

УДК 573.018: 517.982.43

ПРИМЕНЕНИЕ СОГЛАСОВАННЫХ КОЭФФИЦИЕНТОВ ЭЛЛИПТИЧЕСКОГО ПРЕОБРАЗОВАНИЯ ФУРЬЕ ДЛЯ СРАВНЕНИЯ ФОРМ БИОЛОГИЧЕСКИХ СТРУКТУР (НА ПРИМЕРЕ ПРИКРЕПИТЕЛЬНЫХ ОРГАНОВ МОНОГЕНЕЙ)

© 2024 г. А. М. Лях

ФГБУН ФИЦ “Институт биологии южных морей им. А.О. Ковалевского РАН”,
”р-т Нахимова, 2, Севастополь, 299011 Россия

E-mail: me@antonlyakh.ru

Поступила в редакцию 23.08.2023 г.

После доработки 18.02.2024 г.

Принята к публикации 19.02.2024 г.

Эллиптическое преобразование Фурье — распространенный метод описания формы объектов уникальной последовательностью коэффициентов, которые позволяют сравнить формы математическими методами. Однако сырые коэффициенты содержат лишние данные, не связанные с формой, что не обеспечивает корректное сравнение. По этой причине коэффициенты нормируют. Это убирает часть лишних данных, но оставляет информацию о зеркальной симметрии и порядке обхода контуров объектов, закодированные в знаках коэффициентов, которые также мешают сравнению форм. В работе описан алгоритм согласования нормированных коэффициентов, нивелирующий влияние упомянутой информации. На примере прикрепительных органов моногеней показаны преимущества использования согласованных коэффициентов для сравнения форм биологических структур.

Ключевые слова: анализ форм, контур объекта, морфометрия, геометрические признаки, зеркальная симметрия, алгоритм, кластерный анализ, танглемма, паразиты рыб, прикрепительные органы моногеней, плоские черви

DOI: 10.31857/S1026347024040015, **EDN:** VINVUJ

Математический анализ форм организмов и их частей активно используется в современных биологических и палеонтологических исследованиях. Анализ форм помогает оценить изменчивость, сравнить, кластеризовать, классифицировать и идентифицировать формы разнообразных объектов математическими методами (Loncaric, 1998; Фурман и др., 2003; Dryden, Mardia, 2012; Pappas *et al.*, 2014; Васильев и др., 2018; Mitteroecker, Schaefer, 2022). Процедуры сравнения составляют основу для многих методов анализа форм.

Для подготовки анализируемых данных форму описывают наборами чисел. Они служат *количественными признаками* формы, с той или иной степенью точности характеризующими ее особенности. Самые простые количественные признаки — это размеры. Другие наиболее известные признаки получают при помощи методов геометрической морфометрии (ГМ) и анализа контуров (АК).

ГМ описывает форму последовательностью точек (меток), расставленных в геометрически или биологически значимых местах (Васильев и др., 2018; Mitteroecker, Schaefer, 2022). Преобразованные

координаты меток служат признаками формы. АК представляет границу объекта последовательностью кривых, которые при объединении приближаются к его форме. В качестве признаков формы используют характеристики этих кривых (Kuhl, Giardina, 1982; Фурман и др. 2003; Vignop, 2011; Tuset *et al.*, 2020; Salili-James *et al.*, 2022).

Основное отличие методов ГМ от АК заключается в том, что методы АК позволяют описать границу объекта целиком, тогда как ГМ не учитывает участки границы, расположенные между метками. По данным АК можно с любой наперед заданной точностью восстановить форму объекта, а при использовании данных ГМ форма участков между метками остается не определенной. Поэтому считают, что признаки АК более точно отражают геометрические особенности формы объекта.

Распространенным методом АК является *эллиптическое преобразование Фурье* (ЭПФ). Оно описывает контур объекта суммой эллипсов, характеристики которых образуют уникальную последовательность коэффициентов (Kuhl, Giardina, 1982; Ferson *et al.*, 1985; Diaz *et al.*, 1989; Crampton, 1995).

Уникальность означает, что любые отличающиеся формы будут описаны отличающимся набором коэффициентов. По этой причине коэффициенты ЭПФ используют для математического сравнения форм.

Так, коэффициенты ЭПФ применяют для анализа форм объектов живой и неживой природы: от панцирей диатомовых микроводорослей (Wishkerman, Hamilton, 2018), листьев (McLellan, Endler, 1998; Neto *et al.*, 2006) и семян растений (Cervantes *et al.*, 2022), крыльев насекомых (Yang *et al.*, 2015), раковин моллюсков (Crampton, 1995; Tuset *et al.*, 2020), прикрепительных органов паразитических червей (Vignon, 2011; Lyakh *et al.*, 2017) и форм прочих организмов (Lishchenko, Jones, 2021), до форм песчинок (Suzuki *et al.*, 2015) и озер на полюсе спутника Сатурна – Титане (Dhingra *et al.*, 2019).

Следует учесть, что на коэффициенты ЭПФ влияют две группы факторов, определяемых геометрией объекта: первая влияет на значения коэффициентов, вторая – на их знаки. Если их не исключить, то даже одинаковые формы будут описывать разные наборы коэффициентов. Значит такие коэффициенты будут невозможно использовать для корректного сравнения форм. По этой причине вычисленные коэффициенты дополнительно преобразовывают.

Первая группа факторов включает угол поворота, положение и размеры контура, и положение точки, в которой начинается контур. Чтобы убрать влияние этих факторов, коэффициенты специальным образом нормируют (Kuhl, Giardina, 1982; Ferson *et al.*, 1985; Diaz *et al.*, 1989).

Тем не менее нормированные коэффициенты все еще содержат лишние данные, не связанные с формой. Так, закономерность чередования знаков коэффициентов кодирует взаимное расположение частей контура и направление обхода границы контура – по или против часовой стрелки. Это вторая группа факторов. Данные о них невозможно исключить из коэффициентов, но можно уравнивать их влияние на знаки коэффициенты всех анализируемых форм.

Для этого, например, можно использовать абсолютные значения коэффициентов, что уберет информацию о знаках (Crampton, 1995). При этом мы потеряем информацию о форме, так как форма, восстановленная из абсолютных значений коэффициентов, будет отличаться от исходной.

При втором способе изображения объектов располагают так, чтобы исключить зеркально-симметричные отличия между ними. После каждый объект обводят контуром в одном и том же направлении часовой стрелки, начиная построение примерно с одной точки. Тогда, после вычисления нормированных коэффициентов ЭПФ, последовательности чередования знаков коэффициентов в большинстве случаев совпадут. Однако у способа есть недостатки. Во-первых, необходимо потратить дополнительное время на согласованное размещение подобных сторон

объектов. Во-вторых, у объектов со сложными асимметричными формами трудно или даже невозможно определить подобные стороны. Это не позволит расположить их правильным образом, что, в итоге, сохранит в нормированных коэффициентах разницу в знаках.

Третий способ, сохраняющий формы, предполагает ручную корректировку знаков нормированных коэффициентов. Знаки коэффициентов меняют следующие действия с контуром: зеркальное отражение, инверсия порядка обхода и смещение точки начала контура в противоположную часть (Лях, 2019). При помощи этих операций контуры преобразуют до тех пор, пока несколько первых коэффициентов модифицированных контуров не приобретут одинаковые знаки. Это приведет к исчезновению отличий в порядке обхода границы и симметрии контуров. Такой методический прием улучшенной нормализации коэффициентов называется *согласованием коэффициентов*, а сами коэффициенты называются *согласованными* (англ., *harmonized*) (Лях, 2019). Недостаток подхода заключается в трудоемкости процедуры согласования, когда необходимо вручную модифицировать каждый анализируемый контур.

