

УДК 574.587+578.087(282.247.431.2)+595.7(476)

Программное обеспечение PAST – как инструмент анализа данных в фаунистических исследованиях

Г.Г. Сушко

В представленной статье продемонстрированы возможности программного обеспечения PAST 4.04 для анализа фаунистических данных. Для демонстрации использованы материалы собственных исследований ассамблей жужелиц (*Coleoptera*, *Carabidae*) в трех типах сосновых лесов в Белорусском Поозерье. Были показаны методики оценки видового богатства и разнообразия с помощью кривых разрежения, эстиматоров видового богатства и индексов разнообразия, а также методики сравнения видового состава с помощью теста ANOSIM, ординации NMDS и теста SIMPER. Целесообразность данного обзора обусловлена тем, что PAST 4.04 обладает возможностью выполнения большинства основных методов анализа данных фаунистических исследований, а также многих необходимых методов статистической обработки и визуализации данных, не требует написания программного кода и находится в открытом доступе.

Ключевые слова: PAST 4.04, видовое богатство, видовое разнообразие, видовой состав, сравнительный анализ.

The capabilities of the PAST 4.04 software for the analysis of faunal data are presented. For the demonstration, materials from our own studies of assemblies of ground beetles (*Coleoptera*, *Carabidae*) in three types of pine forests in the Belarusian Poozerie were used. Methods for assessing species richness and diversity using rarefaction curves, estimates of species richness and diversity indices were shown, as well as methods for comparing species composition using the ANOSIM test, NMDS ordination and SIMPER test. The expediency of this review is due to the fact that PAST 4.04 has the ability to perform most of the basic methods for analyzing data from faunistic research, as well as many necessary methods of statistical processing and data visualization, does not require writing a program code and is in the public domain.

Keywords: PAST 4.04, species richness, species diversity, species composition, comparative analysis.

Фаунистические исследования, как и любые другие, должны базироваться на четко сформированной доказательной основе. Выводы, построенные на сравнении соотношений в процентах, или полученные только с использованием графических методов в наши дни уже не являются достаточно убедительными. Появление современных компьютерных технологий и разнообразие статистических методов анализа данных позволяет улучшить представление результатов фаунистических исследований и правомерность сформулированных выводов. У многих исследователей, особенно молодых (магистрантов, аспирантов), часто возникают проблемы с выбором метода анализа, но чаще всего основной сложностью является отсутствие возможности использовать дорогостоящие лицензионные пакеты анализа данных. Этот пробел могут восполнить общепризнанное в наши дни бесплатное и общедоступное программное обеспечение, такое как R и Python [1]. Но и здесь возникает проблема, так как для работы требуются навыки программирования и нужно писать программный код для соответствующего типа анализа, что зачастую вызывает у биологов сложности. С другой стороны, и лицензионные пакеты анализа данных не обладают всеми необходимыми инструментами оценки результатов фаунистических исследований, такими как построение кривых накопления видов (*species accumulation curves*), расчет индексов биоразнообразия, эстиматоров видового богатства и др. Большинство необходимых типов анализа данных возможно выполнить с использованием приложения PAST (*PA*laeontological *S*Tatistics) [2]. PAST – это бесплатное программное обеспечение для анализа данных палеонтологических исследований. Однако оно получило широкое распространение для статистического анализа и в других областях биологических исследований, в том числе для оценки биоразнообразия, сравнения видового состава и графической визуализации результатов [2]. Приложение не имеет поддержки на русском языке, но это компенсируется тем, что оно не требует установки и запускается даже с флэш-накопителя.

В связи с этим, цель данной работы – продемонстрировать возможности приложения PAST для анализа фаунистических данных.

Материалы и методы. Материалом для демонстрации возможностей анализа фаунистических данных в PAST 4.04 послужили результаты собственных исследований ассамблей жужелиц (Coleoptera, Carabidae) в трех биотопах, таких как сосняк зеленомошный (PP – *Pinetum pleuroziosum*), сосняк черничный (PM – *Pinetum myrtillosum*) и сосняк багульниковый (PL – *Pinetum ledosum*). В каждом биотопе материал собирался с использованием почвенных ловушек в шестикратной повторности (получено по 6 выборочных совокупностей). Всего выявлено 24 вида. В исходной таблице данных виды расположены в строках, столбцы, в которых указано число зарегистрированных особей, соответствуют выборкам из каждого типа биотопа. Соответственно таблица включает 24 строки и 18 столбцов (по 6 – на каждый тип биотопа в соответствии с числом выборок).

