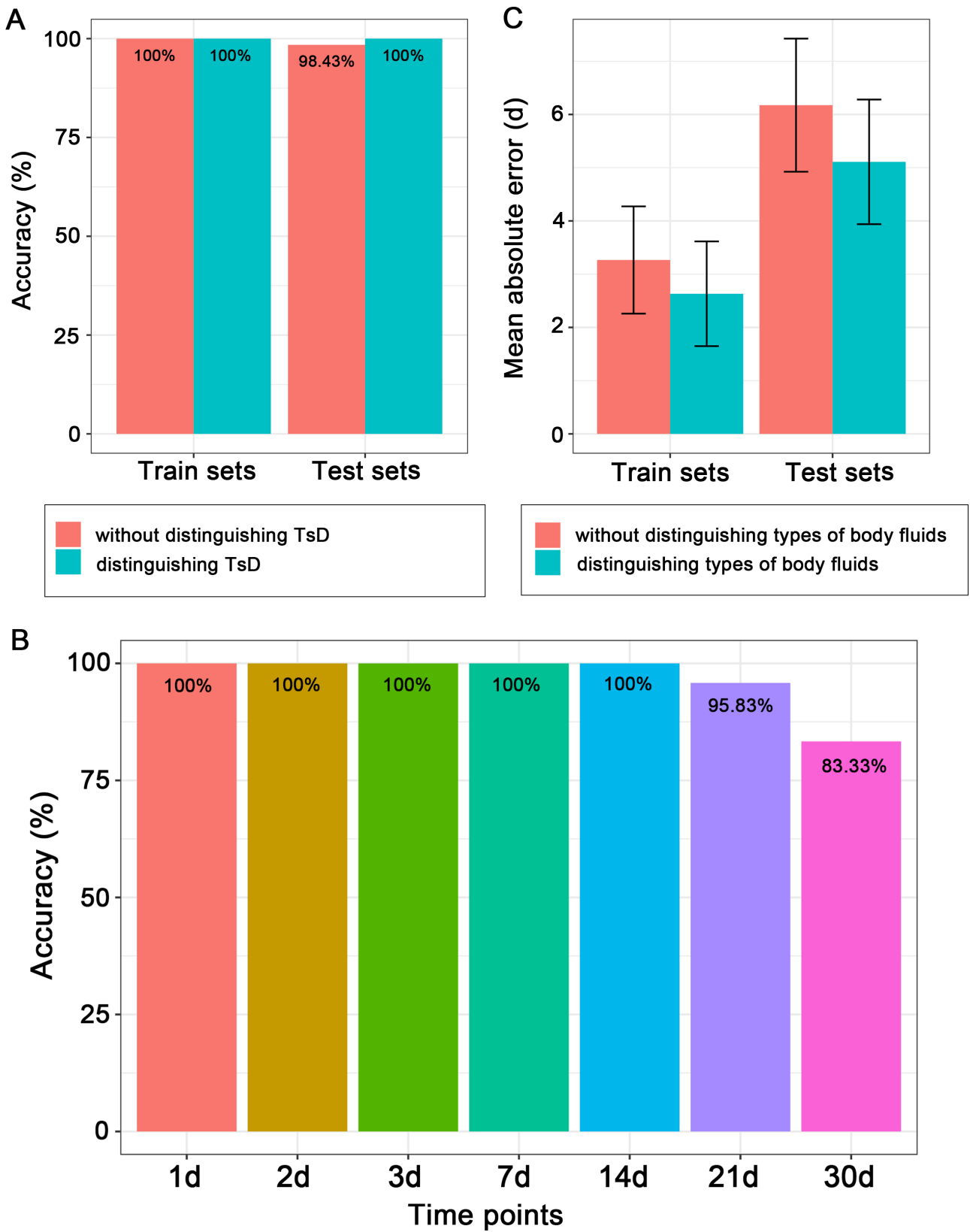
Significant ( $P < 0.001$ ) linear relationships between similarities of microbial communities and intervals of TsD were observed in all four body fluids (Fig. S3). This suggested that the microbial communities of these body fluids changed with a regular pattern and provided a potential chance to predict TsD. The steepest slope was observed in menstrual blood, suggesting the fastest rate of microbial succession.

