Fang Zhang[1] / Ang Shan[1] / Yihui Luan[1]

A novel method to accurately calculate statistical significance of local similarity analysis for high-throughput time series

[1] School of Mathematics, Shandong University, Jinan, 250100, P.R. China, E-mail: yhluan@sdu.edu.cn

Abstract:
In recent years, a large number of time series microbial community data has been produced in molecular biological studies, especially in metagenomics. Among the statistical methods for time series, local similarity analysis is used in a wide range of environments to capture potential local and time-shifted associations that cannot be distinguished by traditional correlation analysis. Initially, the permutation test is popularly applied to obtain the statistical significance of local similarity analysis. More recently, a theoretical method has also been developed to achieve this aim. However, all these methods require the assumption that the time series are independent and identically distributed. In this paper, we propose a new approach based on moving block bootstrap to approximate the statistical significance of local similarity scores for dependent time series. Simulations show that our method can control the type I error rate reasonably, while theoretical approximation and the permutation test perform less well. Finally, our method is applied to human and marine microbial community datasets, indicating that it can identify potential relationship among operational taxonomic units (OTUs) and significantly decrease the rate of false positives.

Keywords: local similarity analysis, moving block bootstrap, statistical significance

DOI: 10.1515/sagmb-2018-0019

1 Introduction

With the rapid development of molecular biological techniques, particularly next generation sequencing and molecular profiling, a mass of high-throughput sequencing data has been generated efficiently and simultaneously in many ecological studies. Metagenomics is one of the hottest topics of these research areas and is widely used to analyze the composition of microbial communities. Benefitting from the sharp decline in the cost of sequencing, time series data can be applied to investigate the temporal and spatial evolution of microbial communities in marine (Fuhrman et al., 2006; Giovannoni & Vergin, 2012), soil (Barberán et al., 2011; Fierer et al., 2010), human (Palmer et al., 2007; Qin et al., 2010; The Human Microbiome Project Consortium, 2012; Shade et al., 2013) and other environments (Andersson et al., 2014; Faust et al., 2015; Shade et al., 2012; Zhou et al., 2014). Many computational and statistical approaches have been proposed to construct taxon cooccurrence and regulatory networks and to analyze gene expression profiles, ranging from correlation analysis combined with permutation tests (Barberán et al., 2011) and multiple regression (Faust et al., 2012; Trosvik, Stenseth & Rudi, 2010) to similarity measurement based on the hypergeometric distribution (Chaffron et al., 2010). Most of these methods consider correlation between different factors (such as genes in gene regulation studies and organisms and environmental factors in ecological studies) across the entire studied time intervals. However, it remains challenging to uncover potential internal relationships or microbe-environment interactions and to understand the variations in microbial communities because the associations in reality might only occur in subintervals or because a time shift occurs between the response of one factor to a change in another. Under these circumstances, methods based on global relationships across the whole time series(for example, methods using Pearson and Spearman correlation coefficients) may fail to reveal these associations.

Fortunately, several methods have been proposed to investigate these associations, such as IdLCR (Pei et al., 2014) and local similarity analysis(LSA) (Ruan et al., 2006). Analogical to the local alignment algorithm used in DNA sequence alignment (Waterman, 1995), Qian et al. (2001) presented a local similarity method to find potential time-delay and local associations in gene expression profile data. Ruan et al. (2006) applied a dynamic programming algorithm to calculate the local similarity score and identified complex associations among microbial organisms and environmental factors in the San Pedro Channel, North Pacific Ocean. Xia et al. (2011)

Yihui Luan is the corresponding author.

extended LSA to understand the associations of replicated time series data and construct the confidence interval of LSA by bootstrapping. Weiss et al. (2016) compared the performance of eight correlation techniques with simulated and real data in response to challenges to microbiome studies and provided specific recommendations for correlation technique usage. Due to its simplicity and effectiveness, LSA has been used in many fields; for example, gene expression profiling (Ji & Tan, 2004; Balasubramaniyan et al., 2005), regulatory network construction (Madeira et al., 2010), cooccurrence identification pattern (Beman, Steele & Fuhrman, 2011) and others (Xia et al., 2011; Gonçalves & Madeira, 2014; Cram et al., 2015).

In practice, it is very important to calculate the statistical significance of LSA; for this, the permutation test has usually been used. Recently, researchers have developed several methods to approximate the statistical significance of LSA (Durno et al., 2013; Xia et al. 2013; 2015; Lagnoux, Mercier & Vallois, 2017). Durno et al. (2013) provided an asymptotic upper bound for the cumulative probability distribution of the LSA statistic, but this is not tight. Xia et al. (2013) proposed an efficient approach to approximate the statistical significance for LSA of two independent identically distributed (i.i.d.) time series. Extending the theories for the tail probability of the range of the sum of Markovian random variables, formulae were presented to approximate the statistical significance of local trend scores (Xia et al., 2015). Lagnoux, Mercier, and Vallois (2017) investigated the accuracy of some classical results on the local score distribution for an i.i.d. model using a Kolmogorov-Smirnov goodness of fit test and demonstrated that considering the length of the local score improved the detection of truly positive sequences. Nevertheless, these methods are only valid when the time series are i.i.d., which is not true for some biological data. In this article, we recommend a new method based on moving block bootstrap, referred to as MBBLSA, to obtain the statistical significance of LSA. Bootstrapping is generally used to construct a confidence interval but is also a powerful tool for hypothesis testing. The moving block bootstrap, introduced by Künsch (1989) and Liu and Singh (1992), is a variation of the common bootstrap, which can keep the stationary property in each resampled block. Simulations show that MBBLSA can control type I error rate reasonably, but type I error rates for other methods are usually larger than the given threshold.

This paper is organized as follows. In the "Methods" section, we present the definition of LSA and its approximate statistical significance as well as our new method, MBBLSA. In the "Simulations and Results" section, we investigate the type I error rate performance for the different methods via simulations, and the results show that our method is more efficient than others. Then, these methods are applied to analyze the marine bacterial communities and the human microbiome from different high-throughput experiments and compare the associations identified by MBBLSA with those identified from theoretical approximations and the permutation test. Conclusions and future work is discussed in the "Conclusions" Section.

2 Methods

2.1 Local similarity analysis

Let X_t and Y_t, $t = 1, 2, \cdots, n$ be two time series with mean 0 and standard deviation σ_X and σ_Y, respectively. The local similarity score is used to find the most similar interval of the same length from each sequence to maximize the similarity between the two time series, i.e. to seek $I = [i, i + k - 1]$ and $J = [j, j + k - 1]$ with $|i - j| \le D$ such that the absolute value of $S_{i,j,k} = \sum_{l=0}^{k-1} X_{i+l}Y_{j+l}$ is maximized with a predefined time delay D, referred to as the local similarity score ($LS(D)$):

$$LS(D) = \max_{0\le i,j,k\le n;|i-j|\le D} \left| \sum_{l=0}^{k-1} X_{i+l}Y_{j+l} \right|. \tag{1}$$

To determine whether the local similarity score is statistically significant or merely obtained incidentally, the statistical significance (P-value) needs to be calculated. Generally, an accurate distribution of local similarity score cannot be obtained; thus the permutation test can be used here. In particular, the local similarity score of X_t and Y_t is computed, denoted as $LS(D)$. Then, the elements of one time series Y_t are fixed and those of X_t are reshuffled. Assume that $X_t^{(k)}$ is the kth permutation of X_t, the local similarity score of $X_t^{(k)}$, and Y_t is denoted as $LS^{(k)}(D)$. Finally, we repeat the above procedure for a large number of times B. Then, the statistical significance of $LS(D)$ is obtained as follows:

$$P_B = \frac{1}{B} \sum_{k=1}^{B} \mathrm{I}\left[LS^{(k)}(D) \ge LS(D)\right], \tag{2}$$

where $I(\cdot)$ is the indicator function.

