

ILLiad TN: 385883



ILL Number: -14610973



Borrower: RAPID:DLM

Lending String:
Patron:

Journal Title: Trudy Zoologicheskogo instituta

Volume: 202

Issue:

Month/Year:

1990

Pages: 20-43

Article Author: Emeljanov, A.F. and A.L.
Lobanov.

Article Title: Construction of phylogenetical
tree with use of computer. I. Method of complete
sorting out of possible cladograms.

Imprint:

series

Call #: D-16411
Location: crls
OCLC#: 1405158
ISSN#: 0206-0477

Mail:
Charge
Maxcost:

Shipping Address:
NEW: Main Library

Fax:
Ariel: 128.175.82.31
Odyssey: 206.107.43.75

Transaction Date: 4/15/2019 1:10:23 PM

PAID

А. Ф. Емельянов, А. Л. Лобанов

**ПОСТРОЕНИЕ ФИЛОГЕНЕТИЧЕСКОГО ДРЕВА
С ИСПОЛЬЗОВАНИЕМ ЭВМ.
I. МЕТОД ПОЛНОГО ПЕРЕБОРА
ВОЗМОЖНЫХ КЛАДОГРАММ**

A. F. Emeljanov, A. L. Lobanov. Construction of phylogenetic tree with use of computer. I. Method of total examination of possible cladograms

Осознание того, что разнообразие жизни на Земле есть следствие эволюционных преобразований и дивергенции от общего предка, что история развития живых организмов есть своего рода вариант генеалогии, привело к возникновению филогенетики и построению филогенетических деревьев как наглядного отображения исторического процесса. Филогения стала теоретическим фундаментом естественной системы. Исключительная практическая значимость такой системы постоянно поддерживает потребность в широких филогенетических исследованиях. В свою очередь деятельность эволюционных систематиков стала важным и самым эффективным приближением к построению правдоподобной филогении (Скарлато, Старобогатов, 1974; Татаринев, 1984; Тахтаджян, 1987).

Традиционно эволюционная систематика основывается на двух принципах выделения таксона:

1. Таксон должен быть монофилетичным, т. е. содержать только потомков единого предка (вида-основателя), и не может содержать представителей, не происходящих от него, хотя бы и внешне схожих.

2. Каждый таксон должен характеризоваться определенным, соответствующим его рангу, единством особенностей, в первую очередь морфологических.

Второй принцип допускает существование дочерних таксонов, если какая-либо часть (или части) потомков вида-основателя проделала более значительную эволюцию и вышла за рамки ранга материнского таксона. Таким образом, каждый таксон монофилетичен и морфологически ограничен. Таксон является субъективным, приблизительным, но пока наилучшим интегральным мерилем количества эволюции.

За свою историю практическая филогенетика, т. е. конструирование филогенетических деревьев, прошла ряд этапов: от умозри-

тельных интуитивных схем, основанных на оценке общего сходства и не являющихся по существу филогенетическими, к обоснованным критериям выбора надежных признаков и выработке приемов доказательства истинного родства, от использования единичных признаков к работе с комплексом разнообразных признаков, от прямолинейных представлений о эволюции отдельных признаков к осознанию роли и значения конвергенций, параллелизмов и реверсий. Филогенетика развивалась вместе с эволюционным учением как его часть и вносила свой вклад в развитие его в целом (Расницын, 1988; Шаталкин, 1988; Wiley, 1981).

Крупным событием двойственного значения стало в этой области появление учения Хеннига или «кладизма», которое он сам именовал «филогенетической систематикой». В. Хенниг (Hennig, 1950, 1966) четко сформулировал правила анализа признаков для выяснения филогении и обозначил необходимые для этого понятия во многом удачными терминами: таковы понятия об «апо-» и «плезиоморфии», о «симплезио-» и «синапоморфии». Благодаря Хеннигу процедуры, которые были ясны немногим более проницательным исследователям, стали всеобщим достоянием.

Главным пороком учения Хеннига является отбрасывание второго принципа эволюционной систематики, ведущее к отрицанию традиционной системы, к новому, по существу неверному и неправомерному (в смысле обращения с терминами) понятию о монофилии. Монофилетический таксон по Хеннигу должен всегда включать **всех** потомков вида-основателя, а это означает невозможность существования дочерних таксонов. Если естественный таксон имеет дочерние группы, то по Хеннигу он парафилетичен, а не монофилетичен. Для сохранения первичного смысла термина «монофилия» П. Эшлок (Ashlock, 1971) предложил монофилию в смысле Хеннига называть «голофилией». Учение Хеннига способствовало большой популяризации филогенетических исследований, однако, с другой стороны, неконструктивные по отношению к традиционной систематике положения его учения наносят ей ощутимый вред, замедляют ее совершенствование, смазывают различие в задачах систематики и филогенетики.

Терминология, приемы и особенности филогенетического анализа

Как явствует из предыдущего изложения, использование терминологии, разработанной Хеннигом, и многих рациональных приемов, применявшихся им, не означает принятия его учения в целом.

Термины Хеннига «плезиоморфный» и «апomorphicный», которые казалось бы дублируют старые понятия «исходный» и «производный», удобны тем, что твердо связаны с конкретной процедурой оценки и использования признаков для выявления филогенетического родства. Следует заметить, что словом «признак» иногда обозначают не только единичный элемент распознавания, но и

серию преемственных состояний единой структуры, морфологического локуса, говоря тогда о единичном как о состоянии признака. «Плезиоморфным» называется состояние признака предков, не сохранившееся у части потомков, и заменившееся у них другим, новым, называемым «апоморфным». Понятия эти относительны и не существуют друг без друга. Одно и то же состояние (кроме морфогенетически конечного) в зависимости от того, с чем оно сравнивается, может рассматриваться как апоморфное или плезиоморфное. Тождественное апоморфное или плезиоморфное состояние признака в нескольких группах, рассматриваемых по отдельности, получает название «синапоморфия» или «симплезиоморфия» соответственно. Производное состояние признака, свойственное всей группе в целом, получает название «автапоморфия», но, если ту же группу, обладающую автапоморфией, мы будем рассматривать как собрание отдельных групп более низкого ранга, то тот же признак уже будет расцениваться как «синапоморфный».

В простейшем случае, когда эволюция идет без осложнений появлениями, исчезновениями, конвергенциями, параллелизмами и реверсиями, каждый не преобразующийся далее признак (апоморфия) характеризует монофилетическую группу целиком и служит для ее опознания. Последовательно объединяя в группы по синапоморфиям минимальные выделенные таксоны от самых узких синапоморфий к все более широким, мы получим филогенетическую кладограмму.

Как показывает опыт филогенетических исследований, эволюция практически никогда не идет столь прямым и простым путем. Определение направления изменений признака, т. е. выявление апоморфного и плезиоморфного состояний, или так называемой его полярности, оказывается очень трудной задачей (Queiroz, 1985). Рассмотрим основные обстоятельства, затрудняющие использование в чистом виде метода последовательных синапоморфий.

1. Невозможность априорно однозначно оценить полярность признакоряда, включающего негативное состояние (отсутствие выражения) признака: предшествует ли отсутствие новообразованию; означает ли это то, что структура (или черта) исчезла, или, что в одних случаях она появляется, а в других исчезает?

2. Невозможность априорного установления полярности целиком позитивного признакоряда, особенно на малых отрезках; то ли от простого к сложному, то ли наоборот и т. п.

3. Нестабильность признака. Многие признаки нестабильны, легко меняют количественные показатели — величину, число элементов. Показали эти могут смещаться по неясным (по крайней мере, пока) для нас причинам в обе стороны и менять направление, преобразуясь в рамках закона гомологических рядов Вавилова. При относительности адаптаций (особенно малых) уверенному толкованию с позиций адаптивного смысла они не поддаются.

4. Параллельное (и конвергентное) возникновение признаков. Оно дает ложную синапоморфию.

5. Повторное появление признаков путем инставрации. Многие признаки могут появляться вновь в результате инставрации, нарушая поступательность развития. При этом плезиоморфия как бы сливается с апоморфией. Чем ближе ветви, тем вероятнее инставрация вместо параллелизма, но и тем труднее их распознать, отличить друг от друга.

