

Алгоритм построения сетей взаимодействия видов в сообществах на основе результатов мониторинговых экологических исследований

Букин Б.С.^{1,3}, Михайлов И.С.^{1,3}, Черкашин Е.А.^{2,3}

¹Лимнологический институт сибирского отделения РАН

²Институт динамики систем и теории управления им. В.М. Матросова сибирского отделения РАН, Россия, Иркутск

³Иркутский научный центр сибирского отделения РАН

bukinyura@mail.ru

Внедрение новых методов мониторинга биологического разнообразия экосистем и сообществ, включающие в себя методы метабаркодирования на основе расшифрованных последовательностей ДНК определенных генетических маркеров, значительно расширили потоки информации о спектрах видов и видовом разнообразии. Во многих случаях, особенно при исследовании сообществ микроорганизмов неизвестно, какие виды взаимодействуют друг с другом, и это взаимодействие носит взаимно положительный или отрицательный характер. В работе мы предлагаем алгоритм, который на основе корреляционного анализа, при достаточно больших выборках информации мониторинговых исследований экосистем, позволяет построить сеть взаимодействия между организмами. В сети отражается информация о том, между какими видами в сообществе существует взаимосвязь, и каков характер этой связи, положительный или отрицательный. Алгоритм реализован на языке программирования R.

Ключевые слова: сети взаимодействия, биологические сообщества, экологический мониторинг, биологические системы, корреляционный анализ.

Algorithm for constructing networks of species interaction in communities based on the results of environmental monitoring studies

Bukin Yu.S.^{1,3}, Mikhailov I.S.^{1,3}, Cherkashin E.A.^{2,3}

¹Limnological Institute Siberian Branch of the Russian Academy of Sciences

²Institute for System Dynamics and Control Theory Siberian Branch of the Russian Academy of Sciences, Irkutsk

³Irkutsk Scientific Center Siberian Branch of the Russian Academy of Sciences

The introduction of new methods for monitoring the biological diversity of ecosystems and communities, including metabarcoding methods based on the decoded DNA sequences of certain genetic markers, significantly increased information flows on species spectra. In many cases, especially in the study of microorganism communities, it is not known which species interact with each other and this interaction is mutually positive or negative. In this paper, we propose an algorithm that, based on correlation analysis, for sufficiently large samples of information of monitoring ecosystems studies. Algorithm makes it possible to build a network of interactions between organisms. The network reflects the information about which species from community interact with each other, and what the nature of this relationship is, whether positive or negative. The algorithm is implemented in the programming language R.

Key words: networks of interaction, biological communities, ecological monitoring, biological systems, correlation analysis.

1. Введение

Исследование и характеристика биоразнообразия сообществ в различных экосистемах является одной из основных задач экологии биосистем [1]. Многими коллективами ученых были накоплены обширные массивы информации о численностях и (или) биомассах различных видов организмов в исследуемых

биологических сообществах. Если для макроорганизмов (многоклеточные растения, насекомые, позвоночные) вопрос о механизмах и способах межвидовых взаимодействий решается достаточно просто путем натуральных наблюдений, то для природных сообществ микроорганизмов (одноклеточные эукариоты и прокариоты) ответ на вопрос о том, какие виды взаимодействуют друг с другом и как влияют друг на друга, может быть

достаточно трудным. Исследования в большинстве случаях осложняются тем, большая часть микроорганизмов природных сообществ не культивируются в лабораторных условиях.

С другой стороны, развитие методов метагеномики в некоторой степени позволило решить вопросы установления таксономических составов сообществ микроорганизмов [2]. Применение ДНК баркодирования позволило достаточно точно установить количественные соотношения таксонов в таких сообществах. В настоящее время в базах данных накопились огромные массивы информации расшифрованных последовательностей ДНК, характеризующие сообщества бактерий, архей и микроэукариотов.

ДНК метабаркодирование в некоторых случаях с успехом применяется и для изучения сообществ многоклеточных организмов.