### Body fluid identification and TsD prediction based on microbial community characteristics

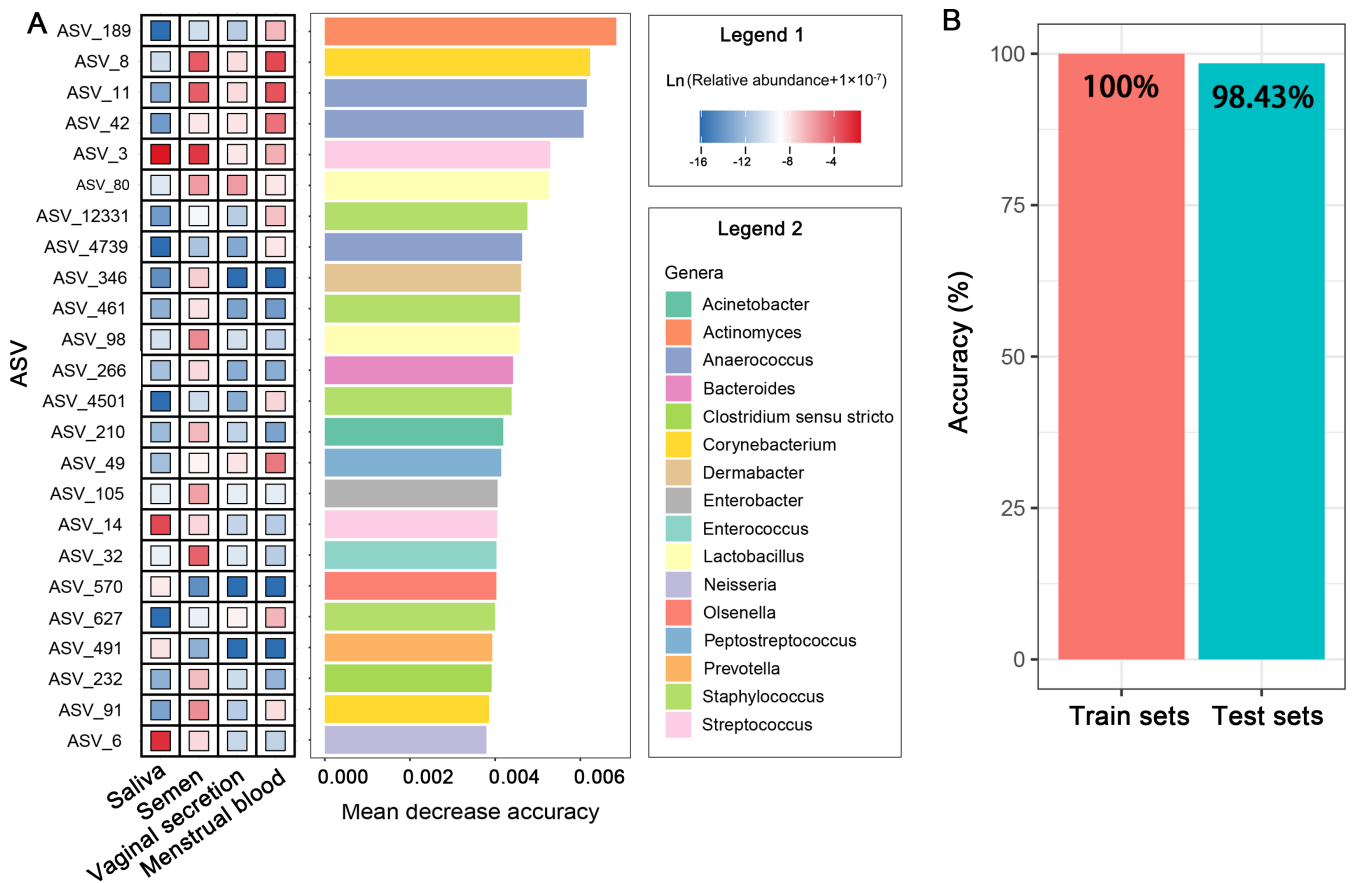
The random forest algorithm was used to construct initial predictive models and select microbial biomarkers. Training sets were used to construct the models, and both training and testing sets were used to test the accuracies of the models. The results showed that the accuracies of body fluid identification reached an extreme value (training sets: 100.00%; testing sets: 98.43%) even without knowing the TsD (Fig. 4A). In addition, fresh body fluids (Day 0) were used to train the model to recognize the types of aged body fluid stains (from Day 1 to Day 30). The results showed that the predictive accuracies for samples collected from Day 1 to Day 14 reached 100.00%. Then, the accuracies decreased at Day 21 (95.83%) and Day 30 (83.33%). This suggested that the model that was only constructed with fresh body fluid data was insufficient to reach a high identification accuracy for aged body fluid stains (Fig. 4B). The accuracies of TsD prediction significantly decreased without distinguishing the types of body fluids, regardless of whether training sets (Wilcoxon test,  $P = 0.004$ ) or testing sets (Wilcoxon test,  $P = 0.032$ ) were used (Fig. 4C).

Therefore, the data used to construct the model of body fluid identification should be collected from various TsD values. However, the TsD prediction model should be constructed separately for each body fluid. Microbial biomarkers were selected based on the minimum error of 10-fold cross-validation with five repeats to construct final prediction models. Finally, 24 ASVs were selected as the optimal biomarker set to construct the model of body fluid identification (Fig. 5A; Fig. S4). Among these biomarkers, ASV\_189 belonging to *Actinomyces* was the most important taxon for classification, followed by ASV\_8 belonging to *Corynebacterium* and ASV\_11 and ASV\_42 belonging to *Anaerococcus*. The highest and lowest relative abundances of these taxa were observed in menstrual blood and saliva, respectively. ASV\_3 and ASV\_14, belonging to *Streptococcus*, had the highest relative abundance in saliva (Fig. 5A).

In addition, 9 ASVs, 677 ASVs, 296 ASVs, and 48 ASVs were selected as the optimal biomarker sets to construct the models of TsD prediction for saliva, semen, vaginal secretion, and menstrual blood, respectively. The most important taxa of the top 20 are ranked. ASV\_61 belonging to *Fusobacterium*, ASV\_376 belonging to *Clostridium sensu stricto 1*, and ASV\_1172 belonging to *Dialister* were the most important taxa to predict TsD of saliva (Fig. S5A), semen (Fig. S5B), and vaginal secretion (Fig. S5C), respectively. These taxa decreased from fresh to Day 30 in the corresponding body fluids. ASV\_78, belonging to *Porphyromonas*, was the most important taxon to predict TsD of menstrual blood. Interestingly, all of the most important taxa ranked in the top 20 increased from fresh to Day 30 in menstrual blood (Fig. S5D).



**FIG 4** The accuracies of predictive models for body fluid identification (A and B) and TsD estimation (C). The predictive accuracies for body fluid identification performed with samples of all time points (without distinguishing TsD) and each time point are shown in A. The accuracies of predictive models constructed with fresh samples to recognize aged body fluids are shown in B. The predictive accuracies for TsD estimation performed with samples of all types of body fluids (without distinguishing types of body fluids) and each type of body fluid are shown in C.