Assume that there are N time series; then, $\frac{N(N-1)}{2}$ pairs of time series in total need to be compared. We want to determine whether the correlation of these pairs are statistically significant, which is a multiple testing problem. Conventional hypothesis testing may lead to a large number of false positives. Many methods have been proposed to resolve this problem, such as use of Bonferroni correction, the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995) and the positive false discovery rate (pFDR) (Storey, 2002). In this paper, pFDR is applied to adjust the p-values generated from the permutation test.

2.2 Statistical significance approximation for local similarity analysis

Recently, Xia et al. (2013) proposed a statistical significance approximation for local similarity analysis based on the asymptotic distribution of the range of the partial sums of i.i.d. and Markovian random variables. Specifically, assume that the means of X_t and Y_t are 0. Let $Z_t = X_t Y_t$. For the i.i.d. time series, assume that the variance of X_t and Y_t are both σ^2. For Markov time series, under the null hypothesis that X_t and Y_t are not associated, Z_t are also Markov random variables. Let ϕ be the stationary distribution of Z_t and

$$\sigma^2 = E_\phi(Z_1^2) + 2\sum_{k=1}^{\infty} E_\phi(Z_1 Z_{k+1}). \tag{3}$$

Following the results shown in Xia et al. (2013), the statistical significance of the local similarity score of X_t and Y_t with maximum time delay D is approximately equal to

$$P(LS(D) \geq s_D) = P\left(\frac{LS(D)}{\sigma\sqrt{n}} \geq \frac{s_D}{\sigma\sqrt{n}}\right) = \mathscr{L}_D\left(\frac{s_D}{\sigma\sqrt{n}}\right), \tag{4}$$

where s_D is the local similarity score of X_t and Y_t, and the tail probability distribution function $\mathscr{L}_D(x)$ is defined as

$$\mathscr{L}_D(x) \approx 1 - 8^{2D+1}\left[\sum_{k=1}^{\infty}\left\{\frac{1}{x^2} + \frac{1}{(2k-1)^2\pi^2}\right\}\exp\left\{-\frac{(2k-1)^2\pi^2}{2x^2}\right\}\right]^{2D+1}. \tag{5}$$

As this method is based on the approximate limit distribution of local similarity score, it is efficient only when the sample size is quite large. However, most time series data in biology are very short due to the cost of experiments and sequencing. Moreover, time series at different points must be i.i.d. in the permutation test and the theoretical approximation method. This assumption is usually violated in reality and time series may depend on the values of previous time points. Therefore, a new method to deal with dependent time series is urgently needed.

2.3 Moving block bootstrap

The permutation test mentioned above is a nonparametric resampling test in which resampling is completed without replacement. Next, we propose the bootstrap resampling method with replacement to obtain the statistical significance of the local similarity analysis. To retain the stationarity of the time series, we use a special nonparametric bootstrap method called the moving block bootstrap (MBB). MBB divides time series values into sequences or blocks of consecutive points. As the order of each point in the block is invariable, the bootstrap resamples can retain stationarity within this block, and this can reduce the influence of autocorrelation on the objective statistics. The permutation test, original bootstrap and moving block bootstrap are compared in Figure 1.

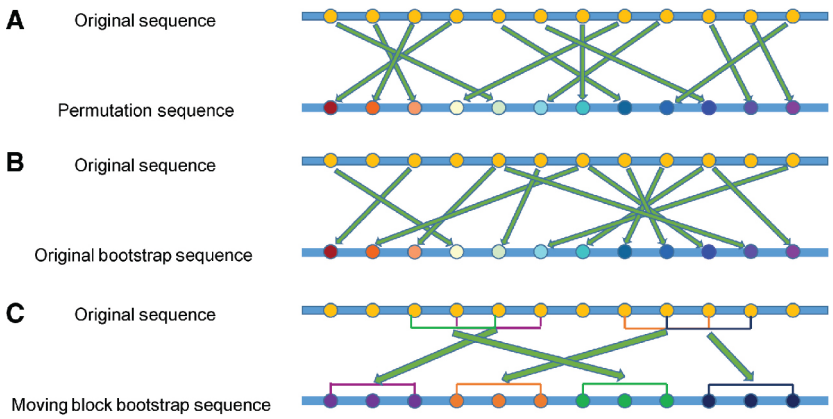

Figure 1: Overview of the differences among the permutation test (**A**), the original bootstrap (**B**) and the moving block bootstrap (**C**). The samples in the permutation test are resampled from the original sequence without replacement but are resampled with replacement in bootstrap method. In the moving block bootstrap, we resampled them from the blocks of same length with replacement.

A method based on MBB is developed to calculate the statistical significance of the local similarity score, and is referred to as the MBBLSA. Given a block length l, as the blocks may overlap, the number of blocks is $n - l + 1$. The MBBLSA procedure is as follows (Mudelsee, 2010):

***step* 1** Calculate the local similarity score of X_t and Y_t, denoted as $LS(D)$.

***step* 2** Set the bootstrap counter $k = 1$.

***step* 3** Start resampling

***step* 3.1** Set counter $c = 1$.

***step* 3.2** Draw random block j^* ($j^* \in 1, 2, \cdots, n - l + 1$) and insert the block data, $\{X_i^{(k)}\}_{i=c}^{c+l-1} = \{X_i\}_{i=j^*}^{j^*+l-1}$. If $X_n^{(k)}$ has been inserted, stop inserting, discard redundant block values and exit.

***step* 3.3** Increase $c \to c + l$, go to *step* 3.2.

***step* 4** Calculate the local similarity score of $X_t^{(k)}$ and Y_t, denoted as $LS^{(k)}(D)$.

***step* 5** Increase counter $k \to k + 1$, then repeat *step* 3 – *step* 4 until $k = B$.

***step* 6** The statistical significance of $LS(D)$ is obtained as follows:

$$P_B = \frac{1}{B} \sum_{k=1}^{B} \mathrm{I}\left[LS^{(k)}(D) \geq LS(D)\right]. \tag{6}$$

The choice of block length l plays a crucial role in deriving the statistical significance because it determines the accuracy of the p-value. Berkowitz and Kilian (2000) described the trade-off problem. If the block size is too small, MBB destroys the time dependency and the average accuracy will decrease. If the block size is too large, there are few blocks and resamples will tend to be similar. Therefore, the average accuracy will also decrease. This means that the block length should not be too small or too large. Sherman, Speed Jr, and Speed (1998) suggest a simple block length selector, which is adapted from Carlstein (1986):

$$l = NI\left\{\left[6^{1/2} \cdot \hat{a} \big/ (1 - \hat{a}^2)\right]^{2/3} \cdot n^{1/3}\right\} \tag{7}$$

where $NI(\cdot)$ is the nearest integer function, and $\hat{a}$ is the estimated autocorrelation coefficient of the AR(1) model fitted to the time series. If time series are generated from an AR(1) model, it can be shown that l in formula (7) is the optimal choice of block length. It is a simplification to utilize this block length selection in realistic problems. However, this simplification produces acceptable results through simulations. In MBBLSA, this block length is applied to step 3.