Для того чтобы установить и исследовать механизмы взаимодействия между видами в сообществах можно использовать корреляционный анализ. Как правило, одно и то же сообщество охарактеризовано либо в сезонной динамике, либо исследователями изучалось несколько повторных проб (образцов) из разных географических районов. Таким образом, результат работ представляет собой таблицу, характеризующую представленность определенного спектра видов в разных биологических образцах. Если иметь достаточное количество таких биологических образцов (10 и более), можно рассчитать корреляционную матрицу, отражающую взаимосвязи между видами в изучаемом сообществе. Положительные коэффициенты корреляции будут свидетельствовать о взаимном позитивном влиянии видов друг на друга. Отрицательные значения коэффициентов корреляции будут говорить о том, что в рассматриваемой паре виды конкурируют за какой-либо общий ресурс, подавляя друг друга. Информацию из корреляционной матрицы можно представить в виде сети взаимодействий (ненаправленного графа, отражающего взаимосвязи между организмами). В свою очередь такой граф может служить шаблоном для разработки динамической модели в виде системы дифференциальных уравнений, описывающей и предсказывающей поведение численности (биомассы) видов с течением времени.

В рамках работы была реализована задача расчетов корреляционной матрицы на основе таблицы представленности видов, статистической коррекции значений коэффициентов корреляции и графической визуализации сети взаимодействий между организмами сообщества.

2. Описание алгоритма

В качестве программного средства для реализации алгоритма построения корреляционной сети была выбрана свободно распространяемая статистическая среда программирования R. Среда R

содержит большое количество встроенных функций для статистической обработки данных, кроме того для R доступен большой набор пакетов, реализующих разные статистические методы.

Общая блок-схема алгоритма приведена на рисунке 1.

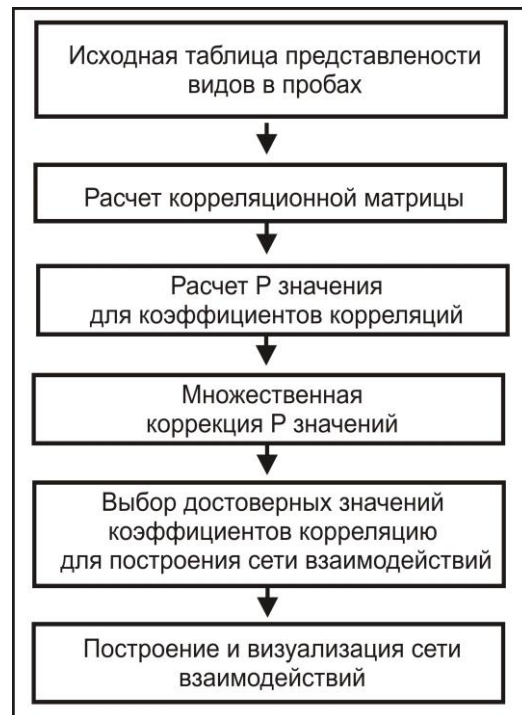


Рис. 1. Блок-схема алгоритма построения корреляционной сети взаимодействия между организмами в сообществе.

Алгоритм включает в себя несколько этапов обработки информации:

1) считывание таблицы, характеризующей представленность таксонов в биологических пробах;

2) тестирование выборок, представленности таксонов в пробах на соответствие закону нормального распределения с помощью теста Шапиро–Уилка;

3) выбор способа расчета коэффициента корреляции:

– коэффициент корреляции Пирсона при нормальных законах распределения таксонов в пробах;

– коэффициент корреляции Спирмена – для ненормальных законов распределения;

4) расчет матрицы значений коэффициентов корреляции;

5) расчет P значений для определения достоверности значений коэффициентов корреляции (с помощью распределения Стьюдента для коэффициентов корреляции Пирсона или с помощью W статистики Спирмена);

6) коррекция P значений на множественное тестирование с помощью метода Бенджамини–Хохберга;

7) замена недостоверных значений ($P > 0.05$) на ноль (недостоверные корреляции);