6. Неустановимость априори порядка перестройки во многих трехзвенных (и более) морфологических рядах — гипотетичность немонотомодальных апоморфий (апоморфные состояния выделены полужирным): $a(\mathbf{b+v})$, или $\mathbf{b(b+a)}$, или $\mathbf{v(a+b)}$. Еще сложнее, если одно из состояний характеризуется отрицательным выражением признака.

7. Неприменимость критерия внешней группы (Farris, 1982). Изолированность признака — утрата плезиоморфии (бывшей симплезиоморфии) в сестринской группе вследствие дальнейшей эммиграции или вымирания. В результате симплезиоморфия с сестринской группой будет выглядеть как автапоморфия избранного таксона или автоплезиоморфия его какой-то части.

Перечисленные затруднения показывают, что по единичным, отдельно взятым рядам признаков нельзя построить достоверную филогению. Однако творческий, дивергентный и необратимый характер макроэволюции в целом позволяет все же выяснить направление эволюции и порядок ветвления в историческом развитии группы путем использования большого и разнообразного комплекса признаков. Отклонения от истинного пути эволюции организма как целого в отдельных рядах признаков не будут совпадать и взаимно компенсируются при достаточном количестве рядов.

Из характера приведенных затруднений также вытекает, что предпочтение надо отдавать позитивным стабильным и сложным признакам, для которых меньше вероятность отклонения от поступательного развития и возможность внезапных резких перемен (Лебедкина, 1980; Le Quesne, 1981).

В практической работе систематика-филогенетика бессознательно и сознательно идет отбор более надежных признаков, т. е. их взвешивание. В первую очередь отбрасываются или не принимаются во внимание признаки нестабильные в пределах низших из подвергающихся анализу единиц: если это виды, то размах индивидуальной изменчивости и его межвидовые соотношения; если это роды, то стабильность в пределах рода и т. д. Признаки, выбранные как годные, также, как правило, явно или неявно оцениваются по их значимости обычно в 2—3-балльной шкале (плохой, хороший, средний). Более дробные оценки, по-видимому, не нужны, так как могут внести излишний субъективизм.

Однако не следует замыкаться только на самых надежных признаках, если мы хотим понять филогению, эволюцию группы во всем многообразии, а не ограничиваться только построением кладограммы. Порядок преобразований менее устойчивых и значительных признаков также не хаотичен и способен дать дополнительные критерии правильности или неправильности полученных выводов. При отсутствии или небольшом числе надежных признаков анализ ненадежных остается единственной опорой.

Процедура реконструкции филогении без применения ЭВМ

Процедура реконструкции филогении на многих этапах тесно связана с процедурой построения системы (Емельянов 1987; Расницын 1988). Рассмотрим ее с самого начала, включая те операции, которые уже проделаны поколениями систематиков прошлого.

1. Предварительная сортировка организмов (видов) на группы по формальному сходству (стихийные народные группирования и доэволюционистская систематика).

2. Освобождение группировок от явно неродственных, ошибочно причисленных представителей, поиск им верного места в системе. Выявление отдельных существенных систематических признаков. Начальная эволюционистская систематика.

Эти два этапа во всех группах можно считать пройденными. Практическая работа начинается с третьего пункта.

3. Массированный поиск и выявление эволюционирующих в пределах исследуемого большого таксона (*ИТ*) признаков и структур, построение многочисленных морфологических рядов по выявленным и ранее известным признакам. На этом этапе следует отвлечься от таксономической структуры *ИТ*, кроме таксонов низшей анализируемой категории. Например, если *ИТ* — семейство, а низший таксон — род, то строить ряды родов, невзирая на трибы и подсемейства.

4. Попутный пересмотр и уточнение объема и состава таксонов низшей анализируемой категории в отношении монофилетичности и соответствия рангу.

5. Разделение *ИТ* на многочисленные мелкие безранговые группировки, отличающиеся повторяющимися сочетаниями выявленных признаков, объединение их последовательно в более крупные группировки более высокого (пока неопределенного) ранга. Возможно, часть из этих группировок впоследствии будет расцелена как трибы, подтрибы, подсемейства.

6. Выявление групп низших порядков (2-го, 1-го), в которых соединяется наибольшее число крайних членов морфологических рядов при том, что другие концы рядов рассеяны в разных группах — т. е. выявление полярности признаков. Плезиоморфны признаки, сходящиеся в немногих низших таксонах, апоморфны признаки, разбросанные по разным низшим таксонам.

7. Использование критерия внешней группы с учетом указанных выше затруднений в выяснении плезиоморфных состояний. Одновременно выявление ошибочно отнесенных к соседним группам отдельных (обычно отклоняющихся) представителей *ИТ*.

8. Построение кладограммы сверху от самых узких синапоморфий к более широким, выявление слабых, наименее обоснованных участков, анализ противоречий с применением взвешивания.

9. Дополнительное выявление новых рядов признаков, если не удастся выполнить программу пункта 8.

10. Если после пункта 9 процедуру пункта 8 все же не удается завершить, производится редукция древа к наиболее надежно обоснованным на разных ранговых уровнях монофилетическим группировкам (это большей частью таксоны) так, чтобы в каждой высшей группировке было 3—4 низших, и проводится сравнительный анализ всех возможных вариантов древа (таких вариантов 3, если группировок 3, или 15, если группировок 4).

11. Если и эти процедуры не дают удовлетворительного результата, то следует еще раз повторить поиск дополнительных признаков, а также применить анализ комбинаций из 5—6 группировок, очень трудоемкий без применения техники.

12. Проверка полученной филогении через анализ согласованности общей картины эволюции как истории последовательных адаптаций, как логичной смены экобиологических особенностей, как географически согласованной картины расселения на современной карте и на фоне палеографических смен, особенно при отсутствии палеонтологических свидетельств, но и при их наличии. Итогом последних операций, завершающих всю работу, должна стать цельная картина эволюции *ИТ*.

Применение ЭВМ в реконструкции филогении

Выполнение вручную описанной выше процедуры наталкивается на серьезные трудности в практической реализации пунктов 8, 10 и 11. При достаточно большом наборе признаков даже самые узкие синапоморфии часто противоречивы и вступают в конфликты. Выделение наименее противоречивых синапоморфий при числе признаков в несколько десятков становится очень трудоемкой задачей даже в том случае, когда *ИТ* содержит всего 4 низших таксона. Прodelанный одним из авторов (Емельянов, 1987) опыт филогенетического анализа группы из 4 таксонов (4 надсемейства цикадовых) с использованием 90 признаков потребовал разработки особых приемов математической оценки конкурирующих вариантов древа и занял около 2 недель только на этапе выбора наиболее вероятного варианта (пункт 8 описанной процедуры), хотя 4 таксона дают всего 15 конфигураций древа. Повторение такого анализа для *ИТ* большего объема выходит за пределы возможностей человека, однако сами приемы оценки вариантов древа при отсутствии единственного и безкон-

фликтного решения показались нам многообещающими и были положены в основу описываемой ниже программы для ЭВМ.

Применение формальных методов (различных алгоритмов, математических приемов и компьютерных программ) для реконструкции филогенетических древ имеет уже большую историю. Достаточно указать на самые известные и широко применяемые подходы (Fitch, Margoliash, 1967; Farris, 1970, 1972; Farris et al., 1970; 1982; Estabrook et al., 1977; 1985; Felsenstein, 1981; Meacham, 1981; Day, 1986). Более подробные сведения о работах в этом направлении можно найти в выполненных недавно обзорах (Felsenstein, 1982; Пименов, 1987; Юшманов, 1987). Уже само большое количество непрекращающихся попыток найти лучший способ нахождения с помощью ЭВМ оптимального древа для реального набора таксонов свидетельствует о трудности этой задачи. Для сколько-нибудь противоречивых исходных данных (а большие наборы признаков для реальных групп таксонов всегда оказываются такими) неоднозначность решения очевидна (Day, Sankoff, 1986), поэтому мы сочли важным начать свою работу в этом направлении с создания программы, позволяющей перебрать все возможные для данного ИТ варианты древа и провести их сравнение по всем интересующим нас показателям. Такая программа является незаменимым инструментом для испытания как своих, так и разработанных другими авторами алгоритмов реконструкции филогении.

