



УДК 599.363

ОПЫТ ИСПОЛЬЗОВАНИЯ СТАТИСТИЧЕСКИХ ТЕСТОВ ДЛЯ КОРРЕКТИРОВКИ МЕТРИЧЕСКИХ ДАННЫХ ПО ИСКОПАЕМЫМ ЗЕМЛЕРОЙКАМ CROCIDOSORICINAE (LIPOTYPHILA: SORICIDAE)

Л.Л. Войта^{1*} и В.С. Зажигин²

¹Зоологический институт Российской академии наук, Университетская наб. 1, 199034 Санкт-Петербург, Россия;
e-mail: leonid.voyta@zin.ru

²Геологический институт Российской академии наук, Пыжевский пер. 7, 109017 Москва, Россия;
e-mail: zazhvol@gmail.com

РЕЗЮМЕ

В статье рассматривается пример использования параметрических (Шапиро-Уилка, Харке-Бера) и непараметрического (Андерсона-Дарлинга) критериев для разведочного анализа на соответствие гипотезе об однородности выборки, и о пригодности признаков для межвидовых сравнений. Тесты Граббса и Туки использовались для определения отклоняющихся значений метрических признаков. Исследовался ископаемый материал, представленный фрагментами костей и изолированными зубами ($n = 218$) неизвестной науке кроцидосорициновой землеройки †Crocidosoricinae indet. (LipotypHila: Soricidae) из среднемиоценового местонахождения Шарга 2 (Монголия). По результатам исследования было определено три промера (MRWc, HC и W(m3)), которые не рекомендуется к использованию при межвидовых сравнениях в процессе работы с Crocidosoricinae indet. Кроме того, исследование характера распределения 25 метрических признаков Crocidosoricinae indet. из местонахождения Шарга 2 показало, что с большой долей вероятности весь материал принадлежит единой генеральной совокупности (локальной популяции), остатки которой накапливались относительно короткое время.

Ключевые слова: дискриминантный анализ, тест Андерсона-Дарлинга, тест Граббса, тест Туки, тест Шапиро-Уилка, тест Харке-Бера, Crocidosoricinae, Soricidae

A PRIMER OF CORRECTION OF QUANTITATIVE DATA OF EXTINCT SHREWS CROCIDOSORICINAE (LIPOTYPHILA: SORICIDAE) USING STATISTIC TESTS

L.L. Voyta^{1*} and V.S. Zazhigin²

¹Zoological Institute, Russian Academy of Sciences, Universitetskaya nab. 1, Saint Petersburg, 199034, Russia;
e-mail: leonid.voyta@zin.ru

²Geological Institute, Russian Academy of Sciences, Pyzhevskii per. 7, Moscow, 109017, Russia; e-mail: zazhvol@gmail.com

ABSTRACT

This article presents an example of using the parametrics (Shapiro-Wilk and Jarque-Bera) and non-parametrics (Anderson-Darling) criteria for exploratory analysis on a sample homogeneity hypothesis, and estimate the suitability of metric features for interspecific comparisons. Grubbs and Tukey tests were used for detecting of the outliers of metric features. The analyzed fossil material is represented by bone fragments and isolated teeth ($n = 218$)

* Автор-корреспондент/Corresponding author

of an unknown crocidosoricine shrew †Crocidosoricinae indet. (Lipotyphla: Soricidae) from the middle Miocene Sharga 2 locality (Mongolia), and. Based on the results of the study, three measurements (MRWc, HC and W(m3)) are not recommended for interspecies comparisons while working with Crocidosoricinae indet. In addition, a homogeneity analysis of 25 metric characters of Crocidosoricinae indet. from Sharga 2 locality showed that all remains are likely to belong to a single general sample (local population), remains of which accumulated during relatively short period.

Key words: discriminant analysis, Anderson-Darling test, Grubbs test, Tukey test, Shapiro-Wilk test, Jarque-Bera test, Crocidosoricinae, Soricidae

ВВЕДЕНИЕ

Одной из задач морфометрического анализа, применяемого в зоологических исследованиях, является определение однородности распределения значений метрических параметров, что должно свидетельствовать о принадлежности измеренных особей одной совокупности (выборке). Статистически однородность может быть оценена с использованием гипотезы о соответствии реального распределения наблюдаемых значений выбранной модели распределения. В случае совпадения модели и реального распределения мы говорим об однородности выборки; в случае не совпадения, следует либо пересмотреть модель, либо признать неоднородность выборки. В свою очередь, неоднородность в узком контексте определяется, как один из возможных вариантов – отклоняющимися значениями, соответствующими, например, экземплярам из другой совокупности (например, особи другого вида), или ошибкам, допущенным при измерении.

При анализе однородности кранио- или одонтометрического материала мелких млекопитающих мы имеем дело с рядом выборок, каждая из которых отдельно тестируется по одному или нескольким признакам. Таким образом, предположительно, мы работаем с одномерным непрерывным распределением, которое, чаще всего, описывается простой моделью нормального распределения, хотя это не единственно возможный вариант и существуют другие – модель логнормального распределения, гамма-распределение и пр. (подробнее см. Пузаченко [Puzachenko] 2004). Логарифмирование данных позволяет уменьшить возможные эффекты асимметрии распределения, и, таким образом, приблизить распределение наблюдаемых значений к нормальному (Пузаченко [Puzachenko] 2004).

Если принять модель нормального распределения за основу, то тестирование гипотезы «анализируемая выборка происходит из генеральной совокупности, имеющей нормальное распределение» может осуществляться с использованием ряда формальных критериев, например, Шапиро-Уилка (Shapiro-Wilk test), Харке-Бера (Jarque-Bera test), Андерсона-Дарлинга (Anderson-Darling test). Если получаемая при помощи теста вероятность ошибки (p-level) оказывается меньше принятого уровня значимости (например, p-level < 0.05), результаты теста считаются «неудовлетворительными» и нулевая гипотеза отклоняется.

Параметрический тест Шапиро-Уилка считается одним из наиболее распространенных критериев проверки корреляции между исходными данными и соответствующим предположением об их нормальном распределении (Shapiro and Wilk 1965; Öztuna et al. 2006; Ghasemi and Zahedials 2012). Параметрический тест Харке-Бера основан на тестировании соответствия асимметрии (skewness) и эксцесса (kurtosis) исходных данных и модели нормального распределения (Gujareti and Porter 2010: 131). Непараметрический тест Андерсона-Дарлинга использует преобразование данных (нормализация среднего и стандартного отклонения), и позволяет тестировать простые гипотезы соответствие наблюдаемого распределения не конкретной модели, а какой-либо из известных простых моделей (нормального, логнормального, гамма-распределения) (Anderson and Darling 1952; Кобзарь [Kobzar'] 2006; Лемешко и Лемешко [Lemeshko and Lemeshko] 2009). Использование в практической работе трех перечисленных тестов позволяет: (а) оценить однородность выборки по соответствию исходных данных модели нормального распределения (критерии Шапиро-Уилка и Харке-Бера), или другим про-

стым моделям (критерий Андерсона-Дарлинга); и согласно этому (б) выявить совокупности, где присутствуют отклоняющиеся значения.