Порядок чередования знаков коэффициентов необходимо уравнивать перед проведением любого анализа. Из всех вышеописанных способов процедура согласования знаков коэффициентов является наилучшей, так как она сохраняет исходные формы и для нее не важно, как располагались и в каком направлении были обведены объекты. От единственного недостатка процедуры, связанного с трудоемким ручным преобразованием контуров, избавит алгоритм, автоматически выполняющий данную операцию. Такой алгоритм впервые описан в настоящей работе.

Цель работы заключалась в проверке эффективности использования согласованных коэффициентов ЭПФ для сравнения форм биологических структур.

В качестве объекта исследования выбраны прикрепительные органы плоских паразитических червей из класса Monogenea (Plathelminthes). Форма и размеры прикрепительных органов этих паразитов – крючков и соединительных пластинок – является важным диагностическим признаком, помогающим различать виды (Быховский, 1957). Для этого форму прикрепительных органов описывают наборами размеров, которые анализируют статистическими методами (Dmitrieva *et al.*, 2007; Герасев и др., 2010; Vignon, 2011). Так как разные исследователи используют несовпадающие схемы промеров, то это приводит к труднопоставимым наборам данных и затрудняет таксономическую идентификацию (Герасев и др., 2010; Vignon, 2011), что стимулирует поиск иных способов описания форм прикрепительных органов.

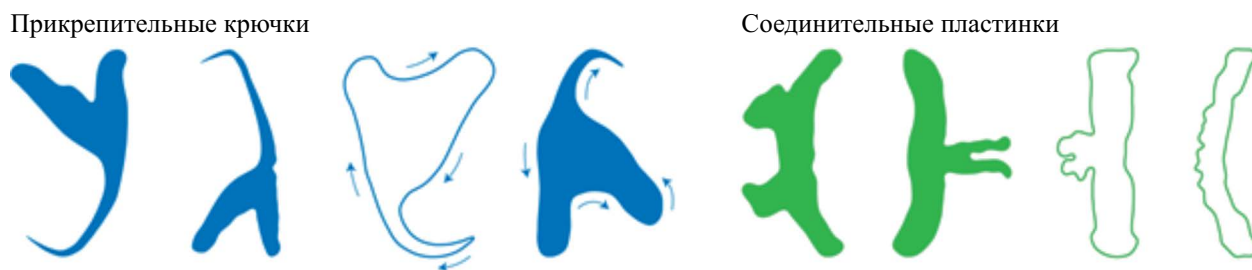


Рис. 1. Контуры прикрепительных органов моногеней. Заливкой обозначен порядок обхода вершин контуров (ориентация контуров): точки закрашенного контура идут против часовой стрелки, а незакрашенного — по часовой.

С другой стороны формы прикрепительных органов моногеней подходят для описания коэффициентами ЭПФ (Vignon, 2011; Lyakh *et al.*, 2017) и тестирования методики сравнения форм. Так, прикрепительный крючок паразитического червя имеет асимметричную форму с ясно выраженными сторонами, которые задают направление острия крючка (рис. 1). Это позволяет визуальную оценку работы алгоритма согласования по однотипному расположению остриев крючков, восстановленных из коэффициентов ЭПФ. Форма соединительной пластинки близка к моносимметричной — зеркально-симметричной относительно продольной оси (рис. 1). У такой формы можно выделить две противоположные стороны — с отростками или прогибом и без, однако две другие стороны отличить невозможно. Эта особенность не позволяет расположить пластинки одинаковым образом, так, чтобы положение их сторон не влияло на коэффициенты. Единственный способ добиться этого — использовать процедуру согласования коэффициентов.

Таким образом прикрепительные органы моногеней являются подходящим модельным объектом для апробации метода ЭПФ и проверки эффективности использования согласованных коэффициентов ЭПФ для сравнения форм биологических структур.

МАТЕРИАЛЫ И МЕТОДЫ

Материалом для работы послужили контуры прикрепительных крючков плоских паразитических червей рода *Ligophorus* Euzet & Seriano, 1977 из класса моногеней (*Monogenea*). Для работы использованы контуры 38 вентральных и 38 дорсальных прикрепительных крючков и 38 вентральных соединительных пластин особей 32 видов (рис. 2).

Контур построен по фотографиям прикрепительного аппарата червей из коллекции паразитов рыб Мирового океана ФИЦ ИнБЮМ (Дмитриева и др., 2016) и по иллюстрациям из паразитологических работ. Оцифровка крючков выполнена в векторном редакторе Inkscape, в котором контуры описаны последовательностью соединенных друг

с другом кривых Безье, дуг окружностей и отрезков прямых. Готовые контуры сохранены в формате масштабируемой векторной графики СВГ (SVG — Scalable Vector Graphics).

Во время построения контуров исследователи не придерживались никаких правил оцифровки: они не исправляли расположение изображенного объекта, произвольно выбирали точку начала контура и направление обводки. То есть вели себя естественным образом, как любой неспециалист, которого попросили обвести изображенные объекты.

В результате был создан набор контуров с нескорректированным влиянием второй группы факторов, влияющих на знаки коэффициентов. По этому набору с помощью компьютерного скрипта было сгенерировано два дополнительных набора с таким же нескорректированным влиянием второй группы факторов. Для этого каждый контур из исходного набора был случайным образом зеркально отражен относительно вертикальной или горизонтальной оси, либо у него было инвертировано направление обхода границы, после чего контур бы помещен в новый набор. Готовые наборы контуров использованы для вычисления коэффициентов ЭПФ и выявления отличий между разными типами коэффициентов.

Коэффициенты ЭПФ вычислены в программе «Эльфурье» (antonlyakh.ru/elfourier/). В качестве входных данных программа использовала СВГ файлы с контурами, а в качестве результата выдала текстовые файлы с нормированными и согласованными значениями коэффициентов. Программа вычислила нормированные коэффициенты по формулам из работ (Ferson *et al.*, 1985; Diaz *et al.*, 1989) и автоматически согласовала их при помощи разработанного алгоритма.

Алгоритм согласования коэффициентов создан путем логического анализа, формализации и математизации положений, высказанных в статьях (Lyakh *et al.*, 2017; Лях, 2019).

После проработки и реализации алгоритма, дальнейшие методические шаги были направлены на выявление различий в результатах сравнения наборов форм, описанных при помощи нормированных, согласованных и абсолютных значений



Рис. 2. Контуры прикрепительных органов видов рода *Ligophorus*, использованные в работе. Контуры выровнены по вертикали и подогнаны к одному размеру при помощи процедуры нормирования, а при помощи алгоритма согласования, описанного в статье, они геометрически согласованы друг с другом.

коэффициентов. Они были выполнены в среде R. Кластерный анализ выполнен с помощью функции `hclust` стандартной библиотеки `stats`, являющейся частью R. Построение дендрограмм и танглеграмм выполнено при помощи функций библиотеки `dendextend`.