2.4 Data normalization

Before calculating the local similarity score of raw data, time series need to be normalized. Here we apply the normal score transformation (Xia et al., 2013) to normalize the sequences. Specifically, let R_k be the rank of X_k in $X_1, X_2, \cdots, X_n$. We take $X_k^{'} = \Phi^{-1}\left(\frac{R_k}{n+1}\right)$, where Φ is the cumulative distribution function of the standard normal distribution. In addition, if sample size n is small, the above transformed data do not necessarily follow a standard normal distribution closely, where the mean is not 0 and the variance is not 1. Therefore, we use the traditional Z-score transformation to scale the transformed data again, such that $\widetilde{X}_k = (X_k^{'} - \overline{X^{'}})/\sigma^{'}$, where $\overline{X^{'}}$ is the mean and $\sigma^{'}$ is the standard deviation of $X_t^{'}$. $\widetilde{X}_t$ is taken as the normalization of X_t. Analogically, we utilize the same normalization for Y_t and obtain $\widetilde{Y}_t$. Then we apply the permutation test, the approximate theoretical statistical significance method and MBBLSA to $\widetilde{X}_t$ and $\widetilde{Y}_t$ to compute the statistical significance of the raw data X_t and Y_t.

3 Simulations and Results

In this section, Monte Carlo experiments are carried out to investigate the type I error rate of MBBLSA and four other methods to evaluate the statistical significance of local similarity scores.

3.1 The compared methods

We select the following five different approaches to detect the relationship between time series data:

- **PCC**: The Pearson correlation coefficient (PCC) between X_t and Y_t is calculated as follows:

$$\rho(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \tag{8}$$

  where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ are the sample means of X_t and Y_t, respectively. If the random variables are from a bivariate normal distribution, the statistics

$$t = \rho\sqrt{\frac{n-2}{1-\rho^2}} \tag{9}$$

  satisfies a Student's t-distribution with degrees of freedom $n-2$ under the null hypothesis that the two time series are not correlated.

- **SRCC**: Let $R_X(i) = rank(X_i)$ be the rank of X_i in $X_1, X_2, \cdots, X_n$ and $R_Y(i) = rank(Y_i)$ is defined analogously. The Spearman rank correlation coefficient (SRCC) between X_t and Y_t is defined as follows:

$$\gamma(X,Y) = \frac{\sum_{i=1}^{n}(R_X(i) - \bar{R}_X)(R_Y(i) - \bar{R}_Y)}{\sqrt{\sum_{i=1}^{n}(R_X(i) - \bar{R}_X)^2}\sqrt{\sum_{i=1}^{n}(R_Y(i) - \bar{R}_Y)^2}} \tag{10}$$

  where $\bar{R}_X = \frac{1}{n}\sum_{i=1}^{n} R_X(i)$ and $\bar{R}_Y = \frac{1}{n}\sum_{i=1}^{n} R_Y(i)$ are the sample means of R_X and R_Y, respectively. Similar to PCC, under the null hypothesis, the distribution of

$$t = \gamma\sqrt{\frac{n-2}{1-\gamma^2}} \tag{11}$$

  follows Student's t-distribution with degrees of freedom $n-2$.

- **Permutation test**: shown in Section 2.1.
- **Theoretical LSA (TLSA)**: shown in Section 2.2.
- **MBBLSA**: shown in Section 2.3.

3.2 Type I error rates for the various approaches

To investigate whether p-values obtained from these statistics are close to the nominal level, three models are used to compare the type I error rates of the five approaches mentioned above:

(1) AR(1) model:

$$\begin{aligned} X_t &= \rho_1 X_{t-1} + \varepsilon_t^X \\ Y_t &= \rho_2 Y_{t-1} + \varepsilon_t^Y \end{aligned} \tag{12}$$

(2) ARMA(1,1) model:

$$\begin{aligned} X_t &= \rho_1 X_{t-1} + \varepsilon_t^X + 0.5\varepsilon_{t-1}^X \\ Y_t &= \rho_2 Y_{t-1} + \varepsilon_t^Y + 0.5\varepsilon_{t-1}^Y \end{aligned} \tag{13}$$

(3) ARMA(1,1)-TAR(1) model:

$$\begin{aligned} X_t &= \rho_1 X_{t-1} + \varepsilon_t^X + 0.5\varepsilon_{t-1}^X \\ Y_t &= \begin{cases} \rho_2 Y_{t-1} + \varepsilon_t^Y, Y_{t-1} \le -1 \\ 0.5 Y_{t-1} + \varepsilon_t^Y, Y_{t-1} > -1 \end{cases} \end{aligned} \tag{14}$$

where $0 < |\rho_1|, |\rho_2| < 1$ and $\varepsilon_t^X, \varepsilon_t^Y$ are independent standard normal random variables. All models are stationary. For each model, we first generated X_1, Y_1 from the standard normal distribution. Then X_t and Y_t, $t = 2, \cdots, 100 + n$ were generated according to these models. Finally, we discarded the first 100 samples and took the others as the true X_t and Y_t. This procedure can guarantee the stationarity of the time series.

We investigate the effect of the autoregressive coefficients ρ_1, ρ_2 and sample size n on the type I error rate in these models. Specifically, we select six different pairs of autoregressive coefficients from -0.5 to 0.8 and sample sizes from 20 to 300. The results are shown in Table 1, Table 2 and Table 3 corresponding to AR(1) (Eq. 12), ARMA(1,1) (Eq. 13) and ARMA(1,1)-TAR(1) (Eq. 14) models, respectively. For simplicity, we let the time delay $D = 0$. Throughout the simulations, we set the prespecified error rate 0.05.

Table 1: Empirical type I error rate for different methods (the third to seventh columns) in the AR(1) model.

ρ_1, ρ_2	n	PCC	SRCC	TLSA	Permutation	MBBLSA
	20	0.1112	0.1092	0.0382	0.1181	0.0601
	40	0.1172	0.1163	0.0732	0.1382	0.0613
−0.5 −0.5	60	0.1229	0.1194	0.0918	0.1468	0.0621
	80	0.1281	0.1231	0.1080	0.1594	0.0642
	100	0.1315	0.1261	0.1183	0.1655	0.0618
	200	0.1250	0.1216	0.1387	0.1714	0.0619
	300	0.1321	0.1282	0.1498	0.1807	0.0623
	20	0.0521	0.0498	0.0104	0.0500	0.0507
	40	0.0554	0.0512	0.0206	0.0530	0.0521
0 0	60	0.0509	0.0513	0.0238	0.0505	0.0501
	80	0.0511	0.0517	0.0292	0.0493	0.0507
	100	0.0490	0.0499	0.0284	0.0475	0.0478
	200	0.0498	0.0481	0.0340	0.0475	0.0493
	300	0.0497	0.0523	0.0392	0.0514	0.0522
	20	0.0720	0.0685	0.0177	0.0698	0.0605
	40	0.0765	0.0726	0.0331	0.0746	0.0600
0.3 0.3	60	0.0735	0.0743	0.0425	0.0791	0.0604
	80	0.0726	0.0694	0.0458	0.0766	0.0581
	100	0.0725	0.0716	0.0533	0.0808	0.0591
	200	0.0699	0.0691	0.0615	0.0818	0.0566
	300	0.0713	0.0718	0.0644	0.0799	0.0558
	20	0.0848	0.0808	0.0225	0.0811	0.0649
	40	0.0904	0.0876	0.0423	0.0915	0.0636
0.3 0.5						