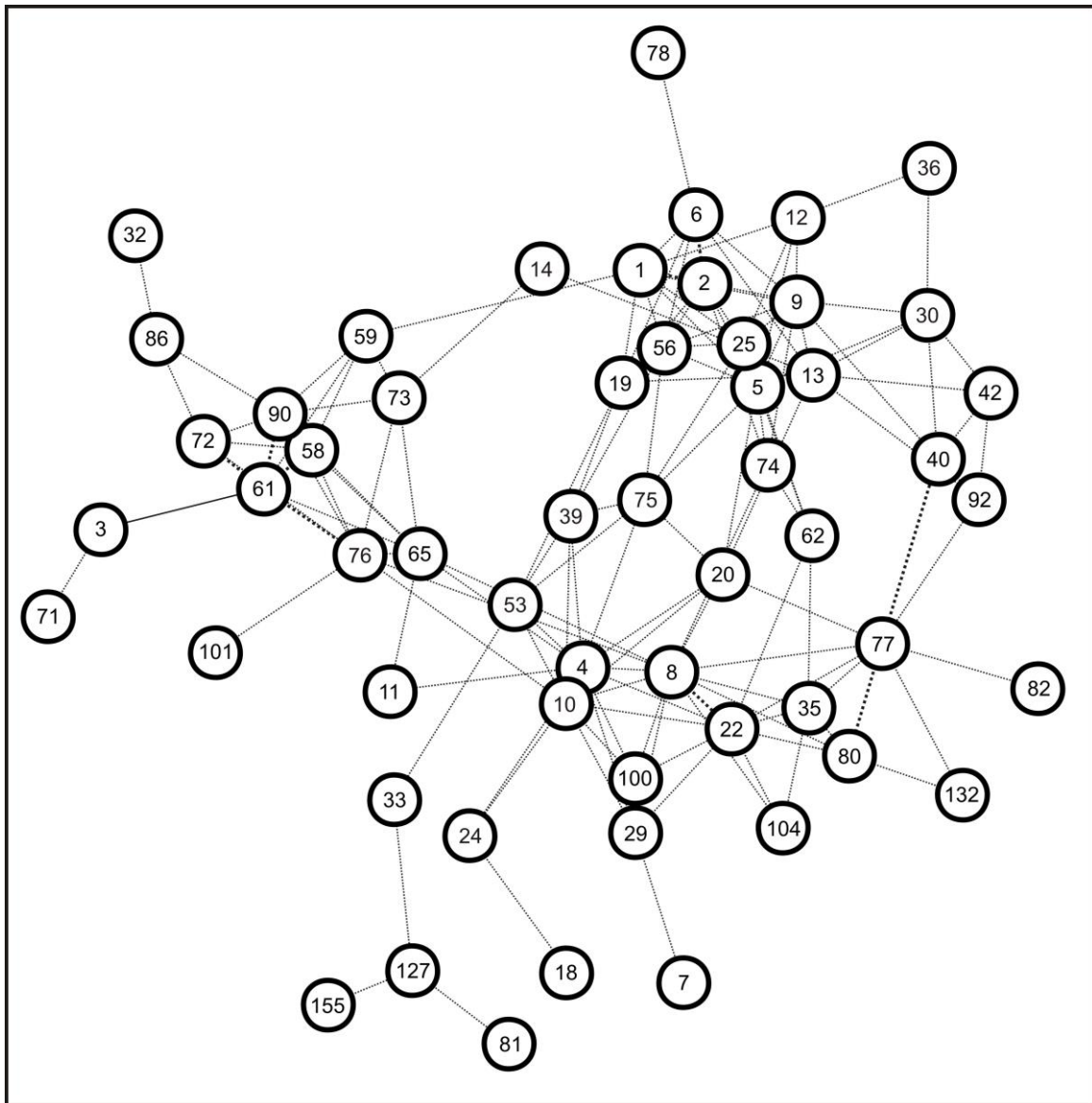


Рис. 2. Пример работы алгоритма построения корреляционной сети взаимосвязей между организмами в сообществе микроорганизмов. Рассматривается бактериальное сообщество экосистемы фотического слоя пелагиали озера Байкал. Цифрами обозначены виды бактерий, пунктирные линии отображают положительные корреляционные связи, сплошные линии – отрицательные корреляционные связи, толщина линий соответствует силе взаимосвязи (чем толще линия, тем больше абсолютное значение коэффициента корреляции).