**FIG 5** The optimal biomarkers were selected to construct final predictive models for body fluid identification and TsD estimation. The optimal biomarkers selected for body fluid identification are shown in A. The predictive accuracies for body fluid identification performed with optimal biomarkers are shown in B.

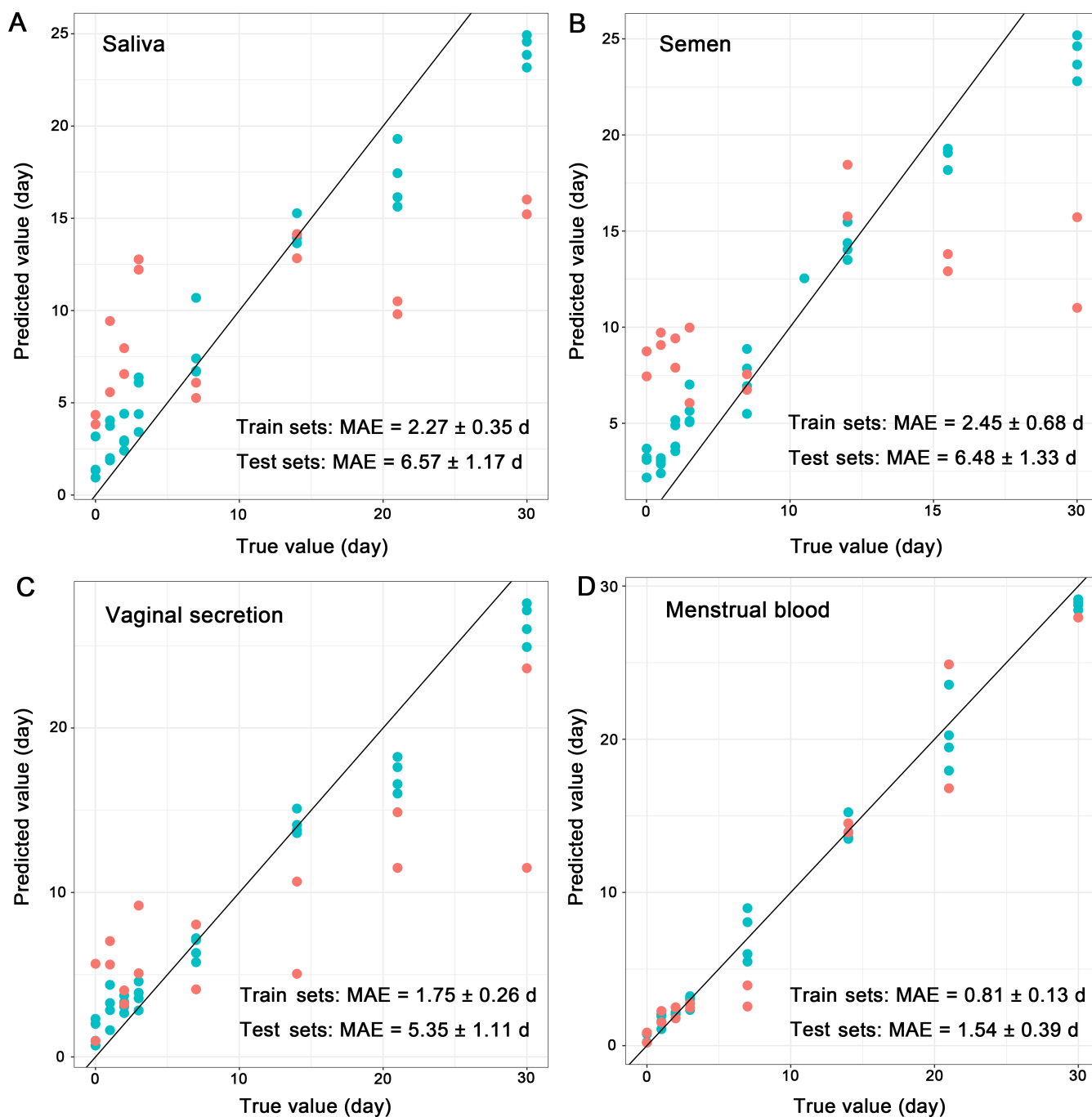
Then, the optimal biomarker sets were used to construct the final predictive model. The results showed that the accuracies of the final model in recognizing body fluids were the same as those of the initial model (training set: 100%, testing set: 98.43%; Fig. 5B). For the models for predicting TsD, the predictive results of the final models showed that the MAEs (mean ± SE) of menstrual blood (training set: 0.81 ± 0.13 d, testing set: 1.54 ± 0.39 d; Fig. 6D) were the lowest, followed by vaginal secretion (training set: 1.75 ± 0.26 d, testing set: 5.35 ± 1.11 d; Fig. 6C). The MAEs of saliva (training set: 2.27 ± 0.35 d, testing set: 6.57 ± 1.17 d; Fig. 6A) and semen (training set: 2.45 ± 0.68 d, testing set: 6.48 ± 1.33 d; Fig. 6B) were relatively higher, especially for the predictive results of the test sets.

**DISCUSSION**

Microbiome-based stain analyses have attracted the attention of the forensics community as a promising tool for body fluid identification and TsD prediction (1). These two questions were investigated separately in previous studies. However, these two questions are not isolated but affect each other. Our current study provides the first assessment of the mutual influence between body fluid identification and TsD prediction and determines an optimal process for microbiome-based analyses that allows for identifying body fluids and estimating the TsD of stains simultaneously.