Для каждого набора контуров были рассчитаны три типа коэффициентов, которые были кластеризованы методом иерархического кластерного анализа. После чего мы выяснили, чем результаты кластеризаций — дендрограммы — отличаются друг от друга.

Во-первых, мы определили, насколько каждый тип коэффициентов устойчив к изменению второй группы характеристик контуров. Для этого мы выявили различия между дендрограммами, построенными по одному и тому же типу коэффициентов, вычисленных по разным наборам контуров. Во-вторых, мы выявили различия между дендрограммами, построенными по разным типам коэффициентов.

Для визуального определения схожести дендрограмм использована *танглеграмма* (Scornavacca, 2011). Это схема, на которой две дендрограммы с одинаковым составом листьев противопоставлены друг другу, и между совпадающими листьями проведены отрезки. Чем больше отрезков соединяет листья на одинаковых ветках обоих деревьев, тем сильнее похожи эти деревья.

Для количественной оценки различий между дендрограммами использован *индекс гамма-ассоциации*

Бейкера (Baker, 1974). Индекс принимает значение от -1 до 1 , где 1 означает, что деревья идеально соответствуют друг другу, а -1 — что они абсолютно разные.

Полученные результаты легли в основу выводов.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

В процессе работы был написан алгоритм автоматического согласования коэффициентов ЭПФ. Он нацелен на исключение различий в знаках первых нормированных коэффициентов анализируемых контуров, что автоматически приводит к их согласованности. Для понимания работы алгоритма дадим некоторые пояснения.

Коэффициенты ЭПФ принято объединять в четверки, называемые *гармониками*. Гармоники нумеруют, начиная с нуля. Коэффициенты, принадлежащие i -ой гармонике, принято обозначать первыми буквами латинского алфавита: $H_i = \{a_i, b_i, c_i, d_i\}$. Таким образом последовательность коэффициентов ЭПФ, описывающих замкнутый контур, представляет собой бесконечную последовательность гармоник:

$$\{H_0 = \{a_0, b_0, c_0, d_0\}, H_1 = \{a_1, b_1, c_1, d_1\},$$

$$H_2 = \{a_2, b_2, c_2, d_2\}, \dots\}$$

Нулевая гармоника H_0 содержит координаты центра объекта, поэтому в анализе она не участвует. Первая гармоника H_1 задает главный эллипс, который очень грубо повторяет форму объекта, но положение его главных осей помогает нормировать коэффициенты. Прочие гармоники содержат сведения о все более и более мелких деталях формы.

После нормирования значения коэффициентов по абсолютной величине не превышают единицы. При этом с увеличением номера гармоники значения коэффициентов, относящихся к ней, стремятся к нулю. Из-за этого разница между коэффициентами первых и, к примеру, тридцатых гармоник достигает нескольких порядков (рис. 3). По этой причине обычно анализируют только коэффициенты первых гармоник, так как они имеют наибольший вес.

Количество анализируемых гармоник подбирают так, чтобы форма контура, восстановленного из этих гармоник, была идентична исходной форме. Тогда считается, что набор гармоник правильно описывает анализируемую форму. Для того, чтобы по гармоникам вновь построить контур, используют обратное ЭПФ. Точность восстановления формы растет вместе с числом использованных гармоник. Необходимое число гармоник оценивают по формулам (Stampton, 1995). Обычно достаточно 20–50 гармоник.

Стандартная процедура нормирования меняет значения коэффициентов так, что это приводит к повороту восстановленных контуров в вертикальное или горизонтальное положение. При этом коэффициенты первой гармоники принимают особые значения: два коэффициента становятся равными нулю, один — единице, еще один

по абсолютной величине не превышает единицы (табл. 1).

Алгоритм согласования исключает разницу в знаках нормированных коэффициентов, соответствующих первым двум гармоникам, так как именно они сильнее всего влияют на результаты математического сравнения форм. Знаки коэффициентов нельзя менять произвольно, иначе это приведет к искажению формы восстановленного контура. Знаки можно менять только по определенным правилам.

Алгоритм считает, что нормированные коэффициенты двух или более контуров согласованы, если у них совпадают знаки двух идентичных, не равных нулю компонент первых гармоник (b_1 и c_1 при вертикальном, a_1 и d_1 при горизонтальном выравнивании контуров), и, как минимум, двух первых компонент вторых гармоник (a_2 и b_2) (табл. 2) (Лях, 2019).

Работа алгоритма основана на применении к нормированным контурам геометрических преобразований (операторов) так, как это делает человек, согласующий контуры (Лях, 2019). Только алгоритм подбирает необходимые операторы автоматически. Операторы — это: O_1 и O_2 — зеркальное отражение относительно вертикальной и горизонтальной оси, O_c — инверсия направления обхода, O_s — сдвиг начальной точки в противоположную часть контура. Эти операторы не меняют форму контура, а влияют на знаки коэффициентов (рис. 4 а, д). Верно и обратное: для имитации действия оператора достаточно по соответствующему правилу поменять знаки коэффициентов.

Использование для согласования коэффициентов только одного правила бывает недостаточным.

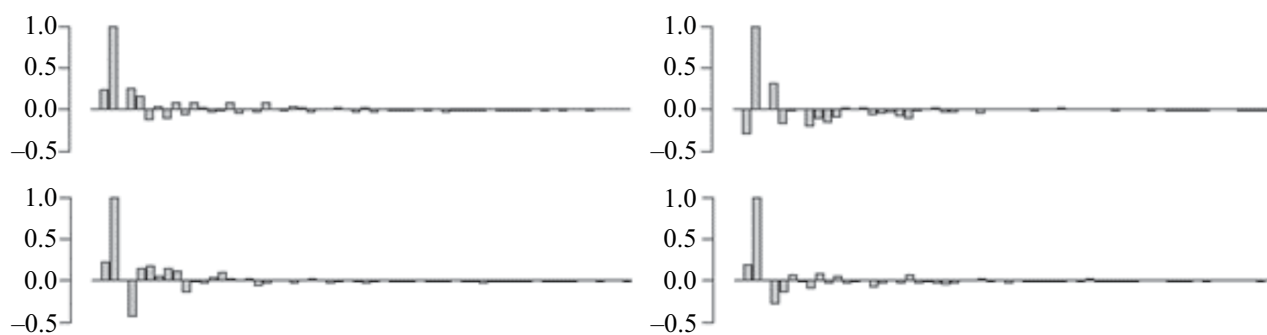


Рис. 3. Значения 60 коэффициентов первых 15 гармоник, описывающих форму четырех прикрепительных крючков. Видно, что с увеличением номера гармоники значения нормированных коэффициентов ЭПФ быстро убывают.

Таблица 1. Значение нормированных коэффициентов первой гармоники

Горизонтально выровненные контуры				Вертикально выровненные контуры			
$a_1 = 1$	$b_1 = 0$	$c_1 = 0$	$ d_1 \leq 1$	$a_1 = 0$	$ b_1 \leq 1$	$c_1 = 1$	$d_1 = 0$

Поэтому правила комбинируют друг с другом. Обозначим символом $\oplus$ комбинацию правил, символом $\times$ положительный коэффициент, а символом $\cdot$ отрицательный или нулевой коэффициент. Тогда комбинация $\oplus$ будет действовать следующим образом:

$\cdot \oplus \cdot = \cdot; \times \oplus \cdot = \times; \cdot \oplus \times = \times; \times \oplus \times = \cdot$

То есть знак итогового коэффициента меняется, когда только один из комбинируемых коэффициентов положительный.