Разработанная нами программа CLAD3 для ЭВМ СМ-4 и СМ-1420 написана на языке Фортран-IV и содержит (без учета комментариев) 700 операторов. При ограничении исходных данных 2000 чисел (произведение числа низших таксонов на число признаков не более 2000) для работы программы достаточно 32 Кбайт оперативной памяти и 256 Кбайт дисковой памяти, что позволяет перенести ее практически на любой персональный компьютер. Ниже описана процедура подготовки данных для программы CLAD3 и рассмотрены основные алгоритмы, положенные в ее основу.

Формализация сведений о признаках таксонов

При использовании машинных методов построения филогенетических древ необходима строгая формализация сведений о признаках избранной группы таксонов. Наибольшего внимания требует установление полярности признаков — выделение плезиоморфного (предкового) и апоморфного (продвинутого) состояний (первое обозначается нулем, а второе — единицей). Разработанные нами алгоритмы рассчитаны на строгую дихотомичность признаков (выделение только двух состояний). При внимательном рассмотрении это положение не будет казаться существенным ограничением. Во-первых, дихотомичность ветвлений филогенетического древа естественна для таксонов высоких рангов, и политомичность

может обсуждаться как альтернатива только для таксонов группы вида. Во-вторых, любой политомический ряд состояний признака легко может быть приведен к дихотомическому. Даже в случае, когда плезиоморфное состояние находится в середине такого ряда и состояния естественно упорядочены, можно перейти к объединению всех апоморфных состояний в одно, используя для них формулировку плезиоморфного состояния с отрицанием «не». Анализ полярности признаков может наталкиваться на существенные трудности. Способы их преодоления уже обсуждались нами выше. Здесь мы остановимся только на необходимости оценки достоверности установленной полярности. Машинные методы позволяют установить наиболее вероятную полярность для небольшой части используемого набора признаков, поэтому не стоит исключать из анализа признак, который кажется филогенетически значимым, даже если достоверность определения его полярности равна нулю, но оценки достоверности установления полярности необходимо сделать с максимально возможной точностью. В описываемой программе предусмотрен диапазон этих оценок от 0.0 до 1.0. Если исследователю трудно использовать непрерывную шкалу, то можно рекомендовать ряд из 4 (0.0; 0.3; 0.7; 1.0) или даже 3 (0.0; 0.5; 1.0) значений.

Не менее сложной является задача оценки филогенетической значимости признаков, т. е. придания им различных весов. Проблема взвешивания признаков всегда вызывала яростные споры у систематиков (Funk, Wheeler, 1986; Wheeler, 1986), но, если раньше большинство склонялось к отказу от взвешивания, то сейчас чаще признается необходимость придания признакам весов в той или иной форме (Farris, 1969; Steen, 1973; Neff, 1986; Shaffer, 1986). Главным доводом противников взвешивания является субъективность весов, которые придают одним и тем же признакам разные систематики. При этом забывают, что при построении филогенетических схем интуитивными, «ручными» методами систематик кладет в основу выбора той или иной схемы каждого узла древа весьма ограниченный набор признаков, придавая им тем самым неопределенный и равный, но очень высокий вес. Нам кажется, что сознательная замена этим же систематиком неосознанных и неопределенных весов на продуманные и определенные при подготовке данных к машинной обработке заметно уменьшает долю субъективности, вносимую им. В основу оценки филогенетической значимости признака (у нас она дается в виде баллов в диапазоне 0... 10) должны лечь представления систематика о сложности признака, фундаментальности его генетической основы, лабильности и неустойчивости. Возможность конвергентного возникновения апоморфий оценивается отдельным показателем. Программа для ЭВМ построена таким образом, что, если исследователю 10-балльная шкала кажется избыточной, то он может использовать любую более редуцированную, вплоть до 2-балльной (или отказаться от разных весов вовсе).

Отдельным показателем дается оценка возможности конвергентного возникновения апоморфий, которая уменьшает значимость перерывов на филогенетическом древе по признакам с большой вероятностью конвергенции. Диапазон возможных значений: 0.0...1.0. Для признаков с низкой достоверностью определения полярности, которые могут претерпевать в ходе машинной обработки инверсию, следует оценить возможность конвергентного возникновения плезiomорфных состояний (оценки в том же диапазоне, но со знаком «минус»).

Наконец, необходимо на наш взгляд особо выделять те признаки, у которых апоморфное состояние является «отрицательным», т. е. связано с потерей какой-либо структуры, с редукцией, с упрощением. Такие признаки помечаются —1 в графе данного показателя. Для признаков с низкой достоверностью определения полярности необходимо отметить «отрицательность» плезiomорфных состояний: +1. У остальных признаков в этой графе ставится 0.

При желании можно отказаться от одной или даже от всех оценок признаков, поставив для всех признаков 0 или 1. В этом крайнем случае отбор древа будет осуществляться всего по двум показателям — общему числу перерывов и числу перерывов у конкретных признаков.

Для оценки устойчивости результатов выбора оптимального древа к изменениям набора признаков нами предусмотрено разделение признаков на группы. Программа может работать как со всем набором признаков, так и с определенными их группами. В табл. 1 к группе 1 отнесены признаки преимагинальных стадий, а к группе 2 — признаки имаго.

В качестве примера приводим (см. табл. 1) формализованное представление всех данных о 90 признаках 4 надсемейств цикадовых, подробно рассмотренных ранее (Емельянов, 1987).

Сжатое представление кладограммы в виде матрицы

Эффективность машинных манипуляций с филогенетическими древами (кладограммами) во многом зависит от способа их машинного представления. Производить операции прямо с рисунком кладограммы ЭВМ не может, поэтому кладограмму обычно описывают в более удобной для машинных операций форме, чаще всего в виде матрицы. К этому машинному представлению предъявляется ряд требований: оно должно быть достаточно компактно, однозначно и полностью передавать всю информацию о кладограмме; обеспечивать быстрое получение той информации о кладограмме, которая многократно требуется в процессе анализа кладограммы. От выполнения последнего требования зависит скорость анализа одной кладограммы, а, следовательно, и предельная величина (объем) группы таксонов, которая может быть

Формализованные сведения о признаках 4 надсемейств цикадовых

Характеристики признаков						Состояния признаков у таксонов			
<i>N</i>	<i>NG</i>	<i>FI</i>	<i>AP</i>	<i>PC</i>	<i>NE</i>	Cicadel- loidea	Fulgo- roidea	Cerco- poidea	Cicado- idea
1	2	1	0.7	0.2	1	0	1	0	0
2	2	1	1.0	0.5	-1	1	0	1	0
3	2	1	0.7	0.2	0	0	1	0	0
4	2	2	1.0	0.2	0	0	1	0	0
5	2	3	1.0	0.2	0	0	0	1	1
6	2	2	0.7	0.2	0	0	0	0	1
7	1	2	0.7	0.2	0	1	1	0	0
8	2	3	1.0	0.2	0	1	0	0	0
9	2	2	0.7	0.2	0	1	0	1	1
10	2	1	0.7	0.2	0	0	0	1	1
11	2	1	0.7	0.2	0	1	0	1	1
12	2	3	1.0	0.2	0	0	1	0	0
13	2	1	1.0	0.8	0	1	1	0	0
14	2	1	0.7	0.8	-1	0	1	0	1
15	2	1	0.7	0.2	0	0	1	0	0
16	2	3	1.0	0.2	-1	0	0	0	1
17	2	1	0.7	0.8	0	0	1	0	0
18	2	3	1.0	0.2	0	0	1	0	0
19	1	2	1.0	0.2	0	1	0	0	0
20	2	1	1.0	0.2	1	0	1	0	0
21	2	2	0.7	0.2	-1	1	1	0	0
22	2	1	0.7	0.8	0	1	1	0	1
23	2	1	0.7	0.2	0	1	1	0	0
24	2	2	0.7	0.2	1	1	0	1	1
25	2	1	1.0	0.2	0	0	1	0	0
26	2	3	1.0	0.2	0	0	1	0	0
27	2	1	0.7	0.2	0	1	1	0	0
28	2	1	0.7	0.8	-1	0	1	1	0
29	2	1	0.7	0.2	-1	0	1	0	0
30	2	2	1.0	0.2	-1	0	1	0	0
31	2	2	1.0	0.2	0	0	0	1	0
32	1	3	1.0	0.2	1	0	0	1	1
33	2	1	0.7	0.2	0	0	0	0	1
34	1	3	0.7	0.2	-1	1	1	0	0
35	2	3	1.0	0.2	1	1	0	0	1
36	2	3	1.0	0.2	0	0	0	1	1
37	2	3	1.0	0.2	0	0	1	0	0
38	2	1	1.0	0.2	-1	0	1	0	0
39	2	2	1.0	0.2	0	1	0	0	0
40	2	2	1.0	0.2	-1	1	1	0	0
41	2	3	1.0	0.2	0	1	1	0	0
42	2	3	0.7	0.2	1	1	0	1	1
43	2	1	1.0	0.2	1	0	0	0	1
44	2	3	1.0	0.2	0	0	1	0	0
45	2	1	1.0	0.8	0	0	1	0	1
46	2	2	1.0	0.2	0	1	0	0	0
47	2	2	1.0	0.2	0	1	1	0	0
48	2	1	0.7	0.2	0	0	0	1	0
49	2	2	1.0	0.2	1	0	0	0	1
50	2	1	0.7	0.8	-1	1	0	0	1
51	2	1	1.0	0.8	-1	1	1	0	1
52	2	2	1.0	0.2	0	0	0	1	1