Результаты и обсуждение. *Методы оценки видового богатства и разнообразия.* Кроме сравнения числа видов в различных местообитаниях для характеристики видового богатства используются такие методы, как разрежение и расчет потенциально возможного числа видов с помощью эстиматоров [3], [4], [5]. Для построения кривых разрежения одного местообитания по выборкам (samples), которые были здесь получены, в PAST предназначена вкладка **Diversity / Sample rarefaction**. Кривая строится с использованием матрицы данных присутствия-отсутствия (1/0) видов в выборках. Можно не преобразовывать матрицу с числом особей видов, так как любое число в ячейке программа рассматривает как присутствие. Таблица должна включать виды в строках (названия видов, как правило, сокращают до 3 букв названия рода и вида), а выборки в столбцах (рисунок 1).

	Habitat	PM	PM	PM	PM	PM	PM
1	• Pte obl	19	18	19	0	0	0
2	• Car gla	8	8	7	0	0	0
3	• Car cor	0	0	0	3	0	0
4	• Car hor	63	73	52	2	3	4
5	• Car gra	1	0	0	0	0	0
6	• Car arv	7	12	13	25	33	18
7	• Cyc car	9	15	22	0	1	0
8	• Pte lep	0	0	0	0	0	0
9	• Pte ver	0	0	0	0	3	1
10	• Pte obl	12	17	21	0	0	5
11	• Pte nig	59	80	63	3	1	0
12	• Pte mel	2	0	1	0	0	0
13	• Pte bet	6	6	8	0	0	0
14	• Pte ngt	3	0	0	0	0	0
15	• Pte rha	1	0	0	0	0	0
16	• Pte dil	0	1	0	0	0	0
17	• Cal mic	33	27	25	19	15	0
18	• Cal err	0	0	0	0	0	0
19	• Ama bru	1	0	1	0	0	0
20	• Ago eri	0	0	0	0	0	0
21	• Syn viv	0	0	0	0	3	0
22	• Lor pil	0	0	0	0	1	0

Рисунок 1 – Пример ввода данных для построения кривой разрежения для одного местообитания (*Pinetum myrtillosum* – PM) по нескольким выборкам (скриншот программы PAST)

Разрежение на основе выборок (также известное как кривая накопления видов) применимо только, когда есть несколько выборок (не менее 5), по которым видовое богатство оценивается как функция числа выборок. PAST использует методику анализа «Мао's tau» и рассчитывает стандартное отклонение. На графике стандартные ошибки преобразованы в 95-процентные доверительные интервалы [3]. Используя результаты наших исследований, построим кривые для ассамблей жужелиц трех биотопов. Для сосняка зеленомошного кривая достаточно плавная, и она приближается к асимтоте (рисунок 2а). Это говорит о том, что видовое богатство может быть выше, чем выявленное нами. Кривая для сосняка черничного имеет вид асимтоты, и она выходит на плато, что указывает на то, что выявленное число видов практически соответствует потенциально возможному в данном местообитании (рисунок 2б). В сосняке багульниковом кривая наиболее сглажена, что свидетельствует о наличии большего числа видов, чем здесь обнаружено (рисунок 2в).

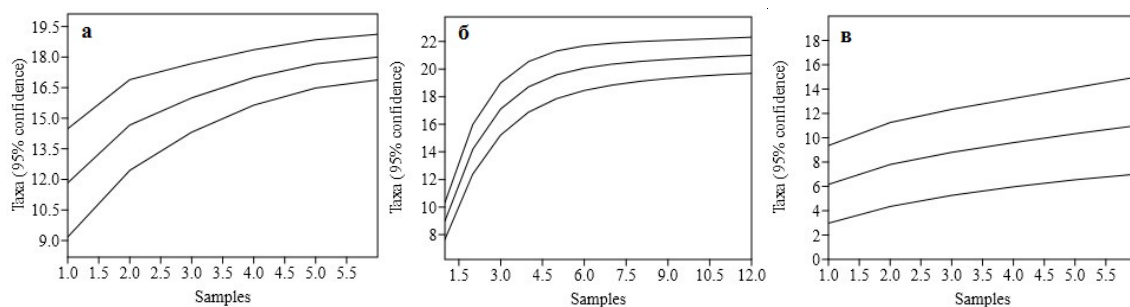


Рисунок 2 – Кривые разрежения для ассамблей жуужелиц сосняка зеленомошного (*Pinetum pleuroziosum*) (а), сосняка черничного (*Pinetum myrtillosum*) (б), сосняка багульникового (*Pinetum ledosum*) (в); Таха – число видов, Samples – число выборок