	60	0.0863	0.0834	0.0546	0.0961	0.0642
	80	0.0880	0.0835	0.0589	0.0967	0.0619
	100	0.0881	0.0828	0.0665	0.1022	0.0625
	200	0.0936	0.0906	0.0843	0.1101	0.0632
	300	0.0904	0.0901	0.0903	0.1110	0.0583
	20	0.1037	0.1006	0.0310	0.1049	0.0694
	40	0.1152	0.1126	0.0644	0.1328	0.0685
0.5 0.5	60	0.1232	0.1183	0.0853	0.1400	0.0676
	80	0.1274	0.1168	0.0951	0.1470	0.0620
	100	0.1273	0.1200	0.1070	0.1516	0.0682
	200	0.1255	0.1199	0.1365	0.1701	0.0630
	300	0.1279	0.1252	0.1480	0.1806	0.0647
	20	0.1591	0.1460	0.0527	0.1517	0.0845
	40	0.1818	0.1744	0.1171	0.1976	0.0820
0.5 0.8	60	0.1877	0.1735	0.1524	0.2280	0.0837
	80	0.1906	0.1788	0.1770	0.2420	0.0786
	100	0.1886	0.1792	0.1904	0.2555	0.0762
	200	0.1997	0.1927	0.2477	0.2956	0.0761
	300	0.1991	0.1887	0.2688	0.3146	0.0681

The first and second columns represent different autoregressive coefficients and sample sizes, respectively. The number of permutations is 1000, the nominal size is 0.05, and the number of replications is 10,000.

Table 2: Empirical type I error rate for different methods (the third to seventh columns) in the ARMA(1,1) model.

ρ_1, ρ_2	n	PCC	SRCC	TLSA	Permutation	MBBLSA
	20	0.0485	0.0516	0.0090	0.0487	0.0478
	40	0.0503	0.0501	0.0194	0.0476	0.0496
−0.5 −0.5	60	0.0507	0.0515	0.0247	0.0507	0.0523
	80	0.0454	0.0463	0.0254	0.0460	0.0460
	100	0.0524	0.0504	0.0314	0.0522	0.0520
	200	0.0506	0.0502	0.0357	0.0504	0.0505
	300	0.0469	0.0482	0.0343	0.0479	0.0488
	20	0.0884	0.0875	0.0239	0.0929	0.0692
	40	0.0885	0.0826	0.0445	0.0956	0.0632
0 0	60	0.0875	0.0790	0.0520	0.0956	0.0577
	80	0.0880	0.0825	0.0612	0.1011	0.0586
	100	0.0882	0.0851	0.0634	0.0998	0.0583
	200	0.0880	0.0848	0.0815	0.1059	0.0568
	300	0.0912	0.0877	0.0920	0.1172	0.0579
	20	0.1363	0.1282	0.0459	0.1375	0.0761
	40	0.1401	0.1298	0.0847	0.1574	0.0681
0.3 0.3	60	0.1403	0.1325	0.1090	0.1757	0.0649
	80	0.1383	0.1323	0.1165	0.1775	0.0602
	100	0.1401	0.1368	0.1356	0.1895	0.0638
	200	0.1350	0.1297	0.1539	0.1931	0.0564
	300	0.1380	0.1339	0.1732	0.2079	0.0554
	20	0.1616	0.1489	0.0584	0.1666	0.0813
	40	0.1625	0.1514	0.1074	0.1929	0.0705
0.3 0.5	60	0.1647	0.1548	0.1297	0.1998	0.0686
	80	0.1667	0.1596	0.1524	0.2178	0.0703
	100	0.1659	0.1570	0.1583	0.2188	0.0643
	200	0.1662	0.1581	0.1942	0.2436	0.0615
	300	0.1663	0.1599	0.2220	0.2633	0.0619
	20	0.1842	0.1682	0.0749	0.1832	0.0814
	40	0.1919	0.1831	0.1406	0.2296	0.0719
0.5 0.5	60	0.2007	0.1942	0.1773	0.2548	0.0691
	80	0.2051	0.1974	0.1998	0.2761	0.0664
	100	0.2012	0.1926	0.2126	0.2806	0.0661
	200	0.2016	0.1935	0.2668	0.3180	0.0620
	300	0.2012	0.1937	0.2827	0.3265	0.0574
	20	0.2471	0.2261	0.1119	0.2424	0.1017
	40	0.2616	0.2475	0.1942	0.3048	0.0850
0.5 0.8	60	0.2558	0.2430	0.2426	0.3414	0.0783

80	0.2708	0.2584	0.2862	0.3719	0.0804
100	0.2620	0.2522	0.3050	0.3830	0.0705
200	0.2737	0.2616	0.3842	0.4456	0.0700
300	0.2624	0.2539	0.4056	0.4573	0.0675

The first and second columns represent different autoregressive coefficients and sample sizes, respectively. The number of permutations is 1000, the nominal size is 0.05, and the number of replications is 10,000.

As shown from Table 1, MBBLSA performs best among the different autoregressive coefficients and sample sizes. The type I error rate of MBBLSA is closer to the prespecified error rate than other methods except for the case of $\rho_1 = 0$ and $\rho_2 = 0$. In this case, all methods work well, as time series are independently bivariate normally distributed. Moreover, the type I error rate of TLSA is somewhat smaller than the nominal size 0.05, indicating that TLSA is conservative, consistent with the findings of Xia et al. (2013). As the statistical significance from TLSA is obtained from the limit distribution of the local similarity score, the approximation underestimates the true significance when the sample size is small. If $\rho_1 \neq 0$, $\rho_2 \neq 0$, i.e. autocorrelation exists in the raw data, the type I error rates of PCC, SRCC and the permutation test are larger than the prespecified error rate even when the sample size is small. Although the type I error rate of TLSA is close to 0.05 for small sample sizes due to the conservative property of TLSA mentioned above, it is greater than 0.05 when the sample size is large. This result demonstrates that the significance of TLSA is not stable across different sample sizes. However, regardless of whether time series are independent, MBBLSA is stable and controls the type I error rate reasonably well.

Analogical results are shown in Table 2 and Table 3 for the ARMA(1,1) and ARMA(1,1)-TAR(1) models, respectively. We find that the type I error rates of PCC, SRCC and the permutation test with $\rho_1 = -0.5$ and $\rho_2 = -0.5$ are also close to the prespecified error rate. In this scenario, X_t in the ARMA(1,1) model are i.i.d.. Therefore, PCC, SRCC and the permutation test can also work well. However, the type I error rates of PCC, SRCC and the permutation test are much larger than the prespecified error rate of 0.05 under all other situations, while the type I error rate of MBBLSA is well controlled under all situations. It is interesting to note that the type I error rate of MBBLSA declines as the sample size increases, especially when the autocorrelation is strong, indicating that MBBLSA can capture more inherent properties when the sample size is large. Finally, simulations for time delay $D \neq 0$ are presented in the supplementary materials.