Отклоняющиеся значения метрических параметров, если они не связаны с неверным «видовым» определением, довольно часто, являются следствием ошибок, допущенных при измерении, либо при вводе данных в электронные таблицы. Поиск таких «ошибок» является еще одной задачей предварительной стадии морфометрического анализа, и может выполняться с использованием ряда специальных тестов, используемых для обнаружения выбросов при одномерном анализе: параметрический критерий Граббса (Grubbs test) обнаруживает отклонение значений от модели нормального распределения, и основан на различии среднего значения выборки и наиболее экстремального значения с учетом среднего квадратического отклонения (Stefansky 1972). Ограничением применения является то, что тест показывает корректные результаты при симметричном (или близком к нему) распределении наблюдаемых данных. Непараметрический тест Туки (Tukey test) также используется для выявления экстремальных значений, но с применением межквартильного диапазона (IQR, interquartile range), или стандартизованного диапазона значений (Tukey 1949). Это позволяет применять его в случаях, когда тип простого непрерывного распределения отличается от нормального.

В настоящем исследовании использование вышеперечисленных подходов к оценке однородности выборки определялось тем, что в нашем распоряжении оказался фрагментированный кранио-и одонтометрический материал вымершей землеройки *Crocidosoricinae* indet. (Lipotyphla: Soricidae) из среднемиоценового местонахождения Шарга 2 (подробнее см. ниже). Имея дело с рецентным материалом, где каждый экземпляр представлен одновременно черепом и нижними челюстями, для оценки однородности (проверки гипотезы соответствия распределений) достаточно использовать какой-либо один метрический параметр (например, наибольшую длину черепа). Ископаемый же материал фрагментирован и каждая группа костных остатков или изолированный зуб может принадлежать экземпляру из другой выборки. В статье мы предлагаем следующий алгоритм определения однородности ископаемой выборки: (а) протестировать соответствие

наблюдаемого и теоретического распределений и оценить пригодность набора метрических признаков для оценки однородности, используя локальные выборки современных видов *Crocidura shantungensis* Miller, 1901 и *Sorex minutus* Linnaeus, 1766; в соответствие с этим (б) тестировать пригодные метрические параметры ископаемой землеройки *Crocidosoricinae* indet. с целью выявления отклоняющихся значений и, таким образом, сформировать суждение об однородности ископаемого материала, т.е. о принадлежности всех исследованных остатков одной выборке, или наличия в ней других форм.

МАТЕРИАЛ И МЕТОДЫ

Изученный материал был представлен 218 костными остатками (черепа и нижней челюсти) и изолированными зубами землеройки из подсемейства *Crocidosoricinae* (Lipotyphla: Soricidae). Работа по определению вида еще не закончена, результаты ее будут опубликованы позднее; здесь для обозначения ископаемого таксона мы используем открытую номенклатуру – *Crocidosoricinae* indet. Материал из среднемиоценовых отложений местонахождения Шарга 2 (пустыня Шаргинская Гоби, Юго-Западная Монголия). Ископаемые остатки были собраны В.С. Зажиным летом 1978 г., и В.С. Зажиным и Е.В. Девяткиным летом 1979 г. Информация об этом местонахождении приводится в публикациях Зажигина и Лопатина (Zazhigin and Lopatin 2000; Lopatin and Zazhigin 2003).

Для сравнения был использован материал по современным видам землероек: манчжурская белозубка, *Crocidura shantungensis* Miller, 1901 (n = 30), и малая бурозубка, *Sorex minutus* Linnaeus, 1766 (n = 30) (см. Приложение 1). Выбор этих видов определялся сходством их размеров с оценочным размером ископаемой землеройки.

Все фрагменты черепа и нижних челюстей, а также зубы фотографировались цифровой камерой (Canon EOS 60D) совмещенной с бинокляром (LOMO Micromed MSP-1) через оптический адаптер (AOT-1C). Морфологические структуры на изображении измерялись с использованием экранного дигитайзера tpsDig ver. 2.01 (Rohlf and Slice 1990). Схема промеров, с некоторыми дополнениями, взята по Reumer (1984): COR – высота венечного отростка нижней челюсти;

НС – высота сочленовного отростка; L – длина зуба (для P4, M3, m3); LML – длина нижнего ряда моляров; LLF – длина нижней фасетки сочленовного отростка; LUF – длина верхней фасетки сочленовного отростка; MBH – высота зубной кости (на уровне протоконоида m2); MRH – высота восходящей ветви нижней челюсти; MRWc – ширина восходящей ветви (измеряется между нижнечелюстным отверстием и концом верхним сочленовного отростка); PMt – положение подбородочного отверстия относительно края основания резца i1; Wios – ширина латеральной стенки подглазничного канала; BL – буккальная длина верхних моляриформных зубов (P4–M2); TAL – длина талонида (для m1–m2); TAW – ширина талонида (ibid.); TRL – длина тригонида (ibid.); TRW – ширина тригонида (ibid.). Выбор анализируемых признаков определялся наличием их у ископаемого таксона. Перед анализом все промеры логарифмировались.

Тестирование выборок на однородность и дискриминантный анализ проводились с использованием статистического пакета PAST ver. 3.15 (Hammer et al. 2001). Результаты тестов Шапиро-Уилка (W), Харке-Бера (JB) и Андерсона-Дарлинга (A) одновременно рассчитываются во вкладке «Normality tests».

Выявление экстремальных значений тех метрических параметров, которые не прошли тестирование однородности, проводилось, вначале визуально на графиках, далее с использованием пакета Statistica 10 (<http://statistica.io>), где применялись тесты Граббса (G) и Туки (q). Для G использовались следующие установки: тестирование симметричного распределения (Two Sided – поиск выбросов со значениями больше и меньше выборочной средней величины); критическое значение теста при уровне значимости $\alpha = 0.05$ (что при $n = 6–30$ соответствует критическим значениям теста 1.944–3.103, и определяет примерно 5% выбросов, или, соответствует 95% доверительному интервалу); количество итераций – 100. Для q использовались установки: тестирование симметричного распределения (Two Sided); критическое значение теста (IQR, interquartile range) – 1 (что соответствует 2σ , и определяет примерно 5% выбросов); количество итераций – 50.

Сокращения: GIN – Геологический институт Российской академии наук; ZISP – Зоологический институт Российской академии наук;

i1 – нижний резец; M/m(1–3) – верхние/нижние коренные зубы (первый–третий); P4 – четвертый верхний премоляр; PZC – персональный каталог ископаемого материала В.С. Зажигина (Personal Zazhigin's Catalog).