The compositions of the microbial community varied obviously among samples from different types of body fluids. The type of body fluids had a greater effect on the compositions of microbial communities in body fluid stains compared with TsD during the exposure experiment of 30 days. Furthermore, significant differences could still





**FIG 6** The predictive accuracies of TsD estimation performed with optimal biomarkers for saliva (A), semen (B), vaginal secretion (C), and menstrual blood (D) were shown using graph of predicted value versus true value. The line represents a perfect prediction of TsD. The cyan points represent the predictive results of train sets, and the pale red points represent the predictive results of test sets.

be observed after exposure for 30 days, although there were shifts in the microbial community for each body fluid.

Most of the test body fluid samples presented a taxonomic composition dominated by the Firmicutes phylum, in agreement with most previous studies (1, 21, 23–25). Only the microbial taxonomic profiles of semen were inconsistent with the research results (dominated by Proteobacteria) of Salzmann et al. (22). However, most studies on seminal microbiotas support our results (23, 26, 27). Of course, this does not mean that the result of Salzmann et al. (22) is incorrect. Human semen microbiome diversity shows spatial

heterogeneity (28). Study populations from different regions may explain the difference in seminal microbial taxonomic compositions.

At the genus level, *Streptococcus* was the dominant taxon in the fresh salivary microbiome, followed by *Neisseria*, *Haemophilus*, *Rothia*, and *Prevotella*. In previous studies, these genera were considered common taxa in the saliva of healthy adults (29–31). These genera maintained their predominance throughout the duration of indoor exposure. *Streptococcus*, *Actinomyces*, and *Prevotella* increased and *Haemophilus* and *Neisseria* decreased along with TsD in relative abundance. These shifts were significant but slight. There was a lack of an obvious dominant genus in seminal microbiotas. Great variations in microbial compositions between individuals were observed. This result is in agreement with a previous study (23). It was difficult to observe a consistent shift at the genus level across the exposure experiment.

*Lactobacillus* was absolutely dominant at the genus level in vaginal secretion. The predominance was maintained across 30 days of exposure. Large amounts of *Lactobacillus* breeding in the vagina may produce an acidic environment to maintain vaginal health (32). Therefore, *Lactobacillus* was one of the dominant genera in fresh menstrual blood. Menstrual blood is a mixture of vaginal secretion and blood (25). The NMDS results showed that menstrual blood was most similar to vaginal secretion. However, there are still obvious variations between these two body fluids to distinguish them. The study of Song et al. (33) suggested that the relative abundance of *Lactobacillus* in the vagina decreased during menses (33). This is consistent with our results. Our samples of menstrual blood were collected from days 2–5 of menstrual cycle. Though, the previous study showed that the vaginal microbiome of most women remained relatively stable throughout the menstrual cycle (34), we are still unable to determine whether menstrual blood could be distinguished from vaginal secretion at the beginning or end of menses. The question will be explored in our further studies. In addition, the most obvious and regular succession of the microbial community was observed in the menstrual blood.

The Bray–Curtis distance between vaginal secretion and menstrual blood increased after exposure for 30 days, and the trend was completely contrary to other pairwise comparisons. In addition to cervical mucus (the main component of vaginal secretion), menstrual blood contains more blood (35). Although vaginal secretion and menstrual blood are excreted through the vagina, quite different nutrient substances determine different directions of succession of microbiology communities. Apart from that, exposure for 30 days diminished the uniqueness of microbial characteristics for each body fluid. The effects of the unique habitat for microbial communities disappeared, and the same exposure conditions and environmental microbial community might be related to convergence on microbial communities of body fluids (36). The previous studies showed that the microbial communities in different body parts of cadavers became similar during the decay process (37). In fact, deposition of body fluids could also be considered a decay process.

However, Bray–Curtis distances of pairwise comparisons at Day 30 still reached a high value to distinguish different body fluids. Dobay et al. (21) exposed body fluids (skin, saliva, peripheral blood, menstrual blood, vaginal fluid, and semen) to indoor conditions for 30 days, while the same samples were processed directly after extraction for comparison. They found that both types of samples were grouped by body site. PCA was performed separately for fresh and aged samples (21). Whether aged samples can be grouped with fresh samples from the same body site is unclear. The question is important. In actual cases, it is difficult to know exactly when the body fluid was deposited. Most studies use the microbial community of fresh body fluids to train a predictive model (2, 19, 20). In our study, the accuracies decreased at Day 21 and Day 30 when only fresh body fluids were used as training data. López et al. (20) trained a deep learning network with data from the HMP and found that the predictive accuracy decreased in aged mock casework samples compared with fresh samples. This suggested that the TsD of body fluids had an effect on the accuracy of body fluid identification based on microbial characteristics.

Therefore, the predictive model for body fluid identification constructed only with fresh samples may not be an optimal solution. Of course, this does not mean that an exact TsD must be known in advance before body fluid identification. The Bray–Curtis distances between samples from different types of body fluids were still significantly greater than those from the same type of body fluids without discriminating TsD. The model was constructed with the integrated data containing fresh and aged samples and reached a high identification accuracy. In other words, we can still recognize body fluids exactly even if we do not know the TsD. In turn, the Bray–Curtis distances between samples from different TsD values are almost completely coincident with the differences between samples from the same TsD without discriminating sample types. However, the differences were significant in each comparison of a specific body fluid. The accuracies of TsD prediction significantly decreased without discriminating the types of body fluids. This suggests that body fluids should be identified before estimating the TsD in microbiome-based stain analyses for forensics.