Table 3: Empirical type I error rate for different methods (the third to seventh columns) in the ARMA(1,1)-TAR(1) model.

ρ_1, ρ_2	n	PCC	SRCC	TLSA	Permutation	MBBLSA
	20	0.0497	0.0502	0.0109	0.0500	0.0525
	40	0.0505	0.0492	0.0198	0.0496	0.0503
−0.5 −0.5	60	0.0488	0.0490	0.0250	0.0496	0.0509
	80	0.0494	0.0513	0.0270	0.0481	0.0488
	100	0.0487	0.0499	0.0304	0.0518	0.0514
	200	0.0537	0.0507	0.0388	0.0532	0.0524
	300	0.0532	0.0538	0.0397	0.0521	0.0532
	20	0.0785	0.0784	0.0188	0.0785	0.0587
	40	0.0756	0.0742	0.0379	0.0819	0.0593
0 0	60	0.0796	0.0771	0.0490	0.0878	0.0594
	80	0.0787	0.0768	0.0520	0.0875	0.0555
	100	0.0851	0.0814	0.0627	0.0954	0.0585
	200	0.0827	0.0810	0.0745	0.1008	0.0565
	300	0.0849	0.0802	0.0784	0.0997	0.0568
	20	0.1036	0.0994	0.0327	0.1037	0.0640
	40	0.1121	0.1106	0.0610	0.1267	0.0654
0.3 0.3	60	0.1133	0.1119	0.0793	0.1311	0.0612
	80	0.1163	0.1123	0.0908	0.1426	0.0618
	100	0.1214	0.1158	0.0984	0.1440	0.0603
	200	0.1193	0.1188	0.1214	0.1577	0.0585
	300	0.1200	0.1158	0.1327	0.1615	0.0557
	20	0.1154	0.1102	0.0371	0.1137	0.0676
	40	0.1276	0.1225	0.0740	0.1370	0.0674
0.3 0.5	60	0.1285	0.1204	0.0912	0.1503	0.0611
	80	0.1326	0.1255	0.1048	0.1595	0.0636
	100	0.1245	0.1196	0.1098	0.1589	0.0592
	200	0.1369	0.1259	0.1400	0.1751	0.0581
	300	0.1350	0.1275	0.1545	0.1854	0.0573
	20	0.1422	0.1316	0.0480	0.1392	0.0791
	40	0.1519	0.1414	0.0944	0.1710	0.0686
0.5 0.5						

	60	0.1615	0.1552	0.1299	0.1968	0.0698
	80	0.1542	0.1427	0.1363	0.1949	0.0616
	100	0.1584	0.1527	0.1527	0.2071	0.0638
	200	0.1589	0.1520	0.1827	0.2252	0.0569
	300	0.1604	0.1516	0.2004	0.2395	0.0566
0.5 0.8	20	0.1830	0.1657	0.0664	0.1729	0.0801
	40	0.2020	0.1870	0.1393	0.2333	0.0776
	60	0.2180	0.2023	0.1809	0.2608	0.0769
	80	0.2110	0.2010	0.2048	0.2792	0.0707
	100	0.2102	0.1961	0.2183	0.2893	0.0668
	200	0.2090	0.1982	0.2694	0.3250	0.0620
	300	0.2146	0.2023	0.3060	0.3498	0.0599

The first and second columns represent different autoregressive coefficients and sample sizes, respectively. The number of permutations is 1000, the nominal size is 0.05, and the number of replications is 10,000.

As time cost of LSA only depends on the length of time series, simulations were performed to obtain the time cost for different sample sizes in the AR(1) model with different methods to evaluate the computing complexity. For each sample size, 10,000 simulations were implemented on a 'Lenovo, Intel Core i7-3770 CPU, 3.4 GHz, 4096 MB RAM' computer, and the average time cost per simulation is shown in Table 4. PCC, SRCC and TLSA consume less time but do not control the type I error well. For the permutation test and MBBLSA, each simulation runs LSA procedures 1000 times as the number of permutations or bootstraps is 1000. Therefore, these methods require much more time than the former three methods. As shown in Table 4, the permutation test has a lower time cost than MBBLSA, but its type I error rate is larger than the prespecified error rate, as mentioned above. Moreover, the growth rate of the time cost of MBBLSA is less than that of the permutation test, demonstrating that MBBLSA will perform better in terms of computing complexity when the time series length is large.

Table 4: The average time cost for different sample sizes for the five methods of PCC, SRCC, TLSA, the Permutation test and MBBLSA in 10,000 simulations.

n	PCC	SRCC	TLSA	Permutation	MBBLSA
20	1.61×10^{-4}	1.98×10^{-4}	4.23×10^{-4}	0.41	3.46
40	1.60×10^{-4}	2.11×10^{-4}	5.11×10^{-4}	0.51	3.50
60	1.59×10^{-4}	2.14×10^{-4}	6.05×10^{-4}	0.65	3.84
80	1.61×10^{-4}	2.17×10^{-4}	7.19×10^{-4}	0.79	4.50
100	1.77×10^{-4}	2.42×10^{-4}	9.60×10^{-4}	0.97	4.79
200	1.72×10^{-4}	2.51×10^{-4}	2.32×10^{-3}	1.92	6.60
300	2.02×10^{-4}	3.30×10^{-4}	4.84×10^{-3}	3.82	9.55

3.3 SPOT dataset

The San Pedro Ocean Time Series (SPOT) dataset was collected following the deep chlorophyll maximum depth at the University of Southern California Microbial Observatory off the coast of Southern California, Los Angeles, USA (Steele et al., 2011). The bacterial community was analyzed using the ARISA molecular finger printing technique. The dataset consists of 10-year monthly sampled operational taxonomic unit (OTU) abundance time series, including 114 time points without replicates. Under the condition that OTU occurs at least 20 times with a minimum relative abundance of 1% and has less than 10 missing values, we selected 31 abundant OTUs from the SPOT dataset. In addition, we chose 8 environmental factors to identify associations between these factors and OTUs. To fill in missing values in the raw data, we applied linear interpolation.

First, the Box-Ljung test (Ljung & Box, 1978) was used to test whether autocorrelation exists in the time series. Given the significance level of 5%, 20 OTUs and 7 environment factors are significantly autocorrelated among all 39 factors, demonstrating that autocorrelation is a widespread statistical property in biological time series. Second, we applied TLSA, the permutation test and MBBLSA with a time delay of $D = 3$ to identify potential local and time-shifted pairwise associations among all OTUs or between environment factors and OTUs, for a total of 713 pairs. In the permutation test and MBBLSA, the number of replicates used was 1000. After obtaining p-values, we utilized pFDR to obtain the modified p-values and q-values using the *qvalue* package in R software (Storey et al., 2015). For a given threshold p-value $P \leq 0.05$ and q-value $Q \leq 0.05$, 133, 243 and 184 pairs of significant associations are found using TLSA, the permutation test and MBBLSA, respectively (Table

5). All significant pairs found by TLSA and MBBLSA are also found using the permutation test, while only 2 pairs found by TLSA are not found to be significant by MBBLSA. Figure 2A illustrates the relationship among the significant associations identified by the three methods, indicating that TLSA is a little conservative if the sample size is small, consistent with that the discussion above. Moreover, 57 pairs that are found significant by the permutation test are not found significant by MBBLSA; these may be false positives caused by autocorrelation in the data. The results shown in Table 5 are obtained when the stricter criteria $P \leq 0.01$ and $Q \leq 0.01$ are applied.