Из четырех операторов согласования ($\emptyset$, $\bullet$, $\times$ и $\cdot$) можно составить 16 уникальных комбинаций, включая нейтральную. Каждой комбинации соответствует единственное правило замены знаков (рис. 4). Все прочие способы изменения знаков коэффициентов повлияют на коэффициенты так, что восстановленный контур изменит форму.

Для согласования коэффициентов двух контуров необходимо подобрать подходящее правило замены знаков. Для этого следует найти *знаковую разность* первых двух гармоник контуров. Она равна

Таблица 2. Соответствие между знаками коэффициентов двух первых гармоник согласованных контуров

Гармоника	H_1		H_2
Выравнивание	вертикальное	горизонтальное	оба случая
Соответствие знаков коэффициентов	—	$[a_1^i] \equiv [a_1^k]$	$[a_2^i] \equiv [a_2^k]$
	$[b_1^i] \equiv [b_1^k]$	—	$[b_2^i] \equiv [b_2^k]$
	$[c_1^i] \equiv [c_1^k]$	—	$[c_2^i] \stackrel{?}{=} [c_2^k]$
	—	$[d_1^i] \equiv [d_1^k]$	$[d_2^i] \stackrel{?}{=} [d_2^k]$

Примечание. Индексы i и k обозначают коэффициенты разных контуров; $[\cdot]$ возвращает знак числа: -1 для отрицательных, 0 или 1 для положительных чисел; $\equiv$ означает совпадение знаков; $\stackrel{?}{=}$ означает “может быть равен, а может и нет”.

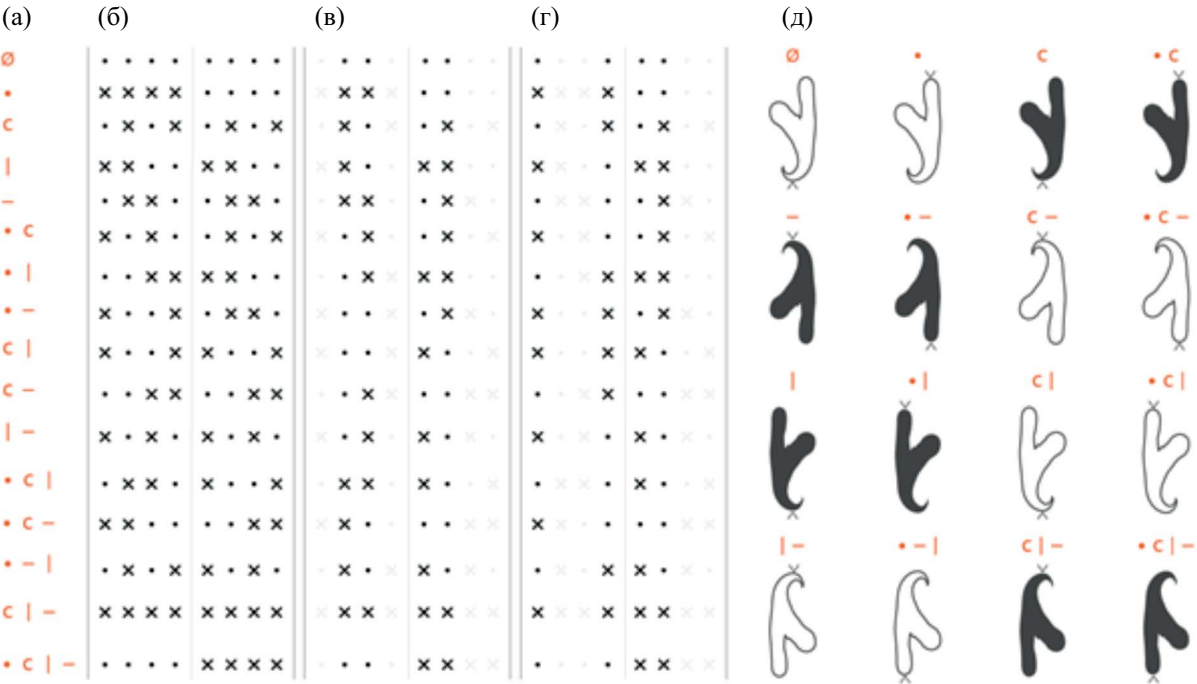


Рис. 4. Правила замены знаков и действие операторов согласования: а — $\emptyset$ — нулевой оператор, не влияющий на контур, $\bullet$ — оператор сдвига начальной точки контура в противоположную часть, $\times$ — оператор инверсии направления обхода, $-$ и $|$ — операторы зеркального отражения относительно горизонтальной и вертикальной осей; б — правила замены знаков коэффициентов; правило записано для первых двух гармоник (восьми коэффициентов), но его следует применить для каждой пары гармоник, начиная с первой; в записи использованы следующие обозначения: $\times$ — инверсия знака, $\cdot$ — знак остается прежним; в, г — маски для поиска операторов, соответствующих знаковой разности согласуемых контуров, при вертикальном (в) и горизонтальном (г) выравнивании контуров; д — результат действия операторов на контуры.

комбинации $\oplus$ символьного представления знаков коэффициентов этих гармоник.

Пусть, например, заданы первые восемь коэффициентов двух контуров A и B , которые выровнены по вертикали (см. табл. 1):

$$A = \{0,00; 0,34; 0,23; 0,00 \mid -0,67; 0,54; 0,77; -0,01\},$$

$$B = \{0,00; -0,17; 0,82; 0,00 \mid 0,42; -0,37; 0,68; -0,03\}.$$

Чтобы найти знаковую разность C , переведем коэффициенты в символьную запись и сложим с помощью $\oplus$. Получим $C = A \oplus B$:

$$A = \{\cdot \times \times \cdot \mid \cdot \times \times \cdot\}$$

$$B = \{\cdot \cdot \times \cdot \mid \times \cdot \times \cdot\}$$

$$C = \{\cdot \times \cdot \cdot \mid \times \times \cdot \cdot\}$$

Знаковая разность служит шаблоном для поиска правила замены знаков, которое согласует контуры. Так как мы согласуем знаки только двух (ненулевых) коэффициентов первой и двух второй гармоники, то в полученной знаковой разности значимыми будут только четыре позиции (отмечены *): у вертикально выровненных контуров — это $\{_ \cdot \cdot _ \mid \cdot \cdot _ _ \}$, у горизонтально выровненных — $\{\cdot _ _ \cdot \mid \cdot \cdot _ _ \}$ (рис. 4 в, г). Тогда, для поиска правила замены знаков следует подобрать такое правило, у которого знаки (символы), стоящие в значимых позициях, будут соответствовать символам в полученной знаковой разности.

Если вернуться к примеру выше, то при вертикальном выравнивании контуров в найденной знаковой разности C значимые знаки будут следующими (подчеркивание $_$ означает незначимый знак коэффициента):

$$C = \{_ \times \cdot _ \mid \times \times _ _ \}$$

Такому шаблону соответствует единственное правило замены знаков — это оператор зеркально-го отражения относительно вертикальной оси O :

$$C \equiv \{\times \times \cdot \cdot \mid \times \times \cdot \cdot\} \equiv O_{\mid}$$

Когда мы применим это правило к контуру B — оно согласует знаки коэффициентов его первых двух гармоник со знаками гармоник контура A :

$$A = \{0,00; 0,34; 0,23; 0,00 \mid -0,67; 0,54; 0,77; -0,01\},$$

$$B \oplus C = \{0,00; 0,17; 0,82; 0,00 \mid -0,42; 0,37; 0,68; -0,03\}.$$

Контур, восстановленный из согласованных коэффициентов $B \oplus C$, будет зеркально отражен по вертикали.