Характеристики признаков						Состояния признаков у таксонов			
<i>N</i>	<i>NG</i>	<i>FI</i>	<i>AP</i>	<i>PC</i>	<i>NE</i>	Cicadel- loidea	Fulgo- roidea	Cerco- poidea	Cicado- idea
53	1	1	1.0	0.2	0	0	0	0	1
54	2	1	1.0	0.8	-1	0	1	1	0
55	2	2	1.0	0.2	0	0	1	0	0
56	2	1	0.7	0.2	0	0	0	0	1
57	2	1	1.0	0.8	0	1	0	0	1
58	2	3	1.0	0.2	0	1	0	0	0
59	2	1	1.0	0.2	0	0	0	0	1
60	2	3	1.0	0.2	1	0	0	1	1
61	2	1	0.7	0.2	0	1	0	1	1
62	2	1	0.7	0.2	-1	1	1	0	1
63	2	1	1.0	0.2	0	0	1	0	0
64	2	3	0.7	0.2	1	1	0	0	0
65	2	2	1.0	0.8	1	1	0	0	1
66	2	3	1.0	0.2	0	0	0	1	1
67	2	1	1.0	0.8	-1	0	1	0	1
68	2	1	0.7	0.8	-1	1	1	0	1
69	2	2	1.0	0.2	0	0	1	0	0
70	2	2	1.0	0.2	0	1	1	0	0
71	2	3	1.0	0.2	0	0	0	1	1
72	2	1	0.7	0.8	0	1	0	0	1
73	2	1	0.7	0.2	0	0	1	1	1
74	1	3	0.7	0.2	1	0	0	1	1
75	2	1	0.7	0.8	0	1	0	0	1
76	2	3	0.7	0.2	0	0	1	1	1
77	2	2	0.7	0.2	0	1	0	1	1
78	2	2	1.0	0.2	0	1	1	0	0
79	2	2	1.0	0.2	0	0	1	0	0
80	1	3	1.0	0.2	1	0	0	1	1
81	1	1	0.7	0.8	1	1	0	0	1
82	2	3	1.0	0.2	1	0	1	0	0
83	2	2	1.0	0.2	0	0	1	0	0
84	2	1	1.0	0.2	-1	0	0	0	1
85	2	1	1.0	0.2	-1	1	0	0	1
86	2	2	1.0	0.2	0	1	1	0	0
87	2	1	1.0	0.2	0	1	0	0	0
88	2	2	0.7	0.2	0	0	0	1	1
89	2	3	1.0	0.2	1	0	1	0	0
90	2	1	0.7	0.2	0	0	0	0	1

Условные обозначения: *N* — номер признака, *NG* — номер группы признаков, *FI* — филогенетическая зависимость, *AP* — достоверность определения полярности, *PC* — вероятность конвергентного возникновения апоморфии, *NE* — наличие отрицательного состояния

обработана программой за приемлемое время. В известных нам способах матричного представления кладограмм последнее требование выполняется плохо, поэтому нами был разработан собственный способ. Кладограмма для *NT* таксонов, содержащая соответственно $NB = NT - 1$ узлов (разветвлений), представляется в виде прямоугольной матрицы размерностью $NB \times NT$. Строка матрицы содержит данные о соответствующем по номеру узле,

а столбец — о таксоне. Каждый узел имеет две выходящие из него ветви, которые рано или поздно оканчиваются листьями, т. е. отражающими таксоны вершинами. Таксоны, входящие в левую ветвь данного узла, помечаются в соответствующих их номерам клетках строки номером узла со знаком «минус», а таксоны правой ветви — номером узла со знаком «плюс» (или без знака). Если таксон не входит в ветви данного узла, то он помечается в этой строке тем же числом, каким был отмечен в предыдущей строке (если в предыдущей строке таксон был единственным в своей ветви, то он помечается здесь и ниже 0). Для иллюстрации в табл. 2 и 3 приведены матрицы, отражающие кладограммы, изображенные на рис. 1 и 2.

Таблица 2

Матрица, отражающая кладограмму на рис. 1

-1	-1	1	1
-2	2	1	1
0	0	-3	3

Таблица 3

Матрица, отражающая кладограмму на рис. 2

1	1	1	-1	1	1	1
-2	-2	2	0	2	2	2
-3	3	2	0	2	2	2
0	0	-4	0	4	4	4
0	0	0	0	-5	5	5
0	0	0	0	0	-6	6

Рассмотрение приведенных матриц показывает, что они позволяют быстро найти ответы на типичные вопросы, возникающие при анализе кладограммы: 1) какие таксоны входят в ветви данного узла? (просмотр одной строки), 2) к каким узлам идут ветви от данного узла? (синхронный спуск со строки данного узла по 2 столбцам, содержащим в этой строке номер узла с разными знаками), 3) какой узел разделяет 2 таксона? (синхронный спуск с 1-ой строки по 2 столбцам) и т. д.

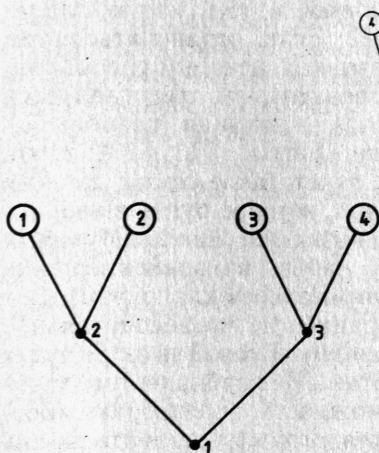


Рис. 1. Кладограмма для 4 таксонов (надсемейства цикадовых).

Темные кружки — узлы; светлые — вершины, соответствующие таксонам.

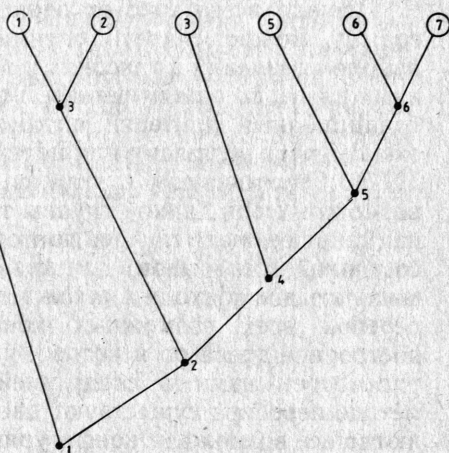


Рис. 2. Кладограмма для 7 цикадовых из 4 надсемейств.

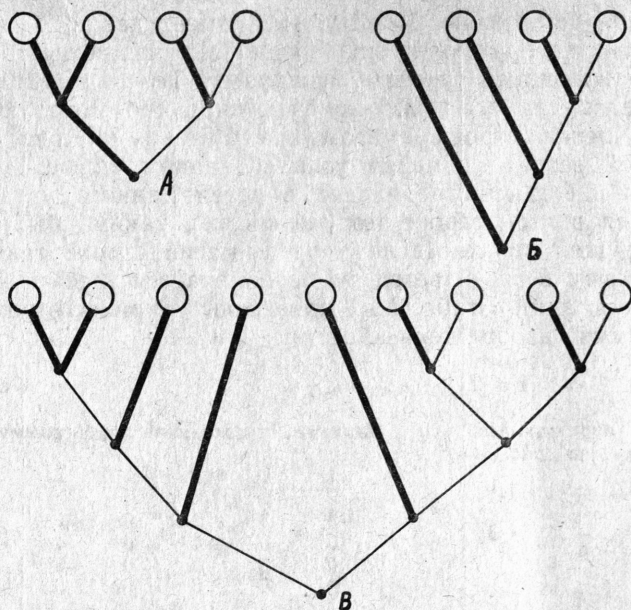


Рис. 3. Схемы кладограмм с 1 (А), 2(Б) и 5(В) перерывами по одному признаку.