Table 5: Significant associations found by TLSA, the permutation test and MBBLSA with different thresholds in real datasets.

Dataset # of factors		SPOT 39	MPHM tongue 41	MPHM feces 45	PML 100
$P \leq 0.05$	TLSA	133	317	469	555
$Q \leq 0.05$	Permutation	243	394	527	1698
	MBBLSA	184	351	437	1068
$P \leq 0.01Q$	TLSA	74	222	353	303
≤ 0.01	Permutation	146	293	404	850
	MBBLSA	91	263	285	474

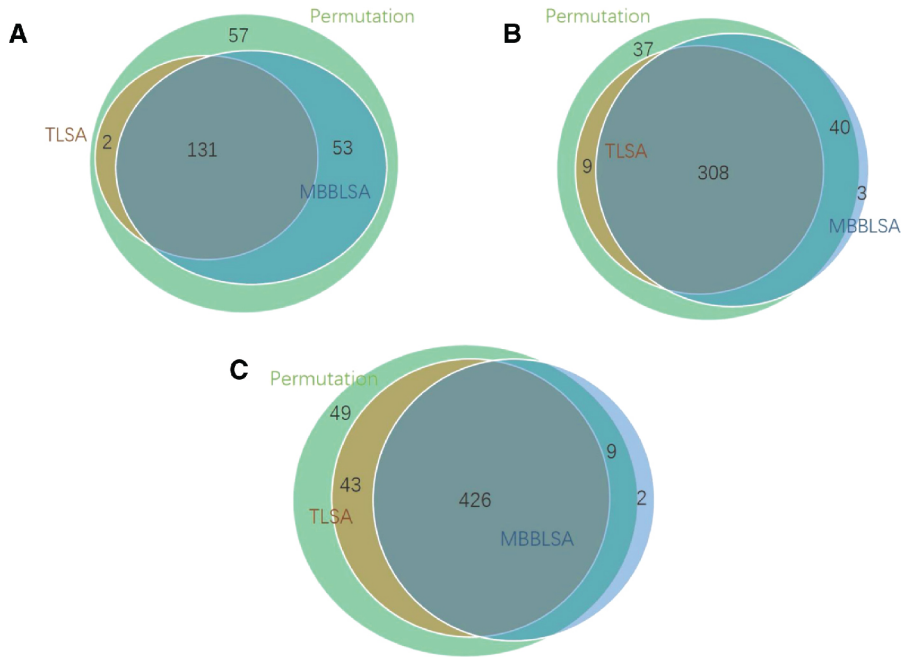

Figure 2: Venn diagram of the relationship among significant pairs found by TLSA, the permutation test and MBBLSA in the SPOT (**A**), MPHM tongue (**B**) and MPHM feces (**C**) datasets. Brown, green and blue colors represent the number of pairs found by TLSA, the permutation test and MBBLSA, respectively.

3.4 MPHM dataset

The Moving Pictures of the Human Microbiome (MPHM) dataset was obtained for two healthy subjects, one male ('M3') and one female ('F4'), both of whom were sampled daily at three body sites (feces, tongue, and skin (left and right palms)) (Caporaso et al., 2011). The set are 130, 135 and 133 daily samples from the feces, tongue and palm of 'F4', and 332, 372 and 357 samples from the corresponding sites of 'M4'. At the taxonomic level of Genus, 335, 373 and 1295 unique operational taxonomic units (OTUs) are found for the feces, tongue and palm (both left and right) sites of 'F4' and 'M3'. Based on the discussion in Caporaso et al. (2011), we only consider 'core' OTUs that were observed in at least 60% of samples taken at the same body site of the same person. For computational feasibility, we selected 41 and 45 core OTUs for the tongue and feces of 'F4' to analyze the relationships between them.

Similar to the SPOT dataset, we first inspected the autocorrelation property in the MPHM dataset. At a significance level of 5%, there are 21 (~51%) and 32 (~71%) significantly autocorrelated OTUs in the tongue and

feces of 'F4', respectively. This results shows that more than 50% of the OTUs are significantly autocorrelated, which verifies the prevalence of autocorrelation in the microbiological community once again. Next, the type I error rate and pFDR threshold were set as 0.05 and 0.01, respectively, to compare the performance of TLSA, the permutation test and MBBLSA with a time delay $D = 3$, as shown in Table 5. When the p-value $P \leq 0.05$ and the q-value $Q \leq 0.05$, 317 (~38.7%), 394 (~48%) and 351 (~42.8%) pairs of significant associations are found among all 820 OTU pairs for the tongue by TLSA, the permutation test and MBBLSA, respectively. However, 469 (~47.3%), 527 (~53.2%) and 437 (~44.1%) pairs of significant associations are found among all 990 OTU pairs for feces. The percentage of significant pairs of OTUs for feces is greater than that in the tongue, possibly because there are more autocorrelated OTUs for feces than those in the tongue. Based on our simulations, autocorrelation may cause false positives to increase in number. However, the use of MBBLSA can avoid the influence of autocorrelation. Venn diagrams showing significant associations of OTUs for the tongue and feces are shown in Figure 2B and C, respectively. Similar to the result obtained with the SPOT dataset, the associations found using the permutation test are almost found by TLSA and MBBLSA. It should be pointed out that the percentage (9.2%) of pairs that are not significant according to MBBLSA among all those found using TLSA for the feces is larger than that (2.8%) found for the tongue, because the strong autocorrelation property in the feces data increases the type I error rate when using TLSA.

3.5 PML dataset

The Plymouth Marine Laboratory (PML) dataset is one of the longest microbial time series that was sampled monthly using high-resolution 16S rRNA tagNGS sequencing at a temperate marine coastal site off Plymouth, UK, for a total of 72 time points (Gilbert et al., 2012). In total, 8794 different bacterioplankton OTUs were identified. We chose 100 abundant OTUs that were present in at least 60% of the time points for the taxonomic level of Genus to analyze their association network.

Among 100 OTUs in the PML dataset, 30 are significantly autocorrelated according to the Box-Ljung test at the significance level of 0.05. As reported in Gilbert et al. (2012), it is obvious that particular OTUs had strong repeatable seasonal patterns. For example, significant seasonal cycles are shown in the autocorrelograms of *Bac*53 and *Bac*54 (Figure 3), and both of their periods are almost 1 year. The first-order autocorrelations of *Bac*53 and *Bac*54 are 0.50 (P-value = 1.84×10^{-5}) and 0.32 (P-value = 0.006), and both are significantly autocorrelated.