В PAST предусмотрена и другая методика оценки видового богатства – сравнение с помощью кривых разрежения ассамблей видов нескольких местообитаний по числу особей (или других количественных данных). Для этого предназначен модуль **Individual rarefaction** во вкладке **Diversity**. Этот модуль предполагает, как и в предыдущем примере, использование таблицы данных о численности, но уже в нескольких местообитаниях. Методика позволяет оценить, сколько видов вы ожидаете обнаружить в выборках даже с не высоким общим количеством особей, сравнить количество видов в выборках разного размера (как правило, мы имеем дело с разным числом видов в выборках из одного местообитания). Используя анализ разрежения для большой выборки, вы можете предположить количество ожидаемых видов для любой меньшей выборки.

Полученный нами график (рисунок 3) содержит информацию обо всех выборках и позволяет сравнивать их размеры, а также видовое богатство.

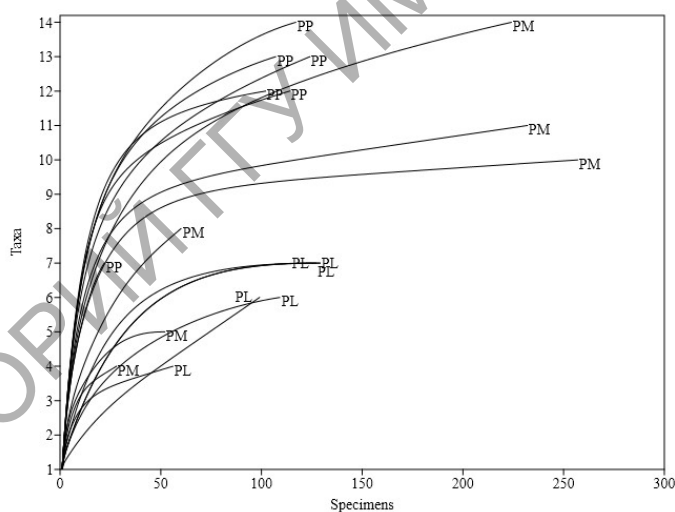


Рисунок 3 – Кривые разрежения для ассамблей жуужелиц нескольких местообитаний (PP – *Pinetum pleuroziosum*, PM – *Pinetum myrtillosum* и PL – *Pinetum ledosum*); Таха – число видов, Specimen – число особей

На рисунке 3 можно увидеть, что в целом число видов в отдельных выборках выше в сосняке зеленомошном, а наименьшее – в сосняке багульниковом. Как и в предыдущем примере, наиболее близкое видовое богатство к потенциально возможному наблюдается в сосняке черничном. Полученные кривые обладают высокой информативностью и в зависимости от целей исследования дают возможность для всестороннего сравнения выборок. Кроме того, модуль позволяет продемонстрировать каждую выборку по отдельности, в том числе и с указанием доверительного интервала (рисунок 4).

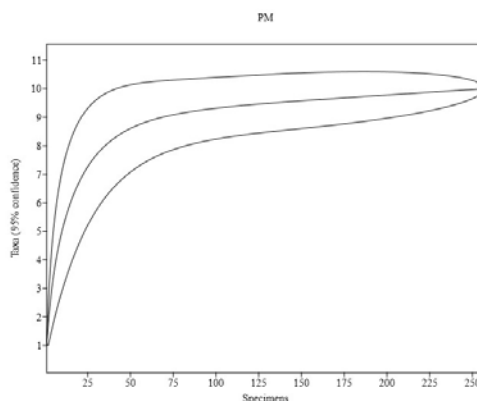


Рисунок 4 – Кривая разрежения для ассамблеи жужелиц сосняка черничного (PM – *Pinetum myrtillosum*) для одной из выборок; Taxa – число видов, Specimen – число особей

Другой способ оценки видового богатства – сравнение полученного в результате исследований числа видов с потенциально возможным видовым богатством установленным с помощью так называемых эстиматоров. В PAST заложены алгоритмы вычисления 4 эстиматоров: Chao 2, jackknife 1, jackknife 2 и bootstrap. Для их вычисления предназначен модуль **Quadrat richness** во вкладке **Diversity**. Анализ выполняется с применением таблицы данных присутствия-отсутствия (1/0) видов в выборках [6]. Можно не преобразовывать матрицу с числом особей видов, так как любое число в ячейке программа рассматривает как присутствие. Таблица должна включать виды в строках, а выборки в столбцах (рисунок 1). Продемонстрируем расчеты на примере выборок полученных в сосняке черничном (рисунок 5).