In addition, the predictive TsD of menstrual blood was more accurate than that of other body fluids. The predictive accuracies for saliva, semen, and vaginal secretion need to be further improved. Asaghiar et al. (38) evaluated two hypoxia-sensitive RNA markers based quantitative PCR (qPCR) to develop a predictive model of TsD for blood, saliva, and semen, resulting in an MAE of 4.2, 2.1, and 5 days of within 28 days of degrading at room temperature, respectively (38). Wang et al. (37) regressed a random forest model with salivary microbiota to predict TsD, resulting in an MAE of 1.41 days within 20 days of exposure (1). The models of these two studies were tested with training data. The predictive accuracies of our study were closed to or even better than those of the previous studies if we used the training sets for testing. However, the accuracies decreased significantly when the testing sets were tested. In general, the data used for validation should be distinguished from the training sets to avoid overfitting (39). A sustainable effect of individual difference was observed in the microbial communities of these three types of body fluids throughout the entire process of exposure. The variations in microbial communities caused by individual differences were even higher than those caused by TsD. A stable individual variation in the microbial community is adverse to TsD prediction but beneficial to individual identification.

The current study still lacked mock casework samples to test our conclusion. The complex environment of the crime scene may cause a deviation from our result. In addition, a longer study may produce a varying result. The predictive accuracies of TsD for saliva, semen, and vaginal secretion need further improvement by including a larger sample size or other biomarkers, such as mRNA. In addition, mixed bodily fluids with one or more body fluids are commonly encountered in crime scenes. However, mixed body fluids were not involved in the current study. Microbial source-tracking methods such as SourceTracker (40) and FEAST (41) can be applied to forensic science and provide a potential solution to identified mixed body fluids. These questions will be explored in our further study. However, the current study provided the first evidence to understand the mutual influence between body fluid identification and TsD prediction in microbiome-based stain analyses. Our study supplied a potential solution to recognize body fluid and estimate the TsD of stains simultaneously based on microbial characteristics for forensics and determined a relatively optimal process for codetection.

## MATERIALS AND METHODS

### Sample collection

Body fluid samples (including saliva, semen, vaginal secretion, and menstrual blood) were collected on sterile cotton swabs from health volunteers. All volunteers were recruited from Shanxi Medical University in March and provided written informed consent. Exclusion factors included taking antibiotics within 1 month. The study was ethically approved by the Medicine Institutional Review Board of Shanxi Medical University (No. 2021GLL049) according to the guidelines of the World Medical

Association and the Declaration of Helsinki. Saliva and semen samples were first self-collected into sterile tubes and then transferred to sterile cotton swabs. Vaginal secretion and menstrual blood samples were self-collected from the vagina directly on sterile cotton swabs.

Then, all samples were placed on a lab bench and exposed to indoor conditions (temperature, 18.4–19.4°C; humidity, 30%–40%). Eight sampling time points (fresh, Day 1, Day 2, Day 3, Day 7, Day 14, Day 21, and Day 30) were set in our experiment. Each body fluid at each time point included six biological replicates corresponding to six volunteers. Saliva samples were collected from three males and three females. In total, 192 samples of body fluid stains were collected. All samples were collected and stored at –80°C before further processing.

### DNA extraction, library preparation, and sequencing

Total genomic DNA was extracted from body fluid stains using the DNeasy Power-Soil Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. Universal primers (505F- GTGCCAGCMGCCGCGGTAA and 806R- GGACTACHVGGGTWTCTAAT) to the variable region 4 of the 16S rRNA gene, which contained Illumina adapters and barcode sequences for sequencing, were amplified. The amplification products were quantified using QuantStudio5.0 (Applied Biosystems, Foster City, USA) and pooled in equimolar concentrations. The pooled sequencing libraries were sequenced using the Illumina NovaSeq platform. Negative controls consisting of six sterile swabs were included during DNA extraction and sequencing.

### Data analysis

We obtained a cluster density of 1,130 k/mm<sup>2</sup> and a passing filter of 88.9%. Totally, 16,335,500 sequences were generated. More than 97.07% of the bases have a Phred score above Q30. Raw data were quality filtered using fastp software (42). Then, the clean data were merged and dereplicated using VSEARCH (43) and USEARCH (44). The UCHIME algorithm (45) was applied to remove chimeric sequences. After trimming and filtering, 14,990,142 reads remained. USEARCH-UNOISE3 (46) was used to identify ASVs. The taxonomic annotation of the ASV sequence was obtained by alignment with the Silva (v138) database using BLAST (47).

The statistics in our study were analyzed using the R software platform (v3.6.0; <http://www.r-project.org/>). The taxonomic compositions of the samples were shown at the phylum and genus levels using the “ggplot2” package (48). Wilcoxon test was performed for the comparison between bacterial compositions of two groups, and the Kruskal–Wallis test was performed for the comparison between multiple groups. NMDS using the “vegan” package (49) was performed and visualized to assess the effects of body fluid origins and TsD on microbial community structures of body fluid stains. Permutational multivariate analysis of variance (PERMANOVA) based on Bray–Curtis distance was used to verify the result of NMDS in statistics. Within-group variation and between-group variation were calculated and compared based on the Bray–Curtis distance to test the discrimination of different grouping manners for samples. The Wilcoxon test with *P* values adjusted according to the Benjamini–Hochberg method (50) was performed to test the significance of differences in beta diversities among various groups. The correlation of microbial community similarities (a value of 1 minus Bray–Curtis distance was ln-transformed) with intervals between pairwise time points was tested by fitting a linear regression.