Согласование нескольких контуров выполняется аналогично. Сначала мы выбираем один из контуров и принимаем знаки его коэффициентов за образец. Далее вычисляем знаковые разности между образцом и коэффициентами остальных контуров и для каждого контура подбираем соответствующее правило замены знаков. В конце применяем это правило к знакам коэффициентов согласуемого контура.

Описанный алгоритм позволяет согласовывать контуры без участия человека.

Сравнение форм прикрепительных органов моногеней, описанных разными типами коэффициентов. Напомним, что каждая разновидность прикрепительных органов моногеней — дорсальные крючки, вентральные крючки и вентральные пластинки — была представлена тремя наборами контуров, которые совпадали формами, но отличались порядком обхода границы и расположением сторон: расположением острий крючков и отростков или вогнутостей пластинок (наборы A , B , C на рис. 5). Для каждого набора были вычислены нормированные коэффициенты ЭПФ и абсолютные и согласованные значения коэффициентов.

В результате получилось, что значения в трех группах нормированных коэффициентов отличались друг от друга. Это показало попарное сравнение дендрограмм, построенных в процессе иерархической кластеризации групп коэффициентов. При сравнении дендрограмм индекс гамма-ассоциации Бейкера был близок к нулю, что говорит о несоответствии дендрограмм. Об этом же сигнализируют танглеграммы (рис. 5). Это значит, что нормирование коэффициентов приводит к неустойчивым, нестабильным результатам, зависящим от начальных характеристик контуров. Таким образом, *нормированные коэффициенты нельзя использовать для корректного сравнения форм.*

Алгоритм согласования коэффициентов привел к противоположному результату. Для каждой разновидности прикрепительных структур три группы согласованных значений коэффициентов совпали друг с другом. Контуры прикрепительных крючков, восстановленные из согласованных коэффициентов, совпали геометрически как по расположению и направлению острия, так и по направлению обхода границы. Расположение сторон контуров соединительных пластинок также совпало за исключением двух случаев (рис. 6). По видимому такое исключительное расположение пластинок лучше выражает их геометрические особенности в числовом виде. Таким образом предложенный алгоритм согласования стабильно выдавал один и тот же набор коэффициентов, не зависящий ни от расположения сторон, ни от порядка обхода обрабатываемых контуров.

Сравнение дендрограмм, построенных по абсолютным и согласованным значениям коэффициентов, показало, что они отличаются друг от друга.

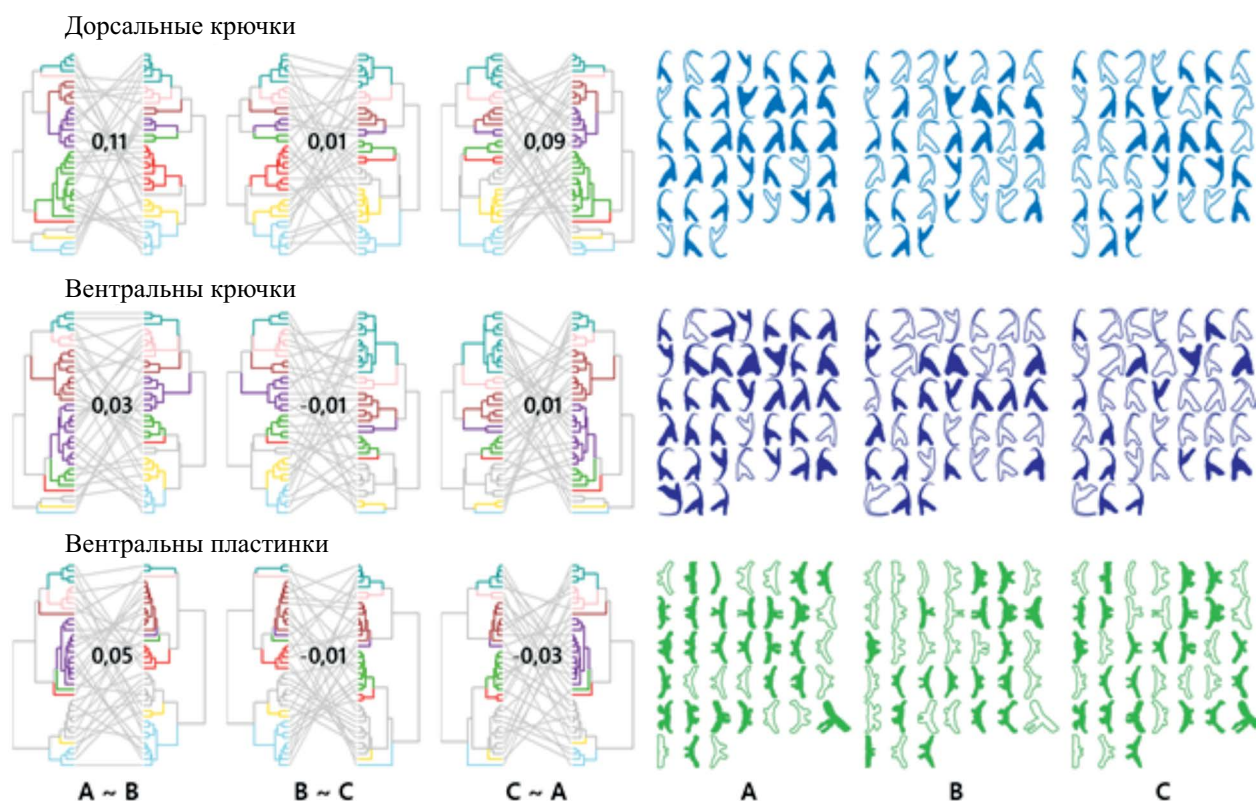


Рис. 5. Сравнение при помощи танглеграмм и индекса гамма-ассоциации Бейкера (число в середине танглеграммы) дендрограмм, построенных по нормированным коэффициентам ЭПФ трех наборов контуров – А, В и С.