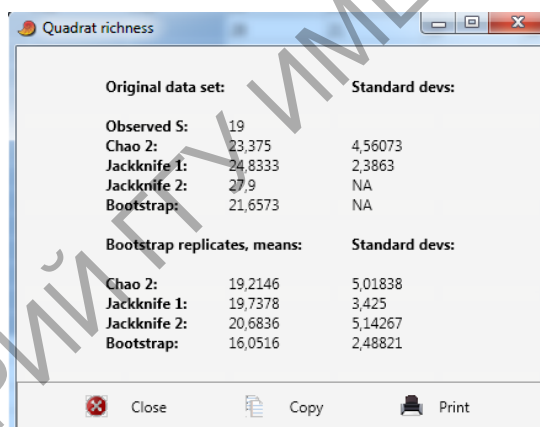


Рисунок 5 – Панель результатов расчетов эстиматоров видового богатства для одного местообитания (PM – *Pinetum myrtillosum*) по нескольким выборкам (скриншот программы PAST)

Результаты расчетов эстиматоров видового богатства разделены на две панели. Сначала по заданному набору выборок вычисляются значения эстиматоров и их стандартные отклонения (только для Chao2 и Jackknife1). Затем эстиматоры вычисляются повторно с помощью 1000 созданных путем случайных перестановок выборок (методика bootstrapping). На второй панели показаны их средние значения и стандартные отклонения (рисунок 5).

В нашем примере общее число видов (Sobs) по всем 6 выборкам составило 19. Тогда как результаты на второй панели показывают, что предполагаемое число видов по значениям разных эстиматоров может быть от 16 до 20 видов в данном местообитании с относительно невысокими стандартными отклонениями. Непараметрические эстиматоры демонстрируют, как правило, завышенные результаты. Считается, что Chao 2 и Jackknife 2 дают оценки наиболее приближенные к реальности [6], [7]. Следовательно, исходя из значений Chao 2 и Jackknife 2, мы приложили достаточные выборочные усилия при сборе материала, и выявленное нами видовое богатство жужелиц отражает потенциально возможное в данном биотопе.

Для оценки видового разнообразия можно использовать ряд различных индексов с помощью модуля **Diversity indices** во вкладке **Diversity**.

Методы оценки видового состава. Следующим этапом анализа фаунистических данных, как правило, является сравнение видового состава. Различия видового состава ассамблей жужелиц 3 местообитаний проанализируем с использованием теста ANOSIM (Analysis Of Similarities). ANOSIM проверяет значимые различия между двумя или более группами на основе определенной меры расстояния. В экологических исследованиях в качестве такой меры используют расстояние Брея-Кертиса. В ходе анализа формируется итоговая статистическая мера R-тест. Если $R = 1$, сообщества очень сильно различаются, если $R = 0$ – полностью сходны. Значения R больше 0,75 принято интерпретировать как хороший уровень различий, R больше 0,5 – как наличие различий, R меньше 0,25 – как едва различимые различия [8]. Также в PAST для сравнения видового состава можно применить и PERMANOVA (Permutational multivariate analysis of variance). Анализ PERMANOVA можно выполнить с использованием вкладки **Multivariate / Tests / one-way PERMANOVA**.

Для выполнения теста ANOSIM использована вкладка **Multivariate / Tests / one-way ANOSIM**. В таблице данных для анализа в столбцах указано число особей разных видов, тогда как в строках – названия выборок из соответствующих биотопов (рисунок 6). Нужно обязательно обозначить тип биотопа как группирующую переменную в атрибутах столбцов в окне ввода данных программы. Перед анализом нужно обратить внимание на выбор меры расстояния или индекса сходства (Similarity index). Изначально программа предлагает меру расстояния Эвклида, которая не всегда подходит для экологических данных. Поэтому нужно выбрать меру расстояния Брея-Кертиса.