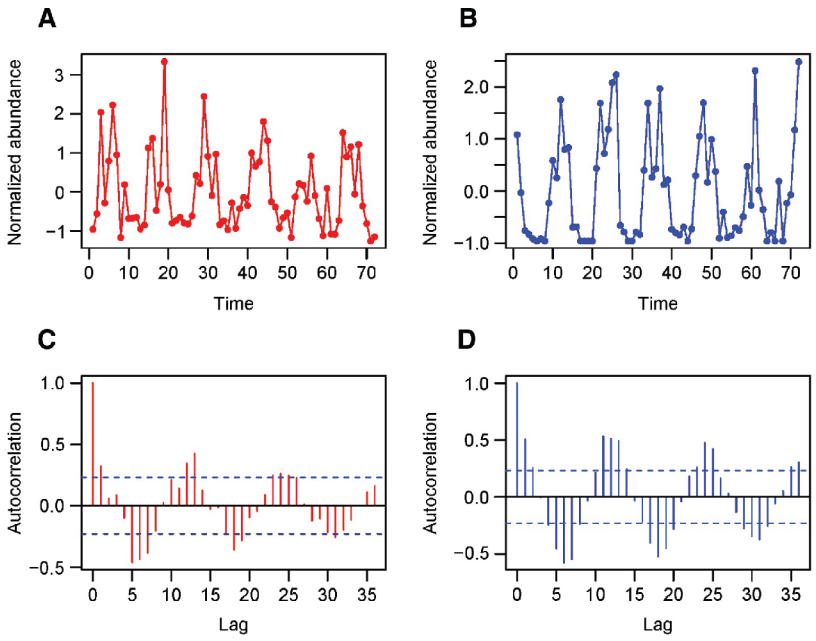

Figure 3: Standardized abundances of *Bac*53 (**A**) and *Bac*54 (**B**) in the PML dataset. The autocorrelograms (**C**,**D**) show the autocorrelation of two time series, each responding to itself, for different lags. Note the significant seasonal variations in the plot of OTUs and their autocorrelograms throughout the 6-year period. The dashed line represents the critical value of statistics $\pm 1.96/\sqrt{n}$, where n is the sample size of the time series. These intervals give the pointwise acceptance area for testing the null hypothesis that the autocorrelation functions of time series are zero at the 5% significance level.

Among 4950 pairwise associations of all 100 OTUs, 555, 1698 and 1068 pairs are significant when using a time delay of $D = 3$ according to TLSA, the permutation test and MBBLSA for a p-value of $P \leq 0.05$ and a q-value $Q \leq 0.05$ (Table 5). It is obvious that fewer significant pairs are found by TLSA, as TLSA is conservative when

the number of samples is small. All pairs found by TLSA are significant according to the permutation test and MBBLSA. All associations found by MBBLSA are also significant according to the permutation test, indicating that the autocorrelation of OTUs appears to result in many false positives when using the permutation test. This suggests that MBBLSA results are in high agreement with those for the permutation test and that the number of false positives can be efficiently reduced.

4 Conclusions

Recent advances in high-throughput sequencing have led to a rapid development in sequencing data and longitudinal studies, investigating variations in microbial communities in a diversity of environments. Many statistical methods have been proposed to gain insight into the temporal and spatial dynamics of biological systems. Compared with i.i.d. data, time series data have many inherent properties, such as autocorrelation; therefore, it is necessary to consider these properties in metagenomics and longitudinal studies.

In this paper, we presented the moving block bootstrap method MBBLSA to examine the statistical significance obtained using local similarity analysis for dependent time series data. Three different dependent time series models were considered to compare the type I error rate performance of MBBLSA with that of other methods, i.e. PCC, SRCC, TLSA and the permutation test. Simulations showed that MBBLSA can control the type I error rate, but the other methods have poorer performance, especially when the sample size is large and autocorrelation exists; the results revealed that only MBBLSA is effective for dependent time series. When the sample size is small, TLSA is a little conservative, but MBBLSA works well. For real datasets obtained in metagenomics studies of microbial communities, MBBLSA could decrease the superfluous relationship when compared with the permutation test and identify more possible associations among OTUs than TLSA.

However, a disadvantage of MBBLSA is that it is time consuming. It is crucial to determine the theoretical statistical significance when using local similarity analysis for dependent time series, and this will be one of our future work. In summary, MBBLSA is an effective statistical approach for detecting potential associations in various biological fields.

Software

The MBBLSA R source code is available at https://github.com/BlueStamford/MBBLSA.

Funding

The research was supported by the Natural Science Foundation of China Grants (Funder Id: 10.13039/501100001809, 11371227, 61432010, 11626247).

Appendix A. Supplementary Materials

The type I error rate performance of three models with different time delays are shown in Supplementary Materials.

References

Andersson, M. G. I., M. Berga, E. S. Lindström and S. Langenheder (2014): "The spatial structure of bacterial communities is influenced by historical environmental conditions," Ecology, 95, 1134–1140.

Balasubramaniyan, R., E. Hüllermeier, N. Weskamp and J. Kämper (2005): "Clustering of gene expression data using a local shape-based similarity measure," Bioinformatics, 21, 1069–1077.

Barberán, A., S. T. Bates, E. O. Casamayor and N. Fierer (2011): "Using network analysis to explore co-occurrence patterns in soil microbial communities," ISME J., 6, 343–351.

Beman, J. M., J. A. Steele and J. A. Fuhrman (2011): "Co-occurrence patterns for abundant marine archaeal and bacterial lineages in the deep chlorophyll maximum of coastal california," ISME J., 5, 1077–1085.