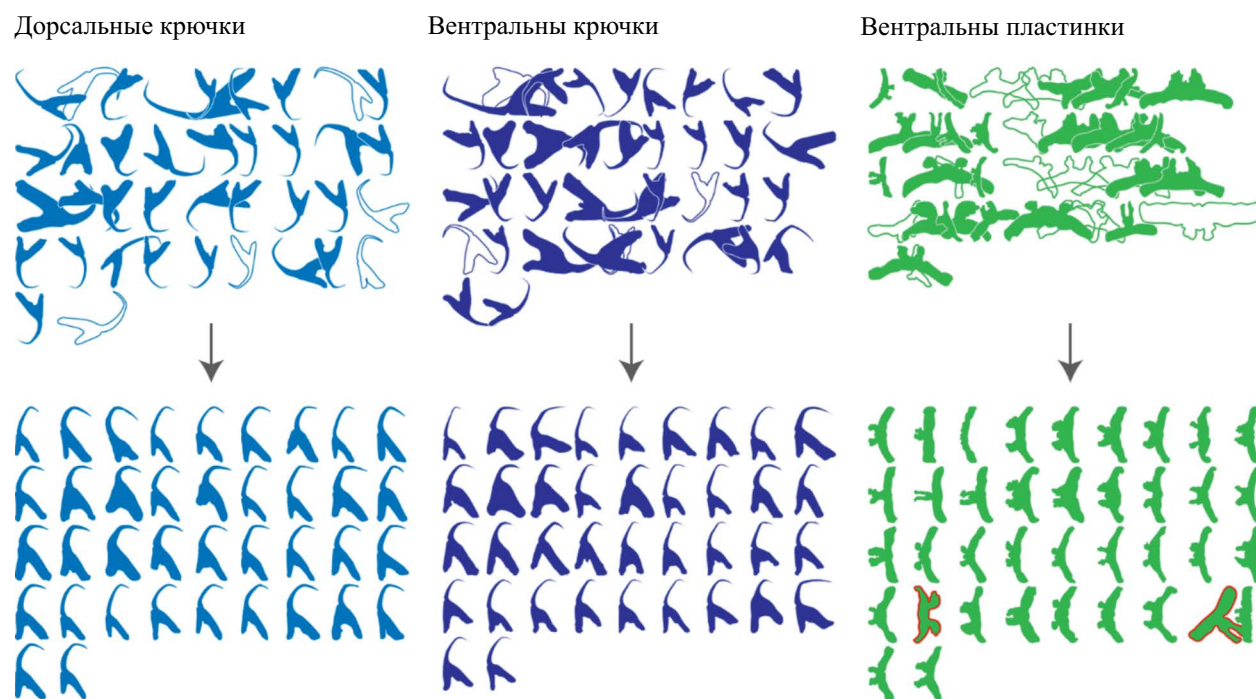


Рис. 6. Алгоритм согласования автоматически преобразовал хаотично расположенные контуры в упорядоченные. В редких случаях положение некоторых контуров (обведены красным) отличалось от остальных, однако это не мешало использовать их числовые описания для сравнения форм.

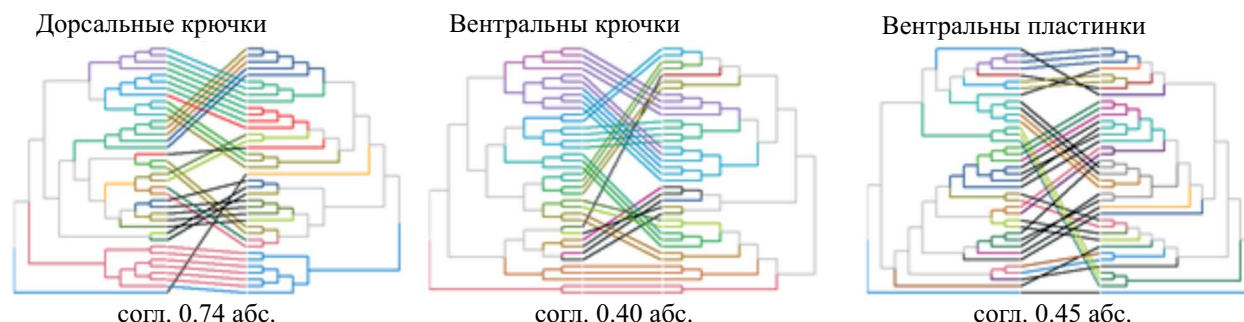


Рис. 7. Сравнение дендрограмм, построенных по согласованным (согл.) и абсолютным (абс.) значениям коэффициентов; числа показывают значения индекса гамма-ассоциации Бейкера.

Наибольшее сходство было между дендрограммами, построенным для дорсальных крючков ($\gamma_{da} = 0.74$), а сходство между двумя другими разновидностями дендрограмм было средним: у вентральных крючков $\gamma_{va} = 0.40$, у вентральных пластинок $\gamma_{vb} = 0.45$ (рис. 7). Очевидно, что результаты кластеризаций согласованных и абсолютных значений коэффициентов не могут быть идентичными, так как среди согласованных коэффициентов встречаются отрицательные значения, отсутствующие у абсолютных значений. Следовательно, *абсолютные значения коэффициентов не могут заменить согласованные коэффициенты*.

Сопоставление дендрограмм по расположению одинаковых видов моногеней показало, что при усечении деревьев на уровне 19 кластеров (половина от количества листьев) абсолютные значения коэффициентов немного лучше улавливали схожесть прикрепительных структур представителей одного вида: из пяти групп моногеней одинаковых видов дендрограммы абсолютных значений в среднем правильно объединили в кластеры 3–4 группы экземпляров видов, а дендрограммы согласованных значений — 2–3. Однако этот вывод нельзя считать окончательным, так как объем проверочных данных был мал: только 5 из 32 видов моногеней были представлены двумя или тремя экземплярами. Для более точного вывода необходимо проанализировать большее число экземпляров каждого вида.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

В целом мы не можем сказать, какая из дендрограмм лучше передает морфологическую схожесть прикрепительных структур. Для этого необходимо располагать идеальной кластеризацией, с которой мы могли бы сравнить полученные результаты. К сожалению, такую идеальную кластеризацию построить невозможно. Это вызвано как техническими причинами — разнообразием методов

кластеризации (Henning, 2022) и разнообразием метрик для сравнения форм (Shen *et al.*, 2013; Bai *et al.*, 2016), из которых невозможно выбрать наилучшую, — так и биологическими причинами — широкой изменчивостью форм структур представителей одного вида, вследствие которой формы видов образуют непрерывные перекрывающиеся спектры.

Однако согласованные коэффициенты сохраняют формы анализируемых объектов неизменными и уравнивают геометрические особенности данных форм, тогда как абсолютные значения коэффициентов искажают оригинальные формы. Вследствие этой особенности согласованные коэффициенты можно использовать для манипуляций с формами. Например, для построения *средних форм объектов из одной группы*. Для этого необходимо найти средние значения коэффициентов группы и восстановить из них среднюю форму (рис. 8).

При помощи этого способа мы можем построить средние формы для кластеров дендрограммы (рис. 9). Такая иерархия форм будет служить для идентификации видов моногеней по форе прикрепительных структур, так как она обеспечивает эффективный поиск форм на основе их сходства с запрашиваемой формой (Srivastava *et al.*, 2005).

Помимо этого средние формы могут использоваться в качестве типовых форм групп объектов, подобных типовым экземплярам видов. Типовые формы можно использовать для выявления характерных (обычных) и аномальных (необычных) форм и последующего анализа причин, приведших к их появлению.