РЕЗУЛЬТАТЫ

Разведочный анализ данных на однородность

***Crocidura shantungensis*.** Результаты тестирования 25 промеров этого вида на соответствие гипотезе об однородности выборки представлены в Табл. 1. Применение критериев выявило семь параметров с неудовлетворительными результатами тестов ($p < 0.05$): критерий Шапиро-Уилка определил несоответствие для MBH ($W = 0.905$), TAW(m1) ($W = 0.913$), L(m2) ($W = 0.921$), TRW(m2) ($W = 0.923$), BL(M1) ($W = 0.921$); критерий Андерсона-Дарлинга определил четыре параметра: MBH ($A = 0.893$), TRW(m1) ($A = 0.777$), TRW(m2) ($A = 0.762$), TAW(m2) ($A = 0.795$); критерий Харке-Бера определил три параметра: MBH ($JB = 5.980$), L(m2) ($JB = 3.845$), BL(M1) ($JB = 6.155$). Визуальная оценка распределения значений на графиках для каждого из семи выявленных признаков показала: значительно отклоняющиеся значения MBH (Рис. 1А); отклоняющиеся значения TRW(m1) (Рис. 1В). Формальная оценка отклоняющихся значений с помощью критерия Граббса подтвердила отклонение значений только по промеру TAW(m1), а критерий Туки – по шести промерам: MBH, TRW(m1), TAW(m1), L(m2), TRW(m2), BL(M1) (Табл. 2).

***Sorex minutus*.** Результаты тестирования этого вида представлены в Табл. 1. Применение критериев выявило шесть несоответствий параметров наблюдаемого распределения теоретическому. Каждый из этих параметров (COR, MBH, MRWc, LLF, L(m3) и BL(P4)) был также проанализирован как визуально, так и формально с помощью тестов Граббса и Туки. Первый критерий подтвердил отклонение по промерам COR и L(m3) (Табл. 3); критерий Туки выявил 21 отклоняющийся экземпляр в разной комбинации для шести промеров (Табл. 3).

Crocidosoricinae indet. Для ископаемой выборки всего было выявлено пять признаков: MRWc, НС, LUF, L(m2) и W(m3) (Табл. 1). Визуальная оценка показала хорошо различимые экс-

Таблица 1. Значения тестов Шапиро-Уилка (W), Андерсона-Дарлинга (A) и Харке-Бера (JB) для 25 метрических признаков *Crocidura shantungensis*, *Sorex minutus* и землеройки †*Crocidosoricinae* indet.

Table 1. Shapiro-Wilk test (W), Anderson-Darling test (A), and Jarque-Bera test (JB) values for 25 metric features of *Crocidura shantungensis*, *Sorex minutus*, and a shrew †*Crocidosoricinae* indet.

Measurements	<i>C. shantungensis</i>			<i>S. minutus</i>			<i>Crocidosoricinae</i> indet.		
	W/n	A/n	JB/n	W/n	A/n	JB/n	W/n	A/n	JB/n
COR	0.964/30	0.323/30	1.236/30	0.751*/29	1.897*/29	130.3*/29	0.9419/19	0.430/19	1.506/19
				0.969	0.245	1.125			
MBH	0.905*/30	0.893*/30	5.980*/30	0.912*/30	0.979*/30	2.810*/30	0.9562/18	0.409/18	0.461/18
	0.927*	0.679	3.400	0.915*	0.994*	2.607			
MRH	0.970/30	0.293/30	1.327/30	0.937/29	0.698/29	1.762/29	0.9436/19	0.428/19	1.369/19
MRWc	0.962/30	0.387/30	0.152/30	0.817*/29	1.495*/29	62.220*/29	0.849*/15	1.108*/15	4.608/15
				0.967	0.238	0.542	0.797*	0.924*	12.63*
Pmt	0.976/30	0.258/30	0.775/30	0.944/30	0.516/30	1.526/30	0.927/7	0.276/7	0.567/7
HC	0.976/30	0.332/30	0.583/30	0.961/30	0.474/30	0.363/30	0.914*/27	0.858*/27	7.108*/27
							0.914*/27	0.858*/27	7.108*/27
LLF	0.943/30	0.547/30	3.563/30	0.942/30	0.760*/30	1.318/30	0.964/30	0.356/30	0.817/30
					0.668				
LUF	0.959/30	0.423/30	0.892/30	0.976/30	0.310/30	1.151/30	0.888*/27	0.804*/27	19.460*/27
							0.959	0.472	0.063
LML	0.943/30	0.544/30	0.950/30	0.963/30	0.342/30	1.162/30	0.983/16	0.131/16	0.533/16
L(m1)	0.964/30	0.268/30	1.971/30	0.940/30	0.533/30	0.892/30	0.977/59	0.293/59	3.034/59
TRL(m1)	0.9708/30	0.334/30	0.888/30	0.970/30	0.248/30	0.666/30	0.982/59	0.363/59	0.317/59
TAL(m1)	0.964/30	0.362/30	0.826/30	0.963/30	0.304/30	1.062/30	0.979/59	0.578/59	0.241/59
TRW(m1)	0.937/30	0.777*/30	5.431/30	0.958/30	0.459/30	0.720/30	0.960/59	0.635/59	3.575/59
		0.482							

Таблица 1. Продолжение.
Table 1. Continued.

Measurements	<i>C. shantungensis</i>			<i>S. minutus</i>			<i>Crocidosoricinae</i> indet.		
	W/n	A/n	JB/n	W/n	A/n	JB/n	W/n	A/n	JB/n
u TAW(m1)	0.913*/30	0.526/30	17.380/30	0.970/30	0.400/30	0.510/30	0.975/59	0.601/59	1.757/59
c	0.964								
u L(m2)	0.921*/30	0.695/30	3.845*/30	0.942/30	0.518/30	1.327/30	0.964*/69	0.942*/69	2.070/69
c	0.947		1.811				0.964*/69	0.942*/69	
TRL(m2)	0.968/30	0.359/30	1.064/30	0.954/30	0.573/30	0.749/30	0.978/65	0.569/65	0.893/65
TAL(m2)	0.947/30	0.510/30	1.766/30	0.938/30	0.601/30	1.585/30	0.978/65	0.562/65	1.895/65
u TRW(m2)	0.923*/30	0.762*/30	2.421/30	0.948/30	0.599/30	0.123/30	0.974/65	0.730/65	0.512/65
c	0.923*	0.762*							
u TAW(m2)	0.936/30	0.795*/30	1.935/30	0.971/30	0.315/30	0.404/30	0.977/65	0.596/65	1.240/65
c		0.795*/							
u L(m3)	0.972/30	0.343/30	0.216/30	0.936/30	0.570/30	6.540*/30	0.971/35	0.551/35	0.779/35
c						0.982			
u W(m3)	0.941/30	0.530/30	1.461/30	0.932/30	0.712/30	0.688/30	0.931*/35	0.799*/35	2.442/35
c							0.920*	0.773*	
Wloc	0.94/30	0.550/30	1.765/30	0.992/30	0.149/30	0.0546/30	0.947/6	0.248/6	0.428/6
u BL(P4)	0.979/30	0.182/30	0.775/30	0.926*/30	0.961*/30	4.778/30	0.978/8	0.165/8	0.204/8
c				0.963	0.563				
u BL(M1)	0.921*/30	0.690/30	6.155*/30	0.939/30	0.720/30	0.992/30	0.978/18	0.219/18	0.414/18
c	0.931		6.256*						
BL(M2)	0.978/30	0.311/30	0.458/30	0.966/30	0.523/30	1.613/30	0.970/17	0.248/17	0.415/17

Примечание: * (серая заливка) – статистически достоверное значение тестов; u – до корректировки; c – после корректировки.