Site ID	Habitat	Car vio	Car gla	Car cor	Car hor	Car gra	Car arv	Cys car
1	PM	19	8	0	63	1	7	9
2	PM	18	8	0	73	0	12	15
3	PM	19	7	0	52	0	13	22
4	PM	0	0	3	2	0	25	0
5	PM	0	0	0	3	0	33	1
6	PM	0	0	0	4	0	18	0
7	PP	13	0	0	6	0	17	2
8	PP	11	0	0	13	0	15	1
9	PP	18	0	0	6	0	11	2
10	PP	18	1	0	7	0	18	2
11	PP	18	0	0	6	0	8	1
12	PP	0	1	2	8	0	2	0
13	PL	0	2	0	0	3	0	0
14	PL	0	0	0	0	4	0	0
15	PL	0	0	0	1	0	0	1
16	PL	0	0	0	0	0	0	0
17	PL	0	0	0	0	2	0	0
18	PL	0	2	0	0	3	0	0
19								
20								
21								
22								

Рисунок 6 – Пример ввода данных для выполнения сравнительного анализа видового состава жужелиц нескольких местообитаний по нескольким выборкам в каждом (скриншот программы PAST)

В нашем примере $R = 0,818$, $p = 0,0001$, что свидетельствует о высоких достоверных различиях видовой состав жужелиц трех биотопов. Для повышения уровня достоверности результатов стандартный p -уровень, который приводится по умолчанию, можно корректировать с помощью поправки Бонферрони, что сделано в рассматриваемом примере. Среди предложенных программой поправок можно также использовать еще одну – sequential Bonferroni, которая является наиболее мощной, но может давать заниженные результаты [4].

Для визуализации различий выборок PAST содержит широкий набор инструментов от кластерного анализа до различных методов ординации (PCA, PCoA, CA, DCA, NMDS и др.), которые можно реализовать с помощью вкладки вкладка **Multivariate**. Однако многие из этих методов могут иметь ограничения для данных фаунистических исследований [9]. Наиболее часто используемым методом является NMDS (Non-metric multidimensional scaling), так как не требует нормального распределения данных и позволяет выбрать из множества различных мер расстояния подходящее.

Для демонстрации данного анализа используем таблицу данных ту же, что и для предыдущего примера, но с некоторыми изменениями. Виды, представленные менее, чем 3 особями в выборках, должны быть исключены из анализа для снижения так называемого «шума». Поскольку многие виды представлены в выборках 1 экземпляром, перед анализом нужно выполнить преобразование данных (логарифмирование, $\log_{10}(x + 1)$) [7], [9]. Для лучшей визуализации можно обозначить выборки одного биотопа определенным символом используя **Row attributes / Symbol**. В нашем примере РМ обозначен точкой, РР – плюсом, РЛ – квадратом.

Для выполнения неметрического многомерного шкалирования использована вкладка **Multivariate / Ordination / Non-metric MDS**. На полученной ординационной диаграмме видна отчетливая дифференциация биотопов по обилию коллектированных видов. Сходством характеризуются биотопы РМ и РР. Биотоп РЛ обладает наибольшими отличиями от остальных (рисунок 7). Это демонстрирует степень различия (градиент) экологических условий анализируемых местообитаний. Концентрация символов, соответствующих выборкам, вокруг определенных биотопов указывает на то, что данные выборки получены из различных генеральных совокупностей.

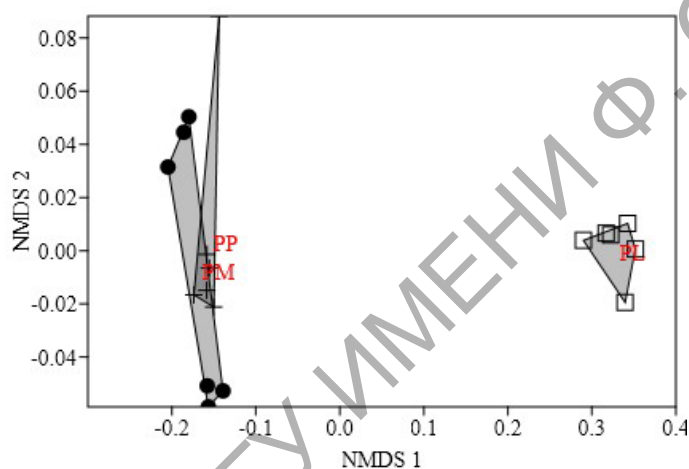


Рисунок 7 – Диаграмма NMDS ординации для ассамблей жужилиц трех местообитаний (PP – *Pinetum pleuroziosum*, РМ – *Pinetum myrtillosum* и РЛ – *Pinetum ledosum*)

Для проверки качества результатов анализа можно построить диаграмму Шепарда с помощью вкладки **Shepard plot** [10].