Перечисленные задачи можно решить только при наличии корректных исходных данных. Правильное осреднение форм возможно только при геометрической схожести контуров, которую обеспечивают согласованные коэффициенты ЭПФ. Предложенный в работе алгоритм автоматически выполняет согласование коэффициентов и геометрии контуров. Это позволяет использовать алгоритм в качестве модуля автоматизированной

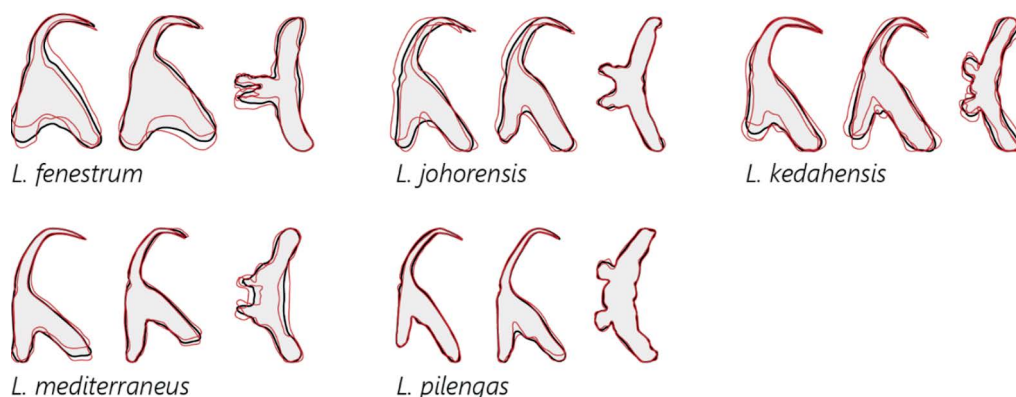


Рис. 8. Средние формы (черные) прикрепительных структур представителей пяти видов рода *Ligophorus* (темно-красные), построенные при помощи осреднения значений согласованных коэффициентов ЭПФ.

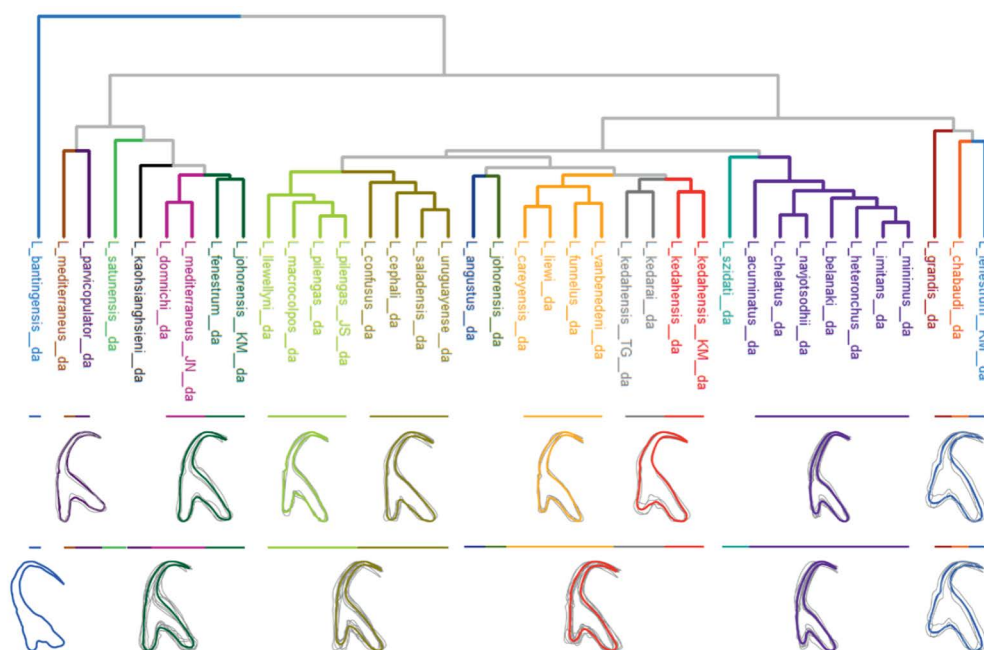


Рис. 9. Дерево средних форм прикрепительных крючков моногеней, которое может служить для идентификации форм. Показаны только два уровня дерева. На первом уровне иерархии средние формы построены при помощи осреднения согласованных коэффициентов ЭПФ, принадлежащих одному кластеру; на втором уровне — при помощи осреднения коэффициентов, принадлежащих соседним кластерам.

системы анализа, распознавания и идентификации форм биологических структур.

Финансирование. Данное исследование было проведено в рамках проекта Государственное задание Федерального исследовательского центра Институт биологии южных морей им. Ковалевского, Россия Академии наук, «Биоразнообразие как основа устойчивое функционирование морских экосистем, критерии и научные принципы его сохранения» (госзадание № 124022400148-4).

БЛАГОДАРНОСТИ

Автор благодарит руководителя отдела экологической паразитологии ФИЦ ИнБЮМ Дмитриеву Е.В. за разрешение использовать контуры прикрепительных структур моногеней в представленной работе. Также автор благодарит рецензента за ценные замечания, которые помогли раскрыть смысл и улучшить читаемость работы.

КОНФЛИКТ ИНТЕРЕСОВ

Автор заявляет, что у него нет конфликта интересов.

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

В данной работе отсутствуют исследования человека или животных.