Note: * (grey filling) – significant value of the tests; u – uncorrected; c – corrected.

Таблица 2. Экземпляры *Crocidura shantungensis* с отклоняющимися значениями семи признаков, выявленные критериями Граббса и Туки. В колонках “lower ...” и “upper ...” представлены номера экземпляров со значением параметра (в скобках), лежащие за пределами нижнего или верхнего 95% доверительного интервала. Полу жирным отмечены скорректированные промеры (в мм).

Table 2. Specimens of *Crocidura shantungensis* with extremal values of seven metric characters which were detected by Grubbs and Tukey tests. Columns “lower ...” and “upper ...” contains specimen numbers with values (in parentheses) that lay without lower or upper 95%-confidence intervals. Corrected values (in mm) are bolded.

Characters	Grubbs test		Tukey test	
	lower of 95% conf. interval	upper of 95% conf. interval	lower of 95% conf. interval	upper of 95% conf. interval
MBH (\bar{m} = 1.02 mm)	–	–	–	ZISP 89489 (1.32 => 1.26). ZISP 89490 (1.26 => 1.24). ZISP 89491 (1.36 => 1.30)
TRW(m1) (\bar{m} = 0.81 mm)	–	–	ZISP 89445 (0.76). ZISP 89475 (0.74 => 0.80)	ZISP 89486 (0.82 => 0.86)
TAW (m1) (\bar{m} = 0.89 mm)	ZISP 89475 (0.76 => 0.89)	–	ZISP 89475 (0.76 => 0.89)	–
L(m2) (\bar{m} = 1.21 mm)	–	–	–	ZISP 89486 (1.32 => 1.30). ZISP 89489 (1.29 => 1.28)
TRW(m2) (\bar{m} = 0.76 mm)	–	–	ZISP 89445 (0.71)	–
TAW (m2) (\bar{m} = 0.78 mm)	–	–	–	–
BL(M1) (\bar{m} = 1.40 mm)	–	–	–	ZISP 89427 (1.53 => 1.44). ZISP 89489 (1.58 => 1.57)

Таблица 3. Экземпляры *Sorex minutus* с отклоняющимися значениями шести признаков, выявленные критериями Граббса и Туки. Детали см. в Табл. 2.

Table 3. Specimens of *Sorex minutus* with extremal values of six metric characters which were detected by Grubbs and Tukey tests. See Table 2 for details.

Characters	Grubbs test		Tukey test	
	lower of 95% conf. interval	upper of 95% conf. interval	lower of 95% conf. interval	upper of 95% conf. interval
COR (\bar{m} = 2.50 mm)	ZISP 84519 (1.78 => 2.39)	–	ZISP 84519 (1.78 => 2.39). ZISP 84544 (2.32). ZISP 84555 (2.34)	–
MBH (\bar{m} = 0.71 mm)	–	–	–	ZISP 84518 (0.98 => 0.94)
MRWc (\bar{m} = 1.83 mm)	–	–	ZISP 84541 (1.65 => 1.81). ZISP 84555 (1.66 => 1.78). ZISP 84561 (1.57 => 1.80)	84520 (2.64 => 1.87). ZISP 84524 (1.97 => 1.82). ZISP 84525 (1.97 => 1.81). ZISP 84530 (2.14 => 1.83). ZISP 84536 (1.95 => 1.87)
LLF (\bar{m} = 0.85 mm)	–	–	ZISP 84524 (0.78 => 0.94). ZISP 84535 (0.76). ZISP 84554 (0.75)	ZISP 84523 (0.97 => 0.94). ZISP 84541 (0.99)
L(m3) (\bar{m} = 0.80 mm)	ZISP 84543 (0.71 => 0.76)	–	ZISP 84543 (0.71 => 0.76). ZISP 84546 (0.76 => 0.80)	ZISP 84536 (0.86 => 0.82)
BL(P4) (\bar{m} = 1.17 mm)	–	–	ZISP 84531 (1.12 => 1.16). ZISP 84532 (1.13). ZISP 84548 (1.13 => 1.17). ZISP 84550 (1.10 => 1.16). ZISP 84556 (1.08 => 1.14)	84518 (1.22). ZISP 84520 (1.24 => 1.20)