Benjamini, Y. and Y. Hochberg (1995): "Controlling the false discovery rate: A practical and powerful approach to multiple testing," J. R. Stat. Soc. B, 57, 289–300.
Berkowitz, J. and L. Kilian (2000): "Recent developments in bootstrapping time series," Econom. Rev., 19, 1–48.
Caporaso, J. G., C. L. Lauber, E. K. Costello, D. Berg-Lyons, A. Gonzalez, J. Stombaugh, D. Knights, P. Gajer, J. Ravel, N. Fierer, J. I. Gordon and R. Knight (2011): "Moving pictures of the human microbiome," Genome Biol., 12, R50.
Carlstein, E. (1986): "The use of subseries values for estimating the variance of a general statistic from a stationary sequence," Ann. Stat., 14, 1171–1179.
Chaffron, S., H. Rehrauer, J. Pernthaler and C. von Mering (2010): "A global network of coexisting microbes from environmental and whole-genome sequence data," Genome Res., 20, 947–959.
Cram, J. A., L. C. Xia, D. M. Needham, R. Sachdeva, F. Sun and J. A. Fuhrman (2015): "Cross-depth analysis of marine bacterial networks suggests downward propagation of temporal changes," ISME J., 9, 2573–2586.
Durno, W. E., Hanson, N. W., Konwar, K. M & Hallam, S. J. 2013, 'Expanding the boundaries of local similarity analysis', *BMC Genomics*, vol. 14, pp. S3–.
Faust, K., J. F. Sathirapongsasuti, J. Izard, N. Segata, D. Gevers, J. Raes and C. Huttenhower (2012): "Microbial co-occurrence relationships in the human microbiome," PLOS Comput. Biol., 8, 1–17.
Faust, K., L. Lahti, D. Gonze, W. M. de Vos and J. Raes (2015): "Metagenomics meets time series analysis: unraveling microbial community dynamics," Curr. Opin. Microbiol., 25, 56–66.
Fierer, N., D. Nemergut, R. Knight and J. M. Craine (2010): "Changes through time: integrating microorganisms into the study of succession," Res. Microbiol., 161, 635–642.
Fuhrman, J. A., I. Hewson, M. S. Schwalbach, J. A. Steele, M. V. Brown and S. Naeem (2006): "Annually reoccurring bacterial communities are predictable from ocean conditions," Proc. Natl. Acad. Sci. USA, 103, 13104–13109.
Gilbert, J. A., J. A. Steele, J. G. Caporaso, L. Steinbrück, J. Reeder, B. Temperton, S. Huse, A. C. McHardy, R. Knight, I. Joint, P. Somerfield, J. A. Fuhrman and D. Field (2012): "Defining seasonal marine microbial community dynamics," ISME J., 6, 298–308.
Giovannoni, S. J. and K. L. Vergin (2012): "Seasonality in ocean microbial communities," Science, 335, 671–676.
Gonçalves, J. and S. Madeira (2014): "Latebiclustering: Efficient heuristic algorithm for time-lagged bicluster identification," IEEE/ACM T. Comput. Bi, 11, 801–813.
Ji, L. and K.-L. Tan (2004): "Mining gene expression data for positive and negative co-regulated gene clusters," Bioinformatics, 20, 2711–2718.
Künsch, H. R. (1989): "The jackknife and the bootstrap for general stationary observations," Ann. Stat., 17, 1217–1241.
Liu, R. Y. and K. Singh (1992): Moving blocks jackknife and bootstrap capture weak dependence, New York: John Wiley, pp. 225–248.
Lagnoux, A., S. Mercier, P. Vallois (2017): "Statistical significance based on length and position of the local score in a model of i.i.d. sequences," Bioinformatics, 33, 654–660.
Ljung, G. M. and G. E. P. Box (1978): "On a measure of lack of fit in time series models," Biometrika, 65, 297–303.
Madeira, S. C., M. C. Teixeira, I. Sa-Correia and A. L. Oliveira (2010): "Identification of regulatory modules in time series gene expression data using a linear time biclustering algorithm," IEEE/ACM T. Comput. Bi, 7, 153–165.
Mudelsee, M. (2010): Climate Time Series Analysis: Classical Statistical and Bootstrap Methods, Dordrecht: Atmospheric and Oceanographic Sciences Library, Springer.
Palmer, C., E. M. Bik, D. B. DiGiulio, D. A. Relman and P. O. Brown (2007): "Development of the human infant intestinal microbiota," PLOS Biol., 5, 1–18.
Pei, Y., Q. Gao, J. Li and X. Zhao (2014): "Identifying local co-regulation relationships in gene expression data," J. Theor. Biol., 360, 200–207.
Qian, J., M. Dolled-Filhart, J. Lin, H. Yu and M. Gerstein (2001): "Beyond synexpression relationships: local clustering of time-shifted and inverted gene expression profiles identifies new, biologically relevant interactions11edited by f. cohen," J. Mol. Biol., 314, 1053–1066.
Qin, J., R. Li, J. Raes, M. Arumugam, K. S. Burgdorf, C. Manichanh, T. Nielsen, N. Pons, F. Levenez, T. Yamada, D. R. Mende, J. Li, J. Xu, S. Li, D. Li, J. Cao, B. Wang, H. Liang, H. Zheng, Y. Xie, J. Tap, P. Lepage, M. Bertalan, J.-M. Batto, T. Hansen, D. Le Paslier, A. Linneberg, H. B. Nielsen, E. Pelletier, P. Renault, T. Sicheritz-Ponten, K. Turner, H. Zhu, C. Yu, S. Li, M. Jian, Y. Zhou, Y. Li, X. Zhang, S. Li, N. Qin, H. Yang, J. Wang, S. Brunak, J. Doré, F. Guarner, K. Kristiansen, O. Pedersen, J. Parkhill, J. Weissenbach, M. Consortium, P. Bork, S. D. Ehrlich and J. Wang (2010): "A human gut microbial gene catalogue established by metagenomic sequencing," Nature, 464, 59–65.
Ruan, Q., D. Dutta, M. S. Schwalbach, J. A. Steele, J. A. Fuhrman and F. Sun (2006): "Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors," Bioinformatics, 22, 2532–2538.
Shade, A., J. S. Read, N. D. Youngblut, N. Fierer, R. Knight, T. K. Kratz, N. R. Lottig, E. E. Roden, E. H. Stanley, J. Stombaugh, R. J. Whitaker, C. H. Wu and K. D. McMahon (2012): "Lake microbial communities are resilient after a whole-ecosystem disturbance," ISME J., 6, 2153–2167.
Shade, A., J. Gregory Caporaso, J. Handelsman, R. Knight and N. Fierer (2013): "A meta-analysis of changes in bacterial and archaeal communities with time," ISME J., 7, 1493–1506.
Sherman, M., F. M. Speed Jr and F. M. Speed (1998): "Analysis of tidal data via the blockwise bootstrap," J. Appl. Stat., 25, 333–340.
Steele, J. A., P. D. Countway, L. Xia, P. D. Vigil, J. M. Beman, D. Y. Kim, C.-E. T. Chow, R. Sachdeva, A. C. Jones, M. S. Schwalbach, J. M. Rose, I. Hewson, A. Patel, F. Sun, D. A. Caron and J. A. Fuhrman (2011): "Marine bacterial, archaeal and protistan association networks reveal ecological linkages," ISME J., 5, 1414–1425.
Storey, J. D. (2002): "A direct approach to false discovery rates," J. R. Stat. Soc. B, 64, 479–498.
Storey, J. D., A. J. Bass, A. Dabney and D. Robinson (2015): qvalue: Q-value estimation for false discovery rate control. R package version 2.6.0.
The Human Microbiome Project Consortium. (2012): "Structure, function and diversity of the healthy human microbiome," Nature, 486, 207–214.
Trosvik, P., N. C. Stenseth and K. Rudi (2010): "Convergent temporal dynamics of the human infant gut microbiota," ISME J., 4, 151–158.
Weiss, S., W. V. Treuren, C. Lozupone, K. Faust, J. Friedman, D. Ye, L. C. Xia, Z. Z. Xu, L. Ursell, E. J. Alm, A. Birmingham, J. A. Cram, J. A. Fuhrman, J. Raes, F. Sun, J. Zhou and R. Knight (2016): "Correlation detection strategies in microbial data sets vary widely in sensitivityand precision." ISME J., 10, 1669–1681.
Waterman, M. S. (1995): Introduction to Computational Biology: Maps, Sequences and Genomes, NY, USA: Chapman and Hall/CRC.

Xia, L. C., J. A. Steele, J. A. Cram, Z. G. Cardon, S. L. Simmons, J. J. Vallino, J. A. Fuhrman and F. Sun (2011): "Extended local similarity analysis (elsa) of microbial community and other time series data with replicates," BMC Syst. Biol., 5, S15.

Xia, L. C., D. Ai, J. Cram, J. A. Fuhrman and F. Sun (2013): "Efficient statistical significance approximation for local similarity analysis of high-throughput time series data," Bioinformatics, 29, 230–237.

Xia, L. C., D. Ai, J. A. Cram, X. Liang, J. A. Fuhrman and F. Sun (2015): "Statistical significance approximation in local trend analysis of high-throughput time-series data using the theory of markov chains," BMC Bioinformatics, 16, 301.

Zhou, J., Y. Deng, P. Zhang, K. Xue, Y. Liang, J. D. Van Nostrand, Y. Yang, Z. He, L. Wu, D. A. Stahl, T. C. Hazen, J. M. Tiedje and A. P. Arkin (2014): "Stochasticity, succession, and environmental perturbations in a fluidic ecosystem," Proc. Natl. Acad. Sci. USA, 111, 836–845.

Supplementary Material: The online version of this article offers supplementary material (DOI: https://doi.org/10.1515/sagmb-2018-0019).