Жирными линиями изображены ветви с апоморфным состоянием признака, *тонкими* — с плезiomорфным.

Подходы к построению оптимального филогенетического дерева

Прежде всего требует решения вопрос о том, какую кладограмму можно считать оптимальной? Если ограничиться тем набором сведений о таксонах и их признаках, который рассмотрен нами выше, то самым четким и естественным нам представляется традиционный критерий оптимальности — принцип наибольшей экономности эволюции или парсимонии (Farris, 1973; 1982; Fitch, 1977). Оптимальным в этом смысле будет такое дерево из всех возможных для данной группы таксонов, которое будет включать наименьшее число случаев повторного возникновения апоморфных состояний (перерывов) для заданного набора признаков. Можно выделить два подхода к нахождению оптимальной кладограммы — перебор всех возможных конфигураций и последовательное построение дерева по некоторому алгоритму. Второй подход будет рассмотрен нами в следующей статье. В разбираемом здесь методе перебора существуют две возможности — перебрать абсолютно все возможные конфигурации кладограмм и выбрать из них оптимальную, или прибегнуть к сокращенному варианту перебора и пропустить группы малоотличающихся кладограмм с целью ускорения процесса перебора. Разработанная нами программа

реализует оба варианта перебора. Возможности полного перебора на ЭВМ СМ-1420 довольно ограничены, так как процесс генерации и анализа очередной кладограммы длится около 1 с. Полное число возможных кладограмм для 5 таксонов равно 105, для 6 — 945, а для 7 — 10395. Перебор кладограмм для группы в 7 таксонов продолжается 3 ч. Для работы с группами большего объема приходится использовать вариант сокращенного перебора. В нашей программе он реализуется путем запрета на изменение (перебор) конфигураций одного или нескольких последних узлов кладограммы. Например, анализ группы из 7 таксонов при исключении из перебора трех последних узлов включает всего 633 кладограммы и длится около 10 мин.

Так как все обсуждаемые ниже способы оценки и сравнения кладограмм опираются на понятие перерыва в эволюции признака, необходимо пояснить его и принятый нами метод подсчета числа перерывов по одному признаку. Перерывом мы называем наличие на кладограмме 2 узлов, из которых выходят 1 или 2 ветви с апоморфным состоянием признака, но при этом к этим узлам ведут ветви с плезиоморфным состоянием признака (или один из узлов является корневым). Если таких узлов 3, то засчитывается наличие 2 перерывов и т. д. На рис. 3, изображены для примера кладограммы с 1, 2 и 5 перерывами.

Алгоритм генерации кладограмм

Выбранный нами способ матричного представления кладограмм позволил создать простой алгоритм генерации всех возможных конфигураций для данного числа таксонов. Дело в том, что обычно используемые программистами генераторы перестановок и сочетаний дали бы избыточное число вариантов кладограмм, идентичных в биологическом смысле (такими являются все варианты, полученные из исходного путем поворотов каких-либо пар ветвей, выходящих из одного узла, на 180°). Это иллюстрирует рис. 4. Значительно отличающиеся друг от друга, на первый взгляд, древа *A* и *B* являются для систематики идентичными как филогенетические схемы. Наш эвристический алгоритм генерирует

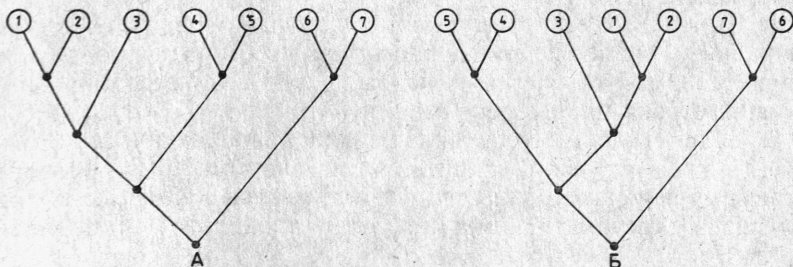


Рис. 4. Две кладограммы, являющиеся идентичными для филогенетики.

матрицы всех принципиально отличающихся кладограмм, среди которых нет идентичных в указанном выше смысле.

Генерация начинается с создания исходной матрицы (табл. 4), в которой на диагонали стоят номера строк со знаком «минус», левее диагонали строки заполнены нулями, а правее — номерами строк без знака «минус». Получение каждой следующей конфигурации выполняется по нижеприведенному алгоритму из предыдущей конфигурации.

Таблица 4

Таблица 5

Исходная матрица,
с которой начинается генерация
набора кладограмм для
группы в 5 таксонов

Матрица, которой заканчивается
генерация кладограмм для
группы в 5 таксонов

-1	1	1	1	1	1	1	1	-1	-1
0	-2	2	2	2	2	2	-2	-1	-1
0	0	-3	3	3	-3	3	0	-1	-1
0	0	0	-4	4	0	0	0	-4	4

1. Номер текущей строки установить равным $NB-1$. (Далее в текущей строке проводятся операции с числами, равными ее номеру. Для краткости в описании алгоритма номер строки с минусом называется просто «минус», а положительный номер строки — «плюс». Сдвигом минуса называется его перемещение вправо или влево на место одного из плюсов, при этом на место минуса ставится плюс. Позиции в строке, занятые нулями или числами, неравными ее номеру, при сдвигах не принимаются во внимание.)

2. Если в текущей строке общее число минусов и плюсов равно 2, то перейти к пункту 6.

3. Начиная с самого правого минуса, сделать попытку сдвига одного минуса на одну позицию вправо. Если число минусов равно числу плюсов и сдвинуть можно только самый левый минус, то перейти к пункту 6 (не сдвигая его). Если удалось сдвинуть самый правый минус, то увеличить номер текущей строки на 1 и перейти к пункту 7. Если сдвиг вообще невозможен, то перейти к пункту 5.

4. Выполнить, если возможно, сдвиг влево всех минусов, расположенных правее перемещенного так, чтобы между сдвинутым в пункте 3 минусом и ними не было плюсов. Увеличить номер текущей строки на 1 и перейти к пункту 6.

5. Если число минусов больше или равно числу плюсов, то перейти к пункту 6. Увеличить число минусов на 1, превратив в минус один из плюсов, и сдвинуть все минусы в крайнее положение слева. Увеличить номер текущей строки на 1 и перейти к пункту 7.

6. Если номер текущей строки больше 1, то уменьшить его на 1 и перейти к 2. Если номер текущей строки равен 1, то прекра-

ДЕРЕВО		НОМЕР	105
	+	T4	4
	+	4	
1	+	T5	5
-1			
1	+	T3	3
	+	2	
	+	T1	1
	+	3	
	+	T2	2

Рис. 5. Машинное изображение кладограммы, соответствующей матрице из табл. 4.

тить выполнение алгоритма — генерация всех возможных конфигураций окончена.

7. Скопировать вышестоящую строку в текущую. Просматривая строку слева направо, найти в ней первый неуникальный элемент (нули не принимать во внимание) и заменить его на минус (номер строки с минусом), а все остальные такие же элементы (только с тем же знаком), стоящие справа, заменить на плюсы.

8. Если номер текущей строки меньше NB (числа узлов), то увеличить его на 1 и перейти к пункту 7. Если номер текущей строки равен NB , то генерация очередного варианта матрицы окончена.

Матрица, которой оканчивается процесс перебора кладограмм для группы из 5 таксонов, приведена в табл. 5.

Для изображения филогенетических древ с помощью печатающего устройства ЭВМ (АЦПУ) нами разработан специальный алгоритм, который не приводится здесь из-за недостатка места. Древo изображается в горизонтальном положении. У конечных вершин печатаются 1—4-буквенные коды таксонов и их номера. В узлах печатаются их номера. На рис. 5 представлено машинное изображение древа, соответствующего матрице табл. 5.