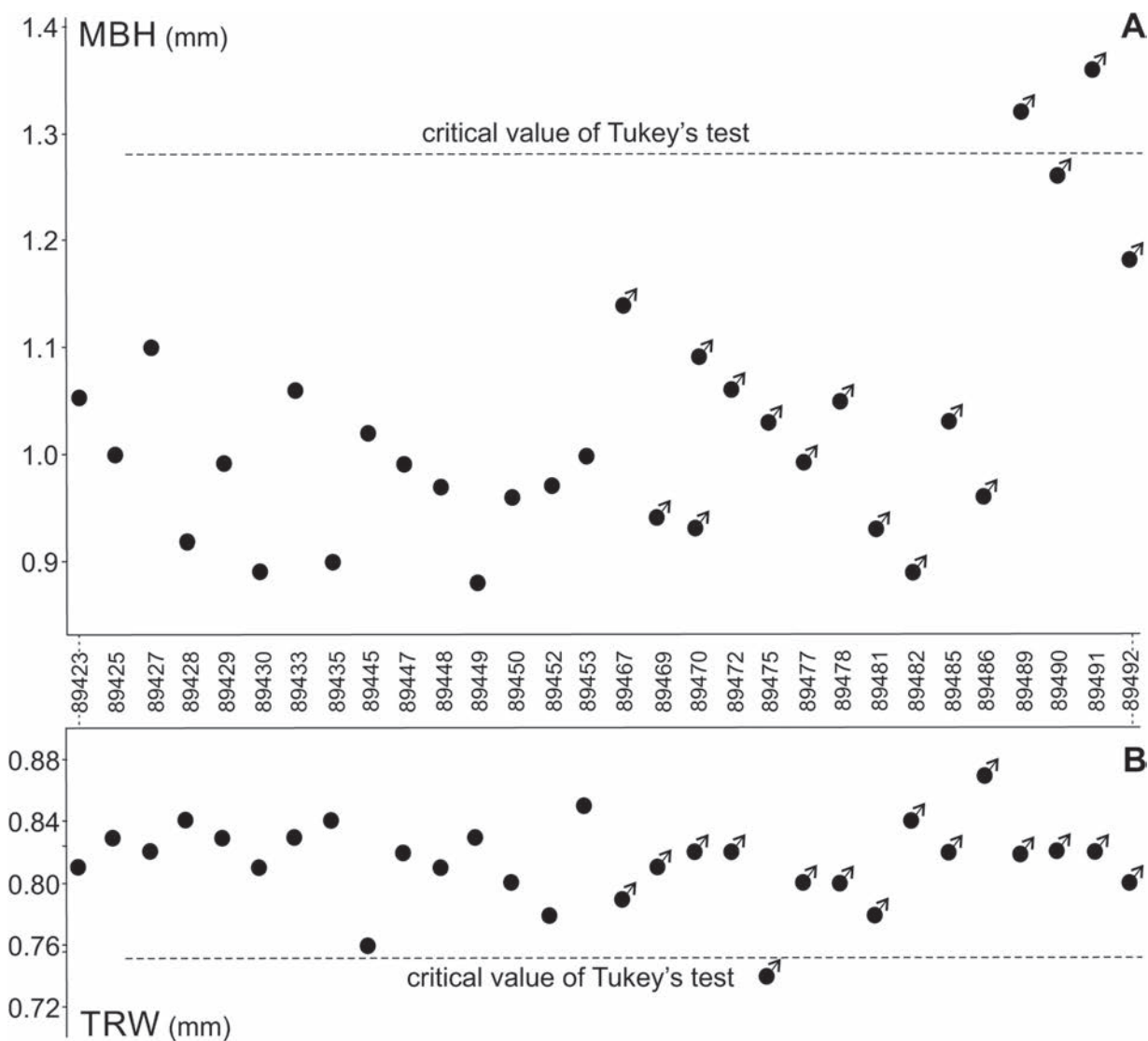


Рис. 1. График распределение значений промеров MBH (А) и TRW(m1) для *Crocidura shantungensis* (30 особей). Черные точки – самки; черные точки со стрелками – самцы; пунктирная линия – 95% доверительный интервал по тесту Туки. Вдоль оси абсцисс номера экземпляров.

Fig. 1. Graph of measures of MBH (A) and TRW(m1) for *Crocidura shantungensis* (30 specimens). Dots – females; dots with arrows – males; dotted line – 95%-confidence interval by Tukey's test. Specimen numbers are along abscissa.

тремальные позиции некоторых экземпляров по трем признакам: MRWc, HC и LUF (Табл. 4). Для L(m2) и W(m3) показано относительно равномерное распределение без заметных отклонений. Критерий Граббса подтвердил отклонение только одного экземпляра по промеру LUF; критерий Туки выявил девять отклоняющихся экземпляров по промерам MRWc, HC, LUF и W(m3) (подробнее см. Табл. 4).

Результаты тестирования и корректировка данных

Как можно увидеть из Табл. 1, не во всех вариантах результаты трех тестов совпадали. Например, параметры TAW(m1), L(m2), BL(M1) для *C. shantungensis*, и L(m3) для *S. minutus*, показали несовпадение наблюдаемого и теоретического распределений по результатам параметрических

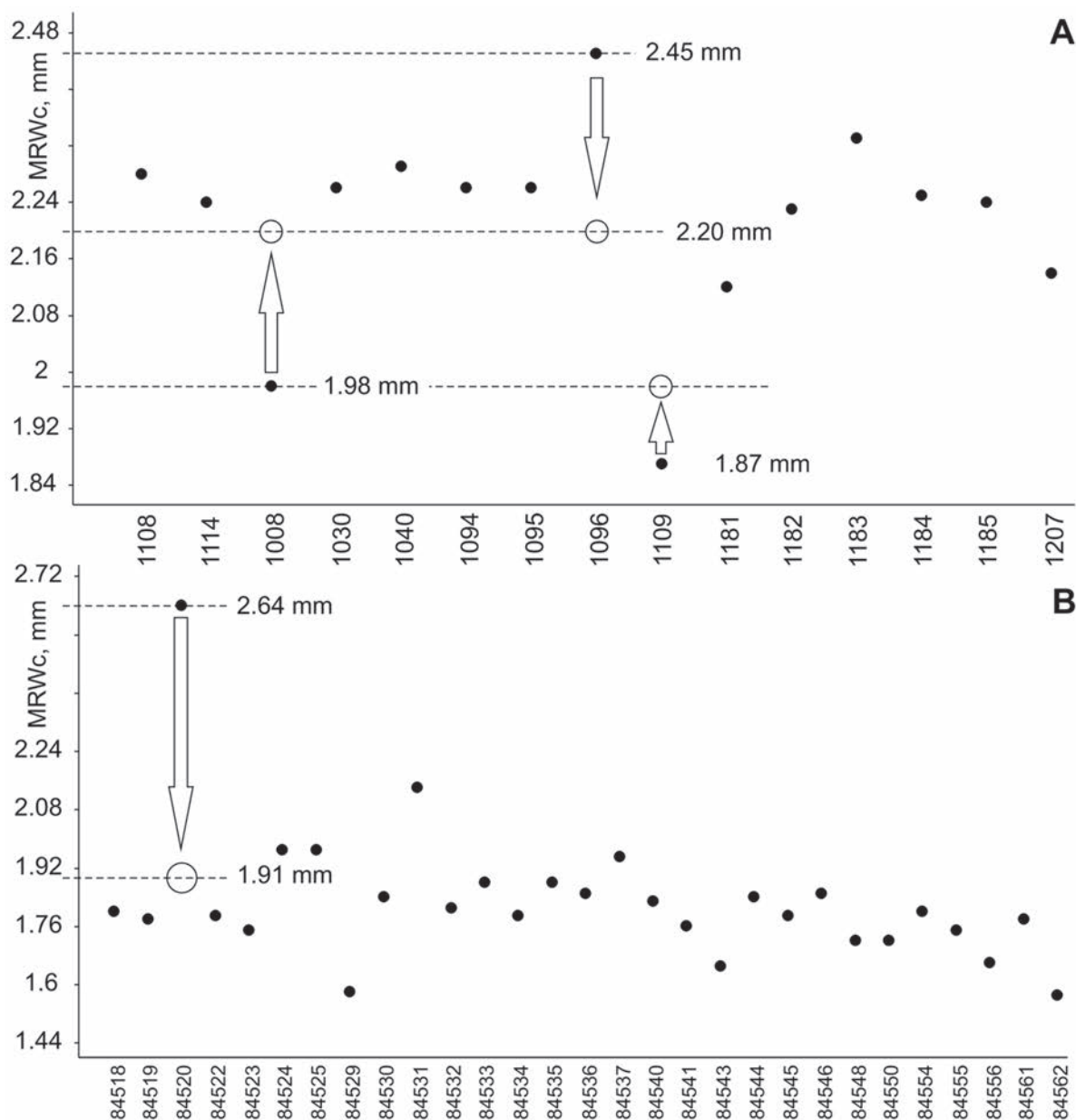


Рис. 2. График распределение значений промера MRWc для *Crocidosoricinae* indet. (15 фрагментов) (А) и *Sorex minutus* (29 особей) (В). Белые стрелки показывают корректировку данных; черные точки – изначальное значение (до корректировки); кружки – скорректированное значение. Другие обозначения см. на Рис. 1.