СПИСОК ЛИТЕРАТУРЫ

- Быховский Б. Е. Моногенетические сосальщики, их система и филогения. М.—Л.: Изд-во АН СССР, 1957. 510 с.
- Васильев А. Г., Васильева И. А., Шкурихин А. О. Геометрическая морфометрия: от теории к практике. М.: КМК. 2018. 471 с.
- Герасев П. И., Дмитриева Е. В., Пугачев О. Н. Методы изучения моногеней (Plathelminthes, Monogenea) на примере паразитов кефалей (Mugilidae) // Зоол.ж. 2010. Т. 89. № 3. С. 1—15.
- Дмитриева Е. В., Лях А. М., Корнийчук Ю. М., Полякова Т. А., Попюк М. П. Электронная коллекция паразитов рыб Мирового океана Института морских биологических исследований им. А. О. Ковалевского // Морской биологический журнал. 2016. Т. 1. № 3. С. 27—31.
https://doi.org/10.21072/mbj.01.3.04
- Лях А. М. Анализ биологических форм на основе согласованных коэффициентов эллиптического преобразования Фурье // Наука Юга России. 2019. Т. 15. № 4. С. 63—70.
- Фурман Я. А., Кревецкий А. В., Передерев А. К., Рожениц А. А., Хафизов Р. Г., Егосина И. Л., Леухин А. Н. Введение в контурный анализ; приложения к обработке изображений и сигналов / Под ред. Я. А. Фурмана. 2 изд., испр. М.: Физматлит. 2003. 592 с.
- Bai X., Donoser M., Liu H., Latecki L. J. Efficient shape representation, matching, ranking, and its applications // Pattern Recogn. Lett. 2016. V. 83. № 3. P. 241—430.
https://doi.org/10.1016/j.patrec.2016.08.007
- Baker F. B. Stability of two hierarchical grouping techniques case 1: Sensitivity to data errors // J. Am. Stat. Assoc. 1974. V. 69. № 346. P. 440—445.
https://doi.org/10.1080/01621459.1974.10482971
- Cervantes E., Rodriguez-Lorenzo J. L., Pozo del D. G., Martin-Gomez J. J., Janousek B., Tocino A., Juan A. Seed silhouettes as geometric objects: new applications of elliptic Fourier transform to seed morphology // Horticulturae. 2022. V. 8. № 10. 974.
https://doi.org/10.3390/horticulturae8100974
- Crampton J. S. Elliptic Fourier shape analysis of fossil bivalves: some practical considerations // Lethaia. 1995. V. 28. P. 179—186.
https://doi.org/10.1111/j.1502-3931.1995.tb01611.x
- Diaz G., Zuccarelli A., Pelligra I., Ghiani A. Elliptic Fourier analysis of cell and nuclear shapes // Comput. Biomed. Res. 1989. V. 22. № 5. P. 405—414.
https://doi.org/10.1016/0010-4809(89)90034-7
- Dhingra R. D., Barnes J. W., Hedman M. M., Radebaugh J. Using elliptical Fourier descriptor analysis (EFDA) to quantify Titan lake morphology // The Astronomical Journal. 2019. V. 158. № 6. P. 1—13.
https://doi.org/10.3847/1538-3881/ab4907
- Dmitrieva E. V., Gerasev P. I., Pron'kina N. V. *Ligophorus llewellyni* n. sp. (Monogenea: Ancyrocephalidae) from the redlip mullet *Liza haematocheilus* (Temminck & Schlegel) introduced into the Black Sea from the Far East // Syst. Parasitol. 2007. V. 67. P. 51—64.
https://doi.org/10.1007/s11230-006-9072-4
- Dryden I. L., Mardia K. V. Statistical shape analysis, with application in R. John Wiley & Sons, Ltd., 2012. 510 p.
- Ferson S., Rohlf J., Koehn R. Measuring shape variation of two-dimensional outlines // Syst. Biol. 1985. V. 34. № 1. P. 59—68.
https://doi.org/10.2307/2413345
- Henning C. An empirical comparison and characterization of nine popular clustering methods // Adv. Data Anal. Classi. 2022. V. 16. P. 201—209.
https://doi.org/10.1007/s11634-021-00478-z
- Kuhl F. P., Giardina C. R. Elliptic Fourier features of a closed contour // Comp. Graph. Image Proc. 1982. V. 18. № 3. 236—258.
https://doi.org/10.1016/0146-664X(82)90034-X
- Lishchenko F., Jones J. B. Application of shape analyses to recording structures of marine organisms for stock discrimination and taxonomic purposes // Front. Mar. Sci. 2021. V. 8. № 667183. P. 1—26.
https://doi.org/10.3389/fmars.2021.667183
- Loncaric A. A survey of shape analysis techniques // Pattern Recogn. 1998. V. 31. №. 8. P. 983—1001.
https://doi.org/10.1016/S0031-2023(97)00122-2
- Lyakh A., Dmitrieva E., Popuyuk M. P., Shikhat O., Melnik A. A geometric morphometric approach to the analysis of the shape variability of the haptor attachment structures of *Ligophorus* species (Platyhelminthes: Monogenea) // Ecologica Montenegrina. 2017. V. 14. P. 92—101.
https://doi.org/10.37828/em.2017.14.10
- McLellan T., Endler J. A. The relative success of some methods for measuring and describing the shape of complex objects // Syst. Biol. 1998. V. 47. № 2. P. 264—281.
https://doi.org/10.1080/106351598260914
- Mitteroecker P., Schaefer K. Thirty years of geometric morphometrics: Achievements, challenges, and the ongoing quest for biological meaningfulness // American Journal of Biological Anthropology. 2022. V. 178. № S74. P. 181—210.
https://doi.org/10.1002/ajpa.24531
- Neto J. C., Mever G. E., Jones D. D., Samal A. K. Plant species identification using Elliptic Fourier leaf shape analysis // Comput. Electr. Agr. 2006. V. 50. № 2. P. 121—134.
https://doi.org/10.1016/j.compag.2005.09.004

- Pappas J. L., Kocielek J. P., Stoermer E. F.* Quantitative morphometric methods in diatom research // *Nova Hedwigia*, Beiheft. 2014. V. 143. P. 281–306.
- Salili-James A., Mackay A., Rodriguez-Alvarez E., Rodrigues-Perez D., Mannack T., Rawlings T. A., Palmer R. A., Todd J., Riutta T. E., Macinnis-Ng. C., Han Z., Davies M., Thorpe Z., Marsland S., Leroi A. M.* Classifying organisms and artefacts by their outline shapes // *J. R. Soc. Interface*. 2022. V. 19. № 195. P. 1–12. <https://doi.org/10.1098/rsif.2022.0493>
- Scornavacca C., Zickmann F., Huson D. H.* Tanglegrams for rooted phylogenetic trees and networks // *Bioinformatics*. 2011. V. 27. P. i248–i256. <https://doi.org/10.1093/bioinformatics/btr210>
- Shen W., Wang Y., Bai X., Wang H., Latecki L. J.* Shape clustering: Common structure discovery // *Pattern Recogn.* 2013. V. 46. P. 539–550. <https://doi.org/10.1016/j.patcog.2012.07.023>
- Srivastava A., Joshi S. H., Mio W., Liu X.* Statistical shape analysis: clustering, learning, and testing // *IEEE T. Pattern Anal.* 2005. V. 27. № 4. P. 590–602. <https://doi.org/10.1109/tpami.2005.86>
- Suzuki K., Fujiwara H., Ohta T.* The evaluation of macroscopic and microscopic textures of sand grains using elliptic Fourier and principal component analysis: Implications for the discrimination of sedimentary environments // *Sedimentology*. 2015. V. 62. № 4. P. 1184–1197. <https://doi.org/10.1111/sed.12183>
- Tuset V. M., Galimany E., Farres A., Marco-Herrero E., Otero-Ferrer J. L., Lombarte A., Ramon M.* Recognising mollusc shell contours with enlarged spines: Wavelet vs Elliptic Fourier analyses // *Zoology*. 2020. V. 140. 125778. <https://doi.org/10.1016/j.zool.2020.125778>
- Vignon M.* Putting in shape – towards a unified approach for the taxonomic description of monogenean haptor hard parts // *Syst. Parasitol.* 2011. V. 79. P. 161–174. <https://doi.org/10.1007/s11230-011-9303-1>
- Wishkerman A., Hamilton P. B.* Shape outline extraction software (DiaOutline) for elliptic Fourier analysis application in morphometric studies // *Appl. Plant Sci.* 2018. V. 6. № 12. e01204. <https://doi.org/10.1002%2Faps3.1204>
- Yang H.-P., Ma C.-S., Wen H., Zhan Q.-B., Wang X.-L.* (2015) A tool for developing an automatic insect identification system based on wing outlines // *Sci. Rep.* 2015. V. 5. № 12786. <https://doi.org/10.1038/srep12786>

Application of harmonized elliptic Fourier transform coefficients for comparing the shapes of biological structures (on the example of the attachment organs of monogenea)

© 2024 A. M. Lyakh

*A.O. Kovalevsky Institute of Biology of the Southern Seas of RAS,
Nakhimov av., 2, Sevastopol, 299011 Russia
E-mail: me@antonlyakh.ru*

Elliptic Fourier transform is a common method of describing the shape of objects by an unique sequence of coefficients that allow comparing the shapes by mathematical methods. However, raw coefficients contain unnecessary data unrelated to the shape, which does not provide a correct comparison. For this reason the coefficients are normalised. This removes some of the superfluous data, but leaves information about mirror symmetry and the order in which the contour vertices are declared, that are encoded in the signs of the coefficients. This also interfere with shape comparison. The paper describes an algorithm for harmonizing the coefficients, leveling the influence of the mentioned information. On the example of attachment organs of monogeneas, the advantages of using harmonized coefficients for comparing the shapes of biological structures are shown.

Keywords: shape analysis, object outlines, morphometry, geometric features, mirror symmetry, algorithm, cluster analysis, tanglegram, parasites, fish parasites, monogenean attachment organs, flatworms.