Методы оценки кладограмм

В соответствии с изложенными выше представлениями об оптимальности кладограмм в основу их формальных оценок нами положены два показателя — общее число перерывов на кладограмме по всем признакам и число признаков, имеющих перерывы. Анализ этих 2 показателей совместно с данными по значимости, достоверности определения полярности, вероятности конвергентного происхождения апоморфий и наличию отрицательных апоморфий дает еще 8 комбинированных показателей. Индексы значимости перерывов, достоверности перерывов и конвергентности перерывов вычисляются как суммы (для всех признаков) произведений числа перерывов одного признака на его соответствующие показатели. Для нормирования индекс значимости делится на

10 (так как значимость оценивается по 10-балльной шкале). Отдельно подсчитывается число перерывов, данных позитивными признаками, у которых апоморфии не связаны с потерей каких-то структур. Для удобства анализа признаков, испытавших перерывы в данной кладограмме, они ранжируются по числу перерывов и распределяются по 3 градациям значимости, достоверности и вероятности конвергенции. Отдельно подсчитывается число позитивных признаков, давших перерывы, и дается их список. Как итоговые показатели, производные от числа признаков с перерывами, используется число прерванных признаков высшей значимости, высшей достоверности и высшей вероятности конвергенции. За исключением двух показателей, связанных с вероятностью конвергенции, у всех остальных лучшую кладограмму характеризуют минимальные значения.

Принципы выбора оптимальной кладограммы

После перебора всех возможных кладограмм программа CLAD3 выдает номера 9 деревьев, оказавшихся лучшими по каждому показателю, и полученные значения этих показателей. К окончательному выбору лучшего дерева нужно подходить с большой внимательностью и осторожностью. Случаи, когда одно дерево по всем показателям значительно лучше других, сравнительно редки. Обычно по разным показателям вперед выходят несколько кладограмм, и есть еще кладограммы, которые лишь немного уступают им. Для окончательного анализа следует отобрать несколько близких по экономности кладограмм, отдавая предпочтение абсолютному числу перерывов и прерванных признаков, индексам значимости и достоверности. Для реализации этого в описываемой программе предусмотрена выдача на печать не только лучших деревьев, но и всех близких к ним, уступающих хотя бы по одному показателю не больше, чем на заданную относительную величину. После внимательного анализа отобранных кладограмм по всем выданным программой аспектам необходим этап биологической интерпретации, на котором проверяется логичность эволюции по лучшим кладограммам, соответствие истории изменения признаков данным по истории формирования арсеалов, палеоклиматическим данным и т. п. Только на основе такого комплексного анализа можно сделать окончательный выбор оптимальной конфигурации дерева.

Использование признаков с неясной полярностью

Выше перечислялись трудности, с которыми часто сталкиваются систематики при определении полярности признака, однако отсутствие точного представления о плезиоморфном и апоморфном состояниях признака не является поводом для отказа от его использования в филогенетическом анализе. Существуют несколь-

ко подходов к использованию таких признаков. Простейший и самый верный заключается в поочередном опробовании всех возможных сочетаний неясных признаков в обоих вариантах полярности. Такой метод увеличивает число перебираемых вариантов кладограмм в 2^{NC} раз, где NC — число признаков с неясной полярностью. Очевидно, что этот метод применим лишь при небольшом числе таких признаков и для небольших групп таксонов. Если доля неясных признаков от общего их числа сравнительно невелика, то возможен более экономный подход — опробовать для каждого признака два варианта перебора всех кладограмм (в одной и другой полярности этого признака) и оставить более оптимальную полярность. Этот способ повышает размерность задачи всего в $2 \cdot NC$ раз. Наконец, одним из авторов (А. Ф. Емельяновым) предложен еще более экономный оригинальный подход, названный нами методом гипотетического предка. Трудоемкость метода не зависит от числа неясных признаков. Метод заключается в искусственном инвертировании состояний всех признаков с неясной полярностью таким образом, чтобы они имели только плезиоморфное состояние для одного из таксонов (этот таксон выступает в роли гипотетического предка). На таком искусственно измененном наборе признаков проводится перебор всех кладограмм по описанным выше алгоритмам. Такая процедура повторяется для всех таксонов. Размерность задачи повышается всего в NT раз (где NT — число таксонов). Так как число таксонов обычно значительно меньше числа признаков, то этот подход дает с предыдущим очень заметную экономию. Из полученных NT лучших кладограмм выбирается самая лучшая, и соответствующий ей вариант инверсии неясных признаков оставляется для окончательного анализа. Заодно этим методом выделяется таксон, обладающий самым большим набором плезиоморфий.

Для реализации метода гипотетического предка нами создана программа CLADIV, которая создает из исходного набора данных о признаках NT вторичных наборов, которые затем подвергаются анализу основной программой CLAD3. Перед инвертированием задается пороговое значение достоверности определения полярности. Инверсии подвергаются только признаки, у которых этот показатель имеет значение, меньшее или равное порогу. Инвертирование выполняется только тогда, когда у заданного таксона имеется апоморфное состояние неясного признака. Оценка значимости признака при инвертировании не изменяется. Достоверность полярности устанавливается равной 0.5. Нелегким является вопрос об изменении значения вероятности конвергенции после инвертирования признака. Этот показатель тесно связан с достоверностью. Если вероятность конвергентного возникновения апоморфий велика, то и достоверность определения полярности должна быть высокой, иначе это противоречит здравому смыслу. Если достоверность установления полярности признака была выше

ЛУЧШИЕ ДЕРЕВЬЯ (ИЗ 15 ПЕРЕБРАННЫХ)

НОМЕР ДЕРЕВА	ПОКАЗАТЕЛЬ И ЕГО ЗНАЧЕНИЕ		
13	ИНДЕКС	ЗНАЧИМОСТИ ПЕРЕРЫВОВ	3.60
13	ИНДЕКС	ДОСТОВЕРНОСТИ ПЕРЕРЫВОВ	20.90
11	ИНДЕКС	КОНВЕРГЕНТН. ПЕРЕРЫВОВ	21.80
13	ОБЩЕЕ	ЧИСЛО ПЕРЕРЫВОВ	26
13	ЧИСЛО	ПОЗИТИВНЫХ ПЕРЕРЫВОВ	16
13	ЧИСЛО	ПРИЗНАКОВ С ПЕРЕРЫВАМИ	26
13	ПРЕРВ. ПРИЗНАКОВ БЫШН. ЗНАЧИМ.		3
13	ПРЕРВ. ПРИЗНАКОВ БЫШН. ДОСТОВ.		26
13	ПРЕРВАНО	ПОЗИТИВНЫХ ПРИЗНАКОВ	16

Рис. 6. Фрагмент распечатки программы CLAD3 — конечный результат перебора всех 15 деревьев для 4 таксонов.

```

=====
ДЕРЕВО НОМЕР      1          ВРЕМЯ: 10:07:41
+----- CL      1
-1
I +----- FL    2
+--2
I +-- CR      3
+--3
+-- CD      4
41 ПЕРЕРЫВОВ ПО 36 ПРИЗНАКАМ:
  2  7  9 11 13 14 21 22 23 24 27 28 34 35 40 41 42 45 47
 50 51 54 57 61 62 65 67 68 70 72 75 77 78 81 85 86
В ТОМ ЧИСЛЕ 25 ПЕРЕРЫВОВ ПО 23 ПОЗИТИВНЫМ ПРИЗНАКАМ:
  7  9 11 13 22 23 24 27 35 41 42 45 47 57 61 65 70 72 75
 77 78 81 86
РАСПРЕДЕЛЕНИЕ ПРИЗНАКОВ ПО ЧИСЛУ ПЕРЕРЫВОВ
С 1 ПЕРЕРЫВАМИ:   2  7  9 11 13 14 21 23 24 27 28 34 35 40
                   41 42 45 47 50 54 57 65 67 70 72 75 77 78
                   81 85 86
С 2 ПЕРЕРЫВАМИ:  22 51 61 62 68
ИНДЕКС ЗНАЧИМОСТИ ПЕРЕРЫВОВ =      6.00
МАЛ. ЗНАЧИМОСТЬ:  2 11 13 14 22 23 27 28 45 50 51 54 57 61
                   62 67 68 72 75 81 85
СР. ЗНАЧИМОСТЬ:   7  9 21 24 40 47 65 70 77 78 86
БЛЖ. ЗНАЧИМОСТЬ: 34 35 41 42
ИНДЕКС ДОСТОВЕРНОСТИ ПЕРЕРЫВОВ =    33.80
МАЛ. ДОСТОВЕРНОСТЬ: 2  7  9 11 13 14 21 22 23 24 27 28 34 35
                   40 41 42 45 47 50 51 54 57 61 62 65 67 68
                   70 72 75 77 78 81 85 86
РАСПРЕДЕЛЕНИЕ ЧИСЛА ПЕРЕРЫВОВ ПО УЗЛАМ:
  31 10 0
ИНДЕКС КОНВЕРГЕНТНОСТИ ПЕРЕРЫВОВ =    19.60
=====

```

Рис. 7. Фрагмент распечатки программы CLAD3 — дерево номер 1 и все данные о нем.