Fig. 2. Graph of measures of MRWc for *Crocidosoricinae* indet. (15 fragments) (A), and *Sorex minutus* (29 specimens) (B). White arrows show the data correction; dots – initial values (before correction); circles – corrected values. See Fig.1 for other details.

тестов Шапиро-Уилка и Харке-Бера, однако, совпадение распределений по результатам непараметрического теста Андерсона. Это, скорее всего, говорит об ином типе наблюдаемого распределения (логнормальное или гамма-распре-

деление), нежели мы предполагали изначально, что некорректно тестируется параметрическими критериями. В данной ситуации с определенной долей осторожности можно опираться на результаты непараметрического теста, поскольку его

Таблица 4. Экземпляры *Crocidosoricinae* indet. (номера PZC) с отклоняющимися значениями пяти признаков, выявленные критериями Граббса и Туки. Детали см. в Табл. 2.**Table 4.** Specimens of *Crocidosoricinae* indet. (PZC numbers) with extremal values of five metric characters, which were detected by Grubbs and Tukey tests. See Table 2 for details.

Characters	Grubbs test		Tukey test	
	lower of 95% conf. interval	upper of 95% conf. interval	lower of 95% conf. interval	upper of 95% conf. interval
MRWc (\bar{m} = 2.21 mm)	–	–	959/1008 (1.98 => 2.20), 959/1109 (1.87 => 1.98), 959/1181 (2.12 => 2.15), 959/1207 (2.14 => 2.17)	959/1040 (2.29), 959/1096 (2.45 => 2.20), 959/1183 (2.33 => 2.31)
HC (\bar{m} = 1.06 mm)	–	–	959/1199 (0.86), 959/1093 (0.88), 959/1184 (0.97)	959/1183 (1.23)
LUF (\bar{m} = 0.55 mm)	959/1040 (0.37)	–	959/1040 (0.37), 959/1093 (0.47), 959/1188 (0.48 => 0.51), 959/1193 (0.49)	959/1092 (0.64), 959/1094 (0.65), 959/1196 (0.66), 959/1187 (0.64)
L(m2) (\bar{m} = 1.24 mm)	–	–	–	–
W(m3) (\bar{m} = 0.56 mm)	–	–	959/1015 (0.51 => excluded), 959/1026 (0.51), 959/1095 (0.53 => 0.57), 959/1096 (0.53 => 0.56), 959/1111 (0.52), 959/1123 (0.53 => excluded)	–

нечувствительность к типу распределения позволяет говорить об однородности выборки.

Ситуация, когда все тесты, включая и непараметрический критерий Андерсона-Дарлинга, не подтверждает гомогенность выборки, предполагает поиск отклоняющихся значений на предмет ошибки измерения (ввода данных), либо неверного «видового» определения. Перепроверка (повторные промеры) всех отклоняющихся экземпляров, выявленных при визуальном просмотре графиков и по формальным критериям Граббса и Туки, позволила выявить некоторое количество ошибок ввода данных. В частности, по промеру MRWc (Рис. 2) были внесены корректировки для всех трех анализируемых землероек: для *C. shantungensis* откорректировано 10 значений промеров из 11 (Табл. 2); для *S. minutus* – 21 значений из 27 (Табл. 3); для *Crocidosoricinae* indet. – 9 из 24. Отдельно для ископаемой землеройки нужно указать, что высота сочленовного отростка нижней челюсти (HC) не требовала корректировки, поскольку все значения были введены без ошибок; такая же ситуация показана для длины верхней фасетки сочленовного отростка (LUF) где из восьми значений было поправлено только одно; а кроме того, при повторном измерении ширины третьего нижнего моляра (W(m3)) было исключено из анализа два зуба, из-

за дефектов лингвальной части их оснований, что занижало значения промеров (Табл. 4).

После проведенной корректировки отклоняющихся значений мы провели повторное тестирование промеров критериями Шапиро-Уилка, Андерсона-Дарлинга и Харке-Бера. Результаты представлены в Табл. 1 (см. строки, помеченные «с»), где видно, что большая часть из первоначального количества промеров с несоответствием наблюдаемого и теоретического распределений показала соответствие. Стало понятно, например, что более крупные самцы *C. shantungensis*, маркированные в нашем исследовании более стертными зубами, отличаются большими значениями МВН, что приводит к асимметрии распределения, как это видно на Рис. 1А. Из семи признаков *C. shantungensis* несоответствие распределений сохранили только TRW(m2) и TAW(m2). Причина этого не понятна. Остальные признаки необходимо считать однородными (Табл. 1).

Для *S. minutus* из семи первоначальных признаков с несоответствием распределений остался только признак МВН (Табл. 1). Причина также не ясна, но это не может быть связано с различиями полов или возраста – выборка более однородна по степени стертости зубов, чем выборка *C. shantungensis*.

Значения признаков НС и L(m2) ископаемой землеройки *Crocidosoricinae* indet. не требовали корректировки, поэтому они сохранили значения тестов без изменений. Из пяти первоначальных признаков только частично скорректированный LUF изменил значение всех тестов (Табл. 1).

Определение m1 и m2

В ряду нижних коренных зубов *Crocidosoricinae* indet. только m3 легко отличается по размерам, тогда как более крупные m1 и m2 слабо отличаются друг от друга по этому критерию. При рассмотрении нескольких линейных показателей этих зубов видно, что при одинаковой средней длине талонидов m1 и m2 (средние значения в обоих случаях = 0.46 мм) длина тригониды m1 составляет 60.4% общей длины зуба, а тригониды m2 – 56.9%. По качественным признакам нижний коренной зуб m1 *Crocidosoricinae* indet. отличается от m2 по длине протокристыды (prcrd), а именно, prcrd m1 короче такового m2; и по углу наклона протокристыды относительно продольной оси зуба, когда prcrd m1 располагается под заметным

углом к продольной оси коронки зуба, тогда как, prcrd m2 располагается почти перпендикулярно этой оси (Рис. 3). Однако, в некоторых случаях, на зубах разной степени стертости, длина и относительный наклон протокристыды не позволяют сразу определить с каким зубом – m1 или m2, мы имеем дело. Тогда, определение изолированного зуба может быть ошибочным.