To measure the effect of the TsD of stain on body fluid identification, classification models for body fluid identification were constructed using the data that included all time points and the data of each time point. Then, the accuracies of the two classification models were compared. Similarly, regression models for TsD prediction were constructed using the data that included all types of body fluids and the data of each body fluid. The accuracies of the regression models were measured by using the MAE. The stratified random sampling method was employed to divide the data into training and testing

sets at a proportion of 2:1 using the “sampling” package (51). The initial models were constructed with a training set using the random forest algorithm from the “randomForest” package (52). The value of “ntree” was set to 1,000. Bacterial ASVs were ranked in order of their feature importance. Furthermore, the random forest algorithm was used to select the optimal biomarker sets according to the order of importance based on the minimum error of 10-fold cross-validation by 100 iterations with five repeats. Each subset of training sets was partitioned based on individual for cross-validation. The final prediction models were constructed using the optimal biomarker sets.

## ACKNOWLEDGMENTS

This work was supported by grants from the National Natural Science Foundation of China (Nos. 82030058 and 82101977), the Fundamental Research Program of Shanxi Province (No. 20210302124671), and the China Postdoctoral Science Foundation funded project (No. 2022M712977).

J.W.Y. and J.Z. conceived and designed the manuscript. D.J.Y., N.G., L.Y.S., Y.Y.W., Y.M.H., Z.M.J., J.L.L., and X.M.Z. collected the samples. J.Z., D.J.Y., T.W., and L.W.Z. processed the samples and sequenced the data. J.Z. analyzed the data and wrote the manuscript. All authors contributed to the article and approved the submitted version.

## AUTHOR AFFILIATIONS

<sup>1</sup>School of Forensic Medicine, Shanxi Medical University, Jinzhong, Shanxi, China

<sup>2</sup>Shanxi Key Laboratory of Forensic Medicine, Jinzhong, Shanxi, China

## AUTHOR ORCIDs

Jun Zhang  <http://orcid.org/0000-0002-3113-4963>

Jiangwei Yan  <http://orcid.org/0000-0002-0359-4508>

## FUNDING

| Funder  | Grant(s)       | Author(s)    |
|---|----------------|--------------|
| <a href="#">MOST   National Natural Science Foundation of China (NSFC)</a>                    | 82030058       | Jiangwei Yan |
| <a href="#">MOST   National Natural Science Foundation of China (NSFC)</a>                    | 82101977       | Jun Zhang    |
| <a href="#">Fundamental Research Program of Shanxi Province</a>                               | 20210302124671 | Jun Zhang    |
| <a href="#">China Postdoctoral Science Foundation (China Postdoctoral Foundation Project)</a> | 2022M712977    | Jun Zhang    |

## AUTHOR CONTRIBUTIONS

Jun Zhang, Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review and editing | Daijing Yu, Data curation, Investigation, Validation | Tian Wang, Data curation, Validation, Writing – review and editing | Niu Gao, Data curation, Validation | Linyu Shi, Data curation, Validation | Yaya Wang, Data curation, Validation | Yumei Huo, Data curation, Validation | Zhimin Ji, Data curation, Validation | Junli Li, Data curation, Validation | Xiaomeng Zhang, Data curation, Validation | Liwei Zhang, Data curation, Validation | Jiangwei Yan, Conceptualization, Funding acquisition, Validation, Writing – review and editing

## DATA AVAILABILITY

All sequence reads generated in this study have been deposited to the NCBI Sequence Read Archive (SRA) under BioProject accession [PRJNA979326](#). Other data and code in this study are available from the corresponding author upon reasonable request.

## ADDITIONAL FILES

The following material is available [online](#).