0.5, то новое значение вероятности конвергенции после инверсии вычисляется как разница между единицей и старым значением. Если достоверность была равна 0.5 или ниже, то значение вероятности конвергенции не изменяется. Если отметка об отрицательных состояниях признака не равна нулю, то ее знак при инверсии изменяется на обратный.

Результаты испытания программы

Программа CLAD3 испытана на приведенном в табл. 1 наборе данных о 4 надсемействах цикадовых. На рис. 6 показан фрагмент распечатки, сделанной программой — конечный результат перебора всех 15 кладограмм. По всем 9 показателям, использованным в программе, лучшим (или не хуже других) оказалось древо номер 13 — то же самое, которое оказалось оптимальным при ручном анализе (Емельянов, 1987: рис. 34 на стр. 90). На рис. 7, 8 и 9 приведены фрагменты распечатки с полными данными о деревьях с номерами 1, 4 и 13. Деревя 1 и 4 обсуждались как наиболее правдоподобные до проведенного в 1987 г. детального анализа.

ДЕРЕВО НОМЕР		4	ВРЕМЯ: 10:08:18																		
+	----- FL	2																			
---	1																				
I	+----- CL	1																			
+	---2																				
I	+-+ CR	3																			
+	---3																				
	+-+ CD	4																			
38 ПЕРЕРЫВОВ ПО 33 ПРИЗНАКАМ:			2	7	13	14	21	22	23	27	28	34	35	40	41	45	47	50	51	54	57
			61	62	65	67	68	70	72	73	75	76	78	81	85	86					
В ТОМ ЧИСЛЕ 22 ПЕРЕРЫВОВ ПО 26 ПОЗИТИВНЫМ ПРИЗНАКАМ:			7	13	22	23	27	35	41	45	47	57	61	65	70	72	73	75	76	78	81
			86																		
РАСПРЕДЕЛЕНИЕ ПРИЗНАКОВ ПО ЧИСЛУ ПЕРЕРЫВОВ																					
C	1 ПЕРЕРЫВАМИ:		2	7	13	14	21	23	27	28	34	35	40	41	45	47					
			50	54	57	65	67	70	72	73	75	76	78	81	85	86					
C	2 ПЕРЕРЫВАМИ:		22	51	61	62	68														
ИНДЕКС	ЗНАЧИМОСТИ ПЕРЕРЫВОВ =		5.40																		
МАЛ.	ЗНАЧИМОСТЬ:		2	13	14	22	23	27	28	45	50	51	54	57	61	62					
			67	68	72	73	75	81	85												
СР.	ЗНАЧИМОСТЬ:		7	21	40	47	65	70	78	86											
БОЛ.	ЗНАЧИМОСТЬ:		34	35	41	76															
ИНДЕКС	ДОСТОВЕРНОСТИ ПЕРЕРЫВОВ =		31.70																		
БОЛ.	ДОСТОВЕРНОСТЬ:		2	7	13	14	21	22	23	27	28	34	35	40	41	45					
			47	50	51	54	57	61	62	65	67	68	70	72	73	75					
			76	78	81	85	86														
РАСПРЕДЕЛЕНИЕ ЧИСЛА ПЕРЕРЫВОВ ПО УЗЛАМ:																					
			34	14	0																
ИНДЕКС	КОНЦЕРТНОСТИ ПЕРЕРЫВОВ =		21.20																		

Рис. 8. Фрагмент распечатки программы CLAD3 — древо номер 4 и все данные о нем.

```

=====
ДЕРЕВО НОМЕР      13          ВРЕМЯ: 10:12:10
  +--- CL         1
  +--2
  I +--- FL       2
  --1
  I +--- CR       3
  +--3
  +--- CD         4
26 ПЕРЕРЫВОВ ПО 26 ПРИЗНАКАМ:
  2  9 11 14 22 24 28 35 42 45 50 51 54 57 61 62 65 67 68
 72 73 75 76 77 81 85
В ТОМ ЧИСЛЕ 16 ПЕРЕРЫВОВ ПО 16 ПОЗИТИВНЫМ ПРИЗНАКАМ:
  9 11 22 24 35 42 45 57 61 65 72 73 75 76 77 81
РАСПРЕДЕЛЕНИЕ ПРИЗНАКОВ ПО ЧИСЛУ ПЕРЕРЫВОВ
С 1 ПЕРЕРЫВАМИ:  2  9 11 14 22 24 28 35 42 45 50 51 54 57
                  61 62 65 67 68 72 73 75 76 77 81 85
ИНДЕКС ЗНАЧИМОСТИ ПЕРЕРЫВОВ =      3.60
МАЛ. ЗНАЧИМОСТЬ:  2  11 14 22 26 45 50 51 54 57 61 62 67 68
                  72 73 75 81 85
СР. ЗНАЧИМОСТЬ:   9 24 65 77
БОЛ. ЗНАЧИМОСТЬ: 35 42 76
ИНДЕКС ДОСТОВЕРНОСТИ ПЕРЕРЫВОВ =    20.99
БОЛ. ДОСТОВЕРНОСТЬ: 2  9 11 14 22 24 28 35 42 45 50 51 54 57
                   61 62 65 67 68 72 73 75,76 77 81 85
РАСПРЕДЕЛЕНИЕ ЧИСЛА ПЕРЕРЫВОВ ПО АЗЛАН:
  26 0 0
ИНДЕКС КОНВЕРГЕНТНОСТИ ПЕРЕРЫВОВ =    21.80

```

Рис. 9. Фрагмент распечатки программы CLAD3 — древо номер 13 и все данные о нем.

Легко заметить, что по всем показателям они гораздо хуже древа 13 и в предлагаемой системе оценок конкурировать с последним не могут.

Выводы

Разработанная и испытанная авторами программа CLAD3, осуществляющая поиск наиболее оптимального древа путем полного перебора всех возможных конфигураций, является прекрасным инструментом для исчерпывающего филогенетического анализа групп, состоящих из 4—8 таксонов. Для таких небольших групп гарантируется нахождение древа, являющегося самым оптимальным из всех возможных по любому из 9 критериев оптимальности, заложенных в программу. Для групп большего объема полный перебор невозможен, так как он занимает даже на ЭВМ очень много времени. Для групп объемом 9—12 таксонов возможен частичный перебор, с нахождением оптимального деления на подгруппы, которые затем могут быть подвергнуты по отдельности полному анализу. Кроме того, программа CLAD3 позволяет проводить на группах в 4—8 таксонов исчерпывающую проверку возможностей любых других алгоритмов конструирования филогенетических древ с дихотомическим ветвлением.