Для проверки предположения о возможных ошибках определения изолированных m1 и m2 мы использовали дискриминантный анализ, данными для которого послужили линейные промеры нижних моляров – L, TRL, TAL, TRW, TAW. В результате мы получили точность классификации выборок m1 и m2 89.29%. Такая точность связана с обнаружением 17 неверно определенных зубов: 9 m1=>m2 и 8 m2=>m1. Корректировка их значений в большинстве случаев выявила «ошибки измерений»; а также в двух случаях ошибку определения зуба: ZISP 104325/1130, ранее определенный как m2, был переопределен как m1; ZISP 104325/1130, изначально определенный как m1, был переопределен как m2. После исправления данных точность классификации повторного дискриминантного анализа достигла 95.97%.

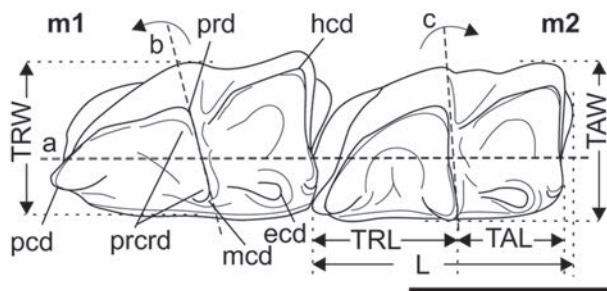


Рис. 3. Рисунок нижних коренных m1–m2 левого зубного ряда таксона *Crocidosoricinae* indet. в проекции жевательной поверхности: a – пунктирная линия показывает продольную ось коронки зуба; b – пунктирная линия со стрелкой вправо показывает значительный передний наклон протокристыды (prcrd) m1 относительно линии a; c – пунктирная линия со стрелкой влево показывает слабый наклон протокристыды m2; ecd – энтоконид; hcd – гипоконид; mcd – метакоид; pcd – параконид; prd – протоконид; L, TAL, TAW, TRL, TRW – см. «Материалы ...». Масштабная линейка = 1 мм.

Fig. 3. Diagrammatic image of left lower molars m1–m2 of *Crocidosoricinae* indet. in occlusal view: a – dotted line shows longitudinal axis of tooth crown; b – dotted line with right-turned arrow shows notable anterior tilt of the protocristid of m1 (prcrd) relative to line a; c – dotted line with left-turned arrow shows weak tilt of the protocristid of m2; ecd – entoconid; hcd – hypoconid; mcd – metaconid; pcd – paraconid; prd – protoconid; L, TAL, TAW, TRL, TRW – see «Material ...». Scale bar = 1 mm.

ОБСУЖДЕНИЕ

В ходе исследования было выявлено, что часть признаков не пригодна для морфометрического анализа межвидовых различий в силу того, что их наблюдаемое распределение не соответствует простым моделям одномерных распределений, которые могут быть протестированы в ходе разведочного анализа критериями Шапиро-Уилка, Андерсона-Дарлина и Харке-Бера. К таким признакам относятся: MBH для *C. shantungensis* и *S. minutus*; TRW(m2) для *C. shantungensis*; MRWc, НС и W(m3) для *Crocidosoricinae* indet. Они могут использоваться при описании изменчивости, но не должны включаться в обсуждение отличительных черт какой-либо формы или вида в ревизиях или при описании нового таксона. Здесь следует отметить, что признак L(m2) ископаемой формы *Crocidosoricinae* indet., с одной стороны, показал несоответствие наблюдаемого и теоретического распределений по результатам тестов Шапиро-Уилка и Андерсона-Дарлинга, что указывает на неоднородность распределения; с другой стороны тесты Граббса и Туки не выявили отклоняющихся

экземпляров по этому признаку. Однако, дискриминантный анализ, использованный нами для поиска ошибок определения m_1 и m_2 , выявил два неверно определенных зуба и целый ряд ошибок измерения. После корректировки данных по результатам дискриминантного анализа повторное тестирование признака $L(m_2)$ показало его однородность: $W = 0.975$ ($p > 0.05$), $A = 0.729$ ($p > 0.05$), $J_B = 1.132$ ($p > 0.05$). Таким образом, можно предположить, что в нашем случае тесты Шапиро-Уилка и Андерсона-Дарлинга показали большую чувствительность к «ошибкам», чем критерии Граббса и Туки.

Таким образом, наше исследование выявило небольшое число ошибок ввода данных, которые повлияли на результаты тестов на однородность, а именно – 1.3% для *C. shantungensis* (10 ошибок на 746 введенных значений); 2.7% для *S. minutus* (20 ошибок на 734 введенных значения); и менее 1% для *Crocidosoricinae* indet. (9 ошибок на 946 введенных значений). Это подтверждает необходимость тщательного разведочного анализа метрических данных стандартными критериями Шапиро-Уилка, Андерсона-Дарлинга и Харке-Бера на соответствие гипотезы об однородности выборок. Кроме того, для фрагментированного палеонтологического материала была показана необходимость проведения дополнительного анализа по поиску ошибок определения остатков, в нашем случае изолированных нижних коренных зубов, с использованием дискриминантного анализа.

Анализ признаков современных видов *C. shantungensis* и *S. minutus*, показал, что в силу особенностей каждой группы землероек, что выражается, например, в проявлении полового диморфизма, характерного для палеарктических белозубок (Zidarova 2015), но не известного для бурозубок (Зайцев и др. [Zaitsev et al.] 2014: 282), нельзя напрямую экстраполировать результаты с рецентного материала на ископаемый. То, что большая часть «неоднородных» признаков всех трех тестируемых землероек не совпадала (Табл. 1), указывает на то, что каждый исследуемый таксон должен тестироваться независимо, и решение об исключении метрических признаков из межвидовых сравнений необходимо принимать отдельно для каждой формы.

Исследование характера распределения 25 метрических признаков *Crocidosoricinae* indet. из местонахождения Шарга 2 показало однород-

ность большей части признаков (Табл. 1). Это, с большой долей вероятности, позволяет говорить о том, что весь исследованный материал по этой форме принадлежит единой генеральной совокупности (локальной популяции), остатки которой накапливались относительно короткое время без существенных метрических изменений. Достаточно короткое время накопления и территориальная ограниченность выборки *Crocidosoricinae* indet. также подтверждается относительной мономорфностью качественных признаков и размером ориктоценоза, который представлен небольшой по площади костеносной линзой в монотонных озерных отложениях местонахождения Шарга 2.