8) использование матрицы коэффициентов корреляции для построения ненаправленного графа;

9) визуализация ненаправленного графа в виде корреляционной сети взаимодействия организмов.

В качестве средств визуализации сети корреляционных взаимодействий использовались пакеты «igraph» [3] и «qgraph» [4] для языка программирования R.

Дополнительно функции пакета «igraph» в алгоритме позволяет рассчитать различные показатели, помощью которых можно определить важность таксонов в сообществах. К этим показателям относятся: центральность таксона – вида; количество связей, формируемых таксоном с другими видами.

3. Апробация алгоритма

Алгоритм был апробирован на данных, опубликованных в работе [5]. Для анализа была взята таблица, характеризующая представленность различных видов бактерий в 27 образцах воды из фотического слоя пелагиали озера Байкал. Информация о представленности видов в пробах определялась с помощью метабаркодинга на основе региона V3–V4 16S рибосомальной РНК. ПЦР продукты, полученные из смеси ДНК микроорганизмов расшифровывались с помощью секвенатора нового поколения GS 454 FLX (Roche, USA) в Лимнологическом институте СО РАН.

Анализ данных метагеномного секвенирования был выполнен с помощью программы Mothur 1.19.0 [6].

После первичного статистического анализа данных по разработанному алгоритму для расчетов коэффициентов корреляции был выбран ранговый критерий Спирмена. Минимально достоверные значения коэффициентов корреляции по абсолютному значению оказались больше, либо равными 0.53 ($\alpha = 0.05$). Реконструированная сеть межвидовых взаимодействий внутри сообщества бактерий приведена на рисунке 2. Сеть включила в себя 60 узлов, соответствующих определенным видам бактерий и 149 связей между этими узлами. Анализ сети показал, что только одна связь из 149 отображает отрицательное значение коэффициента корреляции, все остальные связи были положительными. Это означает, что в исследованном сообществе озера Бакал большинство видов микроорганизмов находятся в мутуалистических взаимоотношениях. Вероятно, в большинстве случаев продукты жизнедеятельности одного вида микроорганизмов являются субстратом для другого вида микроорганизмов. Конкуренция за общие ресурсы практически отсутствует.

4. Благодарности

Работа выполнена при поддержке Интеграционного проекта 4.2 ИНЦ СО РАН «Применение методов NGS-BD (Next Generation Sequencing – Big Data) для решения вопросов экологии» (постановка задачи, корреляционный анализ), бюджетных тем Лимнологического института СО РАН № 0345-2016-0005 (секвенирование), № 0345-2016-0001 (микроскопия).

5. Список литературы

1. Odum E.P. *Fundamentals of ecology*. WB Saunders company, 1959. 560 p.
2. Petrosino J.F., Highlander S., Luna R.A., Gibbs R.A., Versalovic J. Metagenomic pyrosequencing and microbial identification. *Clinical chemistry*. 2009. V. 55. P. 856–866. doi: [10.1373/clinchem.2008.107565](https://doi.org/10.1373/clinchem.2008.107565).
3. Csardi G., Nepusz T. The igraph software package for complex network research. *Inter Journal, Complex Systems*. 2006. V. 1695. P. 1–9.
4. Epskamp S., Cramer A.O., Waldorp L.J., Schmittmann V.D., Borsboom D. qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*. 2012. V. 48. P. 1–18. doi: [10.18637/jss.v048.i04](https://doi.org/10.18637/jss.v048.i04).
5. Mikhailov I.S., Zakharova Y.R., Bukin Y.S., Galachyants Y.P., Petrova D.P., Sakirko M.V., Likhoshway Y.V. Co-occurrence Networks Among Bacteria and Microbial Eukaryotes of Lake Baikal During a Spring Phytoplankton Bloom. *Microbial ecology*. 2018. P. 1–14. doi: [10.1007/s00248-018-1212-2](https://doi.org/10.1007/s00248-018-1212-2).

6. Schloss P.D., Westcott S.L., Ryabin T., Hall J.R., Hartmann M., Hollister E.B., Sahl J.W. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and environmental microbiology*. 2009. V. 75. P. 7537–7541. doi: [10.1128/AEM.01541-09](https://doi.org/10.1128/AEM.01541-09).