- Емельянов А. Ф. Филогения цикадовых (Homoptera, Cicadina) по сравнительно-морфологическим данным // Труды ВЭО, 1987.— Т. 69.— С. 19—109.
- Лебедкина Н. С. Ведущие признаки в филогенетике // Морфологические аспекты эволюции.— М.: Наука, 1980.— С. 53—64.
- Пименов М. Г. Математические методы и вычислительная техника в систематике высших растений // Итоги науки техники.— Серия «Ботаника».— Т. 8.— Вып. 2.— М.: ВИНТИ, 1987.— 96 с.
- Расницын А. П. Филогенетика // Современная палеонтология.— Т. 1.— М.: Недра, 1988.— С. 480—497.
- Скарлато О. А., Старобогатов Я. И. Филогенетика и принципы построения естественной системы // Труды ЗИН АН СССР, 1974.— Т. 53.— С. 30—46.
- Татаринов Л. П. Кладистический анализ и филогенетика // Палеонтол. журн., 1984.— № 3.— С. 3—16.
- Тахтаджян А. Л. Система магнолиофитов.— Л.: Наука, 1987.— 439 с.
- Шаталкин А. И. Биологическая систематика.— М.: Изд-во МГУ, 1988.— 184 с.
- Юшманов С. В. Методы теории графов в эволюции. Построение филогенетических схем / Матем. кибернетика и ее прилож. к биологии.— М.: Наука, 1987.— С. 101—140.
- Ashlock P. D. Monophyly and associated terms // Syst. Zool., 1971.— Vol. 20, N 1.— P. 63—69.
- Day W. H. E. Analysis of quartet dissimilarity measures between undirected phylogenetic trees // Syst. Zool., 1986.— Vol. 35, N 3.— P. 325—333.
- Day W. H. E., Sankoff D. Computational complexity of inferring phylogenies by compatibility // Syst. Zool., 1986.— Vol. 35, N 2.— P. 224—229.
- Estabrook G. F., McMorris F. R., Meacham C. A. Comparison of undirected phylogenetic trees based on subtrees of four evolutionary units // Syst. Zool., 1985.— Vol. 34, N 2.— P. 193—200.
- Estabrook G. F., Strauch J. G., Fiala K. L. An application of compatibility analysis to the blackiths' data on orthopteroid insects // Syst. Zool., 1977.— Vol. 26, N 3.— P. 269—276.
- Farris J. S. A successive approximations approach to character weighting // Syst. Zool., 1969.— Vol. 18, N 4.— P. 374—385.
- Farris J. S. Methods for computing Wagner trees // Syst. Zool., 1970.— Vol. 19, N 1.— P. 83—92.
- Farris J. S. Estimating of phylogenetic trees from distance matrices // Amer. Natur., 1972.— Vol. 106, N 951.— P. 645—668.
- Farris J. S. On the use of the parsimony criterion for inferring evolutionary trees // Syst. Zool., 1973.— Vol. 22, N 2.— P. 250—256.
- Farris J. S. Outgroups and parsimony // Syst. Zool., 1982.— Vol. 31, N 3.— P. 328—334.
- Farris J. S., Kluge A. G., Eckardt M. J. A numerical approach to phylogenetic systematics // Syst. Zool., 1970.— Vol. 19, N 2.— P. 172—191.
- Farris J. S., Kluge A. G., Mickevich M. F. Phylogenetic analysis the monothetic group method, and myobatrachid frogs // Syst. Zool., 1982.— Vol. 31, N 3.— P. 317—327.
- Felsenstein J. Package of programs for inferring phylogenies. Version 1.3. // Tech. Rep., Univ. Wash., Seattle, 1981.— N 23.— 153 p.
- Felsenstein J. Numerical methods for inferring evolutionary trees // Quart. Rev. Biol., 1982.— Vol. 57, N 4.— P. 379—404.
- Fitch W. M., Margoliash E. Construction of phylogenetic trees // Science, 1967.— Vol. 155.— P. 279—284.
- Fitch W. M. On the problem of discovering the most parsimonious tree // Amer. Natur., 1977.— Vol. 111, N 978.— P. 223—257.
- Funk V. A., Wheeler Q. D. Symposium: character weighting, cladistics, and classification // Syst. Zool., 1986.— Vol. 35, N 1.— P. 100—101.
- Hennig W. Grundzüge einer Theorie der phylogenetischen Systematic. Berlin. Deutsch. Zentralverlag., 1950.— 370 S.

- Hennig W.* Phylogenetic systematics.— Urbana. Univ. Illinois Press., 1966.— 263 p.
- Le Quesne W. J.* The uniquely derived character concept and its application to the Delphacidae // *Acta entomol. fenn.*, 1981.— Vol. 38.— P. 41—42.
- Meacham C. A.* A manual method for character compatibility analysis // *Taxon.*, 1981.— Vol. 30, N 3.— P. 591—600.
- Neff N. A.* A rational basis for a priori character weighting // *Syst. Zool.*, 1986.— Vol. 35, N 1.— P. 110—123.
- Queiroz K. de.* The ontogenetic method for determining character polarity and its relevance to phylogenetic systematics // *Syst. Zool.*, 1985.— Vol. 34, N 3.— P. 280—299.
- Shaffer H. B.* Utility of quantitative genetic parameters in character weighting // *Syst. Zool.*, 1986.— Vol. 35, N 1.— P. 124—134.
- Steen W. J., Van der Boontje W.* Phylogenetic versus phenetic taxonomy: a reappraisal // *Syst. Zool.*, 1973.— Vol. 22, N 1.— P. 55—63.
- Wheeler Q. D.* Character weighting and cladistic analysis // *Syst. Zool.*, 1986.— Vol. 35, N 1.— P. 102—109.
- Wiley E. O.* Phylogenetics: the theory and practice of phylogenetic systematics.— N. Y.: John Wiley and Sons, 1981.— 439 p.

Summary

Establishment of the evolution theory allowed to relate classification of organisms to phylogeny introducing the requirement of monophyly of taxa. The first principal proposition concerning morphological certainty and segregation of taxa was formulated as the primary basis of classification. The perfect system assumes the consideration of phylogeny. Principles and methods of phylogeny. were not developed purposefully for a long time and were not formulated clearly. Hennig's theory, or cladism, should be assessed positively as an explicit procedure of clarification of phylogeny and negatively as a cladistic classification alternative to the traditional one.

Hennigian systematics claiming a restricted notion of monophyly excludes existence of daughter taxa of the same and higher rank and reduces classification to a cladogram, i. e. makes it isomorphic to phylogeny. Cladistic classification loses the feature of storage in a condensed form of information on taxa that can be easily retrieved. One should develop and improve Hennigian phylogenetic analysis and abstain from Hennigian principles of classification.

Cladistic phylogenetic analysis is based on construction of morphological series (morphoclines) by separate characters, on determining advanced (apomorphic) and initial (plesiomorphic) state of a character in the series which is called determination of polarity of characters of the morphocline and then on finding apomorphic characters uniting different numbers of the lowest taxa i. e. synapomorphies, that are arranged in the hierarchical order, graphically a cladogram, by means of uniting on the basis of pairs and units of taxa into gradually increasing groups.

Construction of cladogram by synapomorphies encounters certain difficulties: 1. Impossibility to define unequivocally polarity of the negative character. 2. Impossibility to give an a priori isolated assessment of morphocline polarity. 3. Unreliability of determination of variable characters' polarity. 4. Difficulties in identification of false synapomorphies, arising as a result of parallel and convergent evolution. 5. Difficulty of recognition of apomorphies indistinguishable from plesiomorphies arising as a result of instaurations. 6. The difficulty linear arrangement of morphoclines including more than two links. 7. Inapplicability of the criterion of the outer group in taxa isolated by wide hiatuses from the closest ones.

Since the essence of the evolution is origination of the new, divergence and irreversibility, the above difficulties are solved by means of duplicating and weighting i. e. comparing a large number of morphoclines, using of totally positive morphoclines and set of characters. A comparative evaluation (weighting) of characters is of great importance.

The procedure of constructing phylogeny can be divided into several phases:

1. Preliminary sorting of organisms by similarity.
2. Removal of representatives with superficial similarity from the groups.
3. Revealing of as large number of morphoclines as possible.
4. Accompanying clarification of the size of the lower taxa analyzed.
5. Grouping of the lower units by similarity of the state of the character in as large number of morphoclines as possible.
6. Determination of polarity of morphoclines by concentration of plesiomorphies in the few taxa and dispersion of apomorphies in many taxa and differently in each series.
7. Using of the criterion of the outer group.
8. Construction from the top from narrow synapomorphies.
9. Additional search for characters in case of gaps.
10. Reduction of the tree to facilitate weighting of variants of lineage.
11. Repeated additional search for characters to support substantiation.
12. Check of construction of scenario of evolution.

A program for computer is created which allows to examine each variant of the phylogenetic tree possible for the taxa set and to compare them by a number of characters. Criteria of optimum condition that serve for selection of the best variants are based on the concept of the number of duplicates of characters on a particular configuration of the tree (i. e. on the number of repeated arising of apomorphic state of characters). Initial data on taxa are presented as binary characters with polarity of states (i. e. plesiomorphic and apomorphic states) determined by the researcher a priori.

Unlike the traditional approaches rejecting weighting of characters or using one-aspect weighting, a