БЛАГОДАРНОСТИ

Авторы благодарят Е.А. Петрову, М.В.Синицу, Э.М. Смирину за помощь при подготовке рукописи и обсуждение результатов. Отдельно благодарим рецензентов А.Ю. Пузаченко и А.А. Лисовского за неформальный подход к нашим результатам и рекомендациям по улучшению текста. Исследование проведено в рамках государственного задания лаборатории териологии ЗИН РАН № АААА-А17-117022810195-3; и было частично поддержано грантом РФФИ № 16-04-00294-а. В исследовании использованы материалы УФК ЗИН РАН № 2-2.20 <http://www.ckp-rf.ru/usu/73561/>.

ЛИТЕРАТУРА

- Anderson T. W. and Darling D. A. 1952. Asymptotic theory of certain “goodness of fit” criteria based on stochastic processes. *The Annals of Mathematical Statistics*, **23**: 193–212.
- Ghasemi A. and Zahedials S. 2012. Normality test for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, **10**: 486–489.
- Gujareti D.N. and Porter D.C. 2010. *Econometría*. Quinta edición. Derechos Reservados, Mexico, 921 p.
- Hammer Ø. 2002. Morphometrics – brief notes. 49 pp. Paläontologisches Institut und Museum, Zürich. Available from: <http://folk.uio.no/ohammer/past/morphometry.pdf> (accessed 17 June 2015).
- Hummer Ø., Harper D.A.T. and Ryan P.D. 2001. PAST: Paleontological Statistics software package for and data analysis. *Palaeontologica Electronica*, **4**: 1–9.
- Hammer Ø., Harper D.A.T. and Ryan P.D. 2008. PAST: Paleontological Statistics, ver. 1.87. 1–88. Available from: http://www.academia.edu/2683729/PAST_PALaeontological_STatistics_ver_1.87

- Kobzar' A.I. 2006.** Applied mathematical statistics. Fizmatlit, Moscow, 238 p. [In Russian].
- Lemeshko B.Ju. and Lemeshko S.B. 2009.** The models of distributions of statistics of nonparametric acceptance criteria for testing complex hypotheses using maximum likelihood estimates. Part I. *Measurement Techniques*, **6**: 6–11. [In Russian].
- Lopatin A.V. and Zazhigin V.S. 2003.** New Brachyericinae (Erinaceidae, Insectivora, Mammalia) from the Oligocene and Miocene of Asia. *Paleontological Journal*, **37**: 62–75.
- Öztuna D., Elhan A.H. and Tüccar E. 2006.** Investigation of four different normality tests in terms of type 1 error rate and power under different distributions. *Turkish Journal of Medical Sciences*, **36**: 171–176.
- Puzachenko Yu.G. 2004.** Mathematics methods in ecological and geographical studies. Akademia, Moscow, 416 p. [In Russian].
- Reumer J.W.F. 1984.** Ruscinian and early Pleistocene Soricidae (Insectivora, Mammalia) from Tegelen (The Netherlands) and Hungary. *Scripta Geologica*, **73**: 1–173.
- Rohlf F.G. and Slice D.E. 1990.** Extension of the Procrustes method for the optimal superimposition of landmarks. *Systematic Zoology*, **39**: 40–59.
- Shapiro S. S. and Wilk M. B. 1965.** An analysis of variance test for normality. *Biometrika*, **52**: 591–611.
- Stefansky W. 1972.** Rejecting Outliers in Factorial Designs. *Technometrics*, **14**: 469–479.
- Tukey J. 1949.** Comparing Individual Means in the Analysis of Variance. *Biometrics*, **5**: 99–114.
- Zaitsev M.V., Voyta L.L. and Sheftel B.I. 2014.** The mammals of Russia and adjacent territories. Lipotyphlans. Nauka, Saint Petersburg, 391 p. [In Russian].
- Zazhigin V.S. and Lopatin A.V. 2000.** The history of the Dipodoidea (Rodentia, Mammalia) in the Miocene of Asia: 1. *Heterosminthus* (Lophocricetinae). *Paleontological Journal*, **34**: 319–332.
- Zidarova S. 2015.** Is there sexual size dimorphism in shrews? A case study of six European species of the family Soricidae. *Acta Zoologica Bulgarica*, **67**: 19–34.

Представлена 2 июня 2017; принята 26 августа 2017.

ПРИЛОЖЕНИЕ 1.

Список проанализированных экземпляров (фрагментов)

APPENDIX 1.

A list of the analyzed specimens (fragments)

Crocidosoricinae indet. из среднемиоценовых отложений местонахождения Шарга 2 (пустыня Шаргинская Гоби, Юго-Западная Монголия; материал был собран В.С. Зажигиным летом 1978 г., и В.С. Зажигиным и Е.В. Девяткиным летом 1979 г.); информация представлена в следующем порядке: сокращение института, полный номер ЗИН РАН для первого экземпляра, полный номер PZC в скобках для первого экземпляра, последующие номера экземпляров даны без префикса (номера после косой черты совпадают для каталога ЗИН РАН и PZC); n = 218:

Crocidosoricinae indet. from the middle Miocene deposits of locality Sharga 2 (Shargain-Gobi Desert, South-Western Mongolia; material was collected by V.S. Zazhigin in summer of 1978, and V.S. Zazhigin and E.V. Devjatkin in summer of 1979); information is in the following order: institutional acronym, full ZISP number for the first specimen, full PZC number for the first specimen in parentheses, the subsequent specimen numbers are given without prefix (the numbers after slash are the same for catalogue of ZISP and PZC); n = 218.

ZISP: 104323 (GIN 959/1010); 104324/1002 (GIN: 959/1002), /1007, /1009, /1045, /1048–1050, /1055, /1056, /1061, /1065, /1068, /1069 /1071, /1073, /1080, /1097, /1098, /1108, /1114, /1116, /1164); 104325/1001 (GIN: 959/1001), /1003–1006, /1008, /1011–1044, /1046, /1047, /1051–1054, /1057–1060, /1062–1064, /1066, /1067, /1070, /1072, /1074–1078, /1081–1096, /1099–1107, /1109–1113, /1115, /1117–1163, /1165–1168, /1170–1199, /1200–1220.

Crocidura shantungensis из окрестностей озера Хасан, Приморье, Россия (сборы М.В. Охотиной, июнь–октябрь 1970); n = 30.

Crocidura shantungensis from the vicinity of the Hasan Lake, Primor'e, Russia, (collected in June–October 1970 by M.V. Okhotina); n = 30.

ZISP: 89423, 89425, 89427–430, 89433, 89435, 89445, 89447–450, 89452–453, 89467, 89469–470, 89472, 89475, 89477–478, 89481–482, 89485–486, 89489–492.

Sorex minutus из окрестностей п. Оскуй, Новгородская обл., Россия (сборы К.А. Третьякова, июль–октябрь 1997); n = 30.

Sorex minutus from the vicinity of the Oskui Village, Novgorod Province, Russia (collected in July–October 1997 by K.A. Tretjakov); n = 30.

ZISP: 84518–520, 84522–525, 84528–537, 84540–541, 84543–546, 84548, 84550, 84554–556, 84561–562.