Завершающим этапом анализа сходства или различия видового состава нескольких биотопов может быть выявление индикаторных видов или видов, которые способствуют наибольшему различию. В PAST для данных целей предусмотрен тест SIMPER (Similarity of Percentage). SIMPER тест, основанный на сходстве в процентах, является простым методом оценки вклада определенных видов в наблюдаемые различия между группами выборок. Как правило, различия между выборками в целом предварительно оцениваются с помощью ANOSIM. Для SIMPER чаще всего используется мера сходства Брея-Кертиса, но также можно выбирать и другие меры при необходимости. Если анализируется более двух групп, вы можете сравнить любые две группы (парно), или объединить все выборки для выполнения общего анализа (overall multi-group SIMPER). Общее среднее различие (overall average dissimilarity) рассчитывается с использованием всех видов, в то время как различие, характерное для конкретного вида, вычисляется по отдельности [8].

Таблица данных для теста SIMPER сформирована по такому же принципу, как и в предыдущем примере. Для выполнения анализа используем вкладку **Multivariate / Tests / SIMPER**. В таблице результатов SIMPER теста, по данным первых трех столбцов (Average dissimilarity – среднее различие, Contribution % – вклад в различия в %, Cumulative % – накопленные различия в %), можно выявить виды, которые вносят наибольший вклад в гетерогенность видового состава анализируемых биотопов (Average dissimilarity).

Заключение. Таким образом, программное обеспечение PAST 4.04 обладает необходимыми возможностями для анализа фаунистических данных. В их числе методики оценки видового богатства и разнообразия с помощью различных типов кривых разрежения, эстиматоров видового богатства и индексов разнообразия. Кроме того, PAST 4.04 содержит методы многомерного анализа данных, такие как различные типы ординации и непараметрических тестов, которые позволяют выявить достоверные различия видового состава и визуализировать их.

Литература

1. Кабаков, Р. И. R в действии. Анализ и визуализация данных в программе R. / Р. И. Кабаков – М. : ДМК Пресс, 2014. – 588 с.
2. Hammer, Ø. PAST: Paleontological Statistics Software Package for Education and Data Analysis / Ø. Hammer, D. A. T. Harper, P. D. Ryan // *Palaeontologia Electronica*. – 2001. – Vol. 4, № 1. – P. 1–9.
3. Colwell, R. K. Interpolating, extrapolating, and comparing incidence-based species accumulation curves / R. K. Colwell, C. X. Mao, J. Chang // *Ecology*. – 2004. – № 85. – P. 2717–2727.
4. Legendre, P. Numerical Ecology / P. Legendre, L. Legendre. – 2nd ed. – Amsterdam : Elsevier, 1998. – 854 p.
5. Colwell, R. K. Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages / R. K. Colwell, A. Chao, N. J. Gotelli, S.-L. Lin, C. X. Mao, R. L. Chazdon, J. T. Longino // *Journal of Plant Ecology*. – 2012. – № 5. – P. 3–21.
6. Colwell, R. K. Estimating terrestrial biodiversity through extrapolation / R. K. Colwell, J. A. Coddington // *Philosophical Transactions of the Royal Society. Series B*. – 1994. – № 345. – P. 101–118.
7. McCune, B. Analysis of ecological communities / B. McCune, J. B. Grace. – Gleneden Beach : MjMSoftware Design, 2002. – 300 p.
8. Anderson, M. J. PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: what null hypothesis are you testing? / M. J. Anderson, D. C. Walsh // *Ecological monographs*. – 2013. – Vol. 83, № 4. – P. 557–574.
9. Сушко, Г. Г. Методы многомерного анализа данных в синэкологии насекомых / Г. Г. Сушко // *Журнал Белорусского государственного университета. Экология*. – 2020. – № 1. – С. 38–45.
10. Borcard, D. Numerical Ecology with R. / D. Borcard, F. Gillet, P. Legendre. – Wien : Springer Nature, 2018. – 306 p.

Витебский государственный
университет им. П.М. Машерова

Поступила в редакцию 01.02.2021