## Supplemental Material

**Supplemental figures (Spectrum02480-23-s0001.pdf).** Fig. S1 to S5.

**Supplemental material (Spectrum02480-23-s0002.xlsx).** STORMS checklist.

## REFERENCES

- Wang J, Cheng X, Zhang J, Liu Z, Cheng F, Yan J, Zhang G. 2022. Estimating the time since deposition (TsD) in saliva stains using temporal changes in microbial markers. *Forensic Sci Int Genet* 60:102747. <https://doi.org/10.1016/j.fsigen.2022.102747>
- Diez López C, Montiel González D, Haas C, Vidaki A, Kayser M. 2020. Microbiome-based body site of origin classification of forensically relevant blood traces. *Forensic Sci Int Genet* 47:102280. <https://doi.org/10.1016/j.fsigen.2020.102280>
- Hanssen EN, Avershina E, Rudi K, Gill P, Snipen L. 2017. Body fluid prediction from microbial patterns for forensic application. *Forensic Sci Int Genet* 30:10–17. <https://doi.org/10.1016/j.fsigen.2017.05.009>
- Weber AR, Lednev IK. 2020. Crime clock – analytical studies for approximating time since deposition of bloodstains. *Forensic Chem* 19:100248. <https://doi.org/10.1016/j.forc.2020.100248>
- Haas C, Hanson E, Anjos MJ, Bär W, Banemann R, Berti A, Borges E, Bouakaze C, Carracedo A, Carvalho M, et al. 2012. RNA/DNA co-analysis from blood stains—results of a second collaborative EDNAP exercise. *Forensic Sci Int Genet* 6:70–80. <https://doi.org/10.1016/j.fsigen.2011.02.004>
- Haas C, Hanson E, Anjos MJ, Banemann R, Berti A, Borges E, Carracedo A, Carvalho M, Courts C, De Cock G, et al. 2013. RNA/DNA co-analysis from human saliva and semen stains—results of a third collaborative EDNAP exercise. *Forensic Sci Int Genet* 7:230–239. <https://doi.org/10.1016/j.fsigen.2012.10.011>
- Haas C, Hanson E, Anjos MJ, Ballantyne KN, Banemann R, Bhoelai B, Borges E, Carvalho M, Courts C, De Cock G, et al. 2014. RNA/DNA co-analysis from human menstrual blood and vaginal secretion stains: results of a fourth and fifth collaborative EDNAP exercise. *Forensic Sci Int Genet* 8:203–212. <https://doi.org/10.1016/j.fsigen.2013.09.009>
- Haas C, Hanson E, Banemann R, Bento AM, Berti A, Carracedo A, Courts C, Cock GD, Drobnic K, Fleming R, et al. 2015. RNA/DNA co-analysis from human skin and contact traces—results of a sixth collaborative EDNAP exercise. *Forensic Sci Int Genet* 16:139–147. <https://doi.org/10.1016/j.fsigen.2015.01.002>
- Hanson EK, Lubenow H, Ballantyne J. 2009. Identification of forensically relevant body fluids using a panel of differentially expressed microRNAs. *Anal Biochem* 387:303–314. <https://doi.org/10.1016/j.ab.2009.01.037>
- Sauer E, Reinke AK, Courts C. 2016. Differentiation of five body fluids from forensic samples by expression analysis of four microRNAs using quantitative PCR. *Forensic Sci Int Genet* 22:89–99. <https://doi.org/10.1016/j.fsigen.2016.01.018>
- Park JL, Kwon OH, Kim JH, Yoo HS, Lee HC, Woo KM, Kim SY, Lee SH, Kim YS. 2014. Identification of body fluid-specific DNA methylation markers for use in forensic science. *Forensic Sci Int Genet* 13:147–153. <https://doi.org/10.1016/j.fsigen.2014.07.011>
- Silva D, Antunes J, Balamurugan K, Duncan G, Alho CS, McCord B. 2016. Developmental validation studies of epigenetic DNA methylation markers for the detection of blood, semen and saliva samples. *Forensic Sci Int Genet* 23:55–63. <https://doi.org/10.1016/j.fsigen.2016.01.017>
- Bird TAG, Walton-Williams L, Williams G. 2019. Time since deposition of biological fluids using RNA degradation. *Forensic Sci Int Genet Suppl Ser* 7:401–402. <https://doi.org/10.1016/j.fsigs.2019.10.028>
- Salzmann AP, Russo G, Kreutzer S, Haas C. 2021. Degradation of human mRNA transcripts over time as an indicator of the time since deposition (TsD) in biological crime scene traces. *Forensic Sci Int Genet* 53:102524. <https://doi.org/10.1016/j.fsigen.2021.102524>
- Zhang J, Liu W, Simayijiang H, Hu P, Yan J. 2023. Application of microbiome in forensics. *Genom Proteom Bioinform* 21:97–107. <https://doi.org/10.1016/j.gpb.2022.07.007>
- Sender R, Fuchs S, Milo R. 2016. Revised estimates for the number of human and bacteria cells in the body. *PLoS Biol* 14:e1002533. <https://doi.org/10.1371/journal.pbio.1002533>
- Human Microbiome Project Consortium. 2012. Structure, function and diversity of the healthy human microbiome. *Nature* 486:207–214. <https://doi.org/10.1038/nature11234>
- Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JL, Knight R. 2009. Bacterial community variation in human body habitats across space and time. *Science* 326:1694–1697. <https://doi.org/10.1126/science.1177486>
- Hanssen EN, Liland KH, Gill P, Snipen L. 2018. Optimizing body fluid recognition from microbial taxonomic profiles. *Forensic Sci Int Genet* 37:13–20. <https://doi.org/10.1016/j.fsigen.2018.07.012>
- Diez López C, Vidaki A, Ralf A, Montiel González D, Radjabzadeh D, Kraaij R, Uitterlinden AG, Haas C, Lao O, Kayser M. 2019. Novel taxonomy-independent deep learning microbiome approach allows for accurate classification of different forensically relevant human epithelial materials. *Forensic Sci Int Genet* 41:72–82. <https://doi.org/10.1016/j.fsigen.2019.03.015>
- Dobay A, Haas C, Fucile G, Downey N, Morrison HG, Kratzer A, Arora N. 2019. Microbiome-based body fluid identification of samples exposed to indoor conditions. *Forensic Sci Int Genet* 40:105–113. <https://doi.org/10.1016/j.fsigen.2019.02.010>
- Salzmann AP, Arora N, Russo G, Kreutzer S, Snipen L, Haas C. 2021. Assessing time dependent changes in microbial composition of biological crime scene traces using microbial RNA markers. *Forensic Sci Int Genet* 53:102537. <https://doi.org/10.1016/j.fsigen.2021.102537>
- Yao T, Han X, Guan T, Wang Z, Zhang S, Liu C, Liu C, Chen L. 2020. Effect of indoor environmental exposure on seminal microbiota and its application in body fluid identification. *Forensic Sci Int* 314:110417. <https://doi.org/10.1016/j.forsciint.2020.110417>
- Pramanick R, Nathani N, Warke H, Mayadeo N, Aranha C. 2021. Vaginal dysbiotic microbiome in women with no symptoms of genital infections. *Front Cell Infect Microbiol* 11:760459. <https://doi.org/10.3389/fcimb.2021.760459>
- Mei S, Zhao M, Liu Y, Zhao C, Xu H, Fang Y, Zhu B. 2022. Evaluations and comparisons of microbial diversities in four types of body fluids based on two 16S rRNA gene sequencing methods. *Forensic Sci Int* 331:111128. <https://doi.org/10.1016/j.forsciint.2021.111128>
- Mändar R, Punab M, Korrovits P, Türk S, Ausmees K, Lapp E, Preem JK, Oopkaup K, Salumets A, Truu J. 2017. Seminal microbiome in men with and without prostatitis. *Int J Urol* 24:211–216. <https://doi.org/10.1111/iju.13286>
- Altmäe S, Franasiak JM, Mändar R. 2019. The seminal microbiome in health and disease. *Nat Rev Urol* 16:703–721. <https://doi.org/10.1038/s41585-019-0250-y>
- Ma ZS, Li L. 2018. Semen microbiome biogeography: an analysis based on a Chinese population study. *Front Microbiol* 9:3333. <https://doi.org/10.3389/fmicb.2018.03333>
- Hall MW, Singh N, Ng KF, Lam DK, Goldberg MB, Tenenbaum HC, Neufeld JD, G Beiko R, Senadheera DB. 2017. Inter-personal diversity and temporal dynamics of dental, tongue, and salivary microbiota in the healthy oral cavity. *NPJ Biofilms Microbiomes* 3:2. <https://doi.org/10.1038/s41522-016-0011-0>
- Zaura E, Keijsers B, Huse SM, Crielaard W. 2009. Defining the healthy "core microbiome" of oral microbial communities. *BMC Microbiol* 9:259. <https://doi.org/10.1186/1471-2180-9-259>
- Zhang J, Liu W, Shi L, Liu X, Wang M, Li W, Yu D, Wang Y, Zhang J, Yun K, Yan J. 2023. The effects of drug addiction and detoxification on the

- human oral microbiota. *Microbiol Spectr* 11:e0396122. <https://doi.org/10.1128/spectrum.03961-22>
32. Gajer P, Brotman RM, Bai G, Sakamoto J, Schütte UME, Zhong X, Koenig SSK, Fu L, Ma ZS, Zhou X, Abdo Z, Forney LJ, Ravel J. 2012. Temporal dynamics of the human vaginal microbiota. *Sci Transl Med* 4:132ra52. <https://doi.org/10.1126/scitranslmed.3003605>
  33. Song SD, Acharya KD, Zhu JE, Deveney CM, Walther-Antonio MRS, Tetel MJ, Chia N. 2020. Daily vaginal microbiota fluctuations associated with natural hormonal cycle, contraceptives, diet, and exercise. *mSphere* 5:e00593-20. <https://doi.org/10.1128/mSphere.00593-20>
  34. Chaban B, Links MG, Jayaprakash TP, Wagner EC, Bourque DK, Lohn Z, Albert AY, van Schalkwyk J, Reid G, Hemmingsen SM, Hill JE, Money DM. 2014. Characterization of the vaginal microbiota of healthy Canadian women through the menstrual cycle. *Microbiome* 2:23. <https://doi.org/10.1186/2049-2618-2-23>
  35. Dash HR, Das S. 2018. Microbial degradation of forensic samples of biological origin: potential threat to human DNA typing. *Mol Biotechnol* 60:141–153. <https://doi.org/10.1007/s12033-017-0052-5>
  36. Carrier TJ, Reitzel AM. 2018. Convergent shifts in host-associated microbial communities across environmentally elicited phenotypes. *Nat Commun* 9:952. <https://doi.org/10.1038/s41467-018-03383-w>
  37. Zhang J, Wang M, Qi X, Shi L, Zhang J, Zhang X, Yang T, Ren J, Liu F, Zhang G, Yan J. 2021. Predicting the postmortem interval of burial cadavers based on microbial community succession. *Forensic Sci Int Genet* 52:102488. <https://doi.org/10.1016/j.fsigen.2021.102488>
  38. Asghiar F, Williams GA. 2019. Evaluating the use of hypoxia sensitive markers for body fluid stain age prediction. *Forensic Sci Int Genet Suppl Ser* 7:644–646. <https://doi.org/10.1016/j.fsigs.2019.10.123>
  39. Varma S, Simon R. 2006. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics* 7:91. <https://doi.org/10.1186/1471-2105-7-91>
  40. Knights D, Kuczynski J, Charlson ES, Zaneveld J, Mozer MC, Collman RG, Bushman FD, Knight R, Kelley ST. 2011. Bayesian community-wide culture-independent microbial source tracking. *Nat Methods* 8:761–763. <https://doi.org/10.1038/nmeth.1650>
  41. Shenhav L, Thompson M, Joseph TA, Briscoe L, Furman O, Bogumil D, Mizrahi I, Pe'er I, Halperin E. 2019. FEAST: fast expectation-maximization for microbial source tracking. *Nat Methods* 16:627–632. <https://doi.org/10.1038/s41592-019-0431-x>
  42. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
  43. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. 2016. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4:e2584. <https://doi.org/10.7717/peerj.2584>
  44. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>
  45. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27:2194–2200. <https://doi.org/10.1093/bioinformatics/btr381>
  46. Prodan A, Tremaroli V, Brolin H, Zwinderman AH, Nieuwdorp M, Levin E. 2020. Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. *PLoS One* 15:e0227434. <https://doi.org/10.1371/journal.pone.0227434>
  47. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41:D590–D596. <https://doi.org/10.1093/nar/gks1219>
  48. Wickham H, Chang W, RStudio. 2016. ggplot2: create elegant data visualisations using the grammar of graphics. *Book of Abstracts*, Cham.
  49. Solymos P, Stevens M, Wagner H. 2013. Package vegan: community ecology package
  50. Dinno A. 2015. Nonparametric pairwise multiple comparisons in independent groups using Dunn's test. *Stata J* 15:292–300. <https://doi.org/10.1177/1536867X1501500117>
  51. Hulliger B. 2006. On two aspects of outlier treatment: univariate vs. multivariate approach and transformation of data. In *Proceedings of Q2006, the European conference on quality in survey statistics*
  52. Liaw A, Wiener MC. 2002. Classification and regression by randomForest. *R news* 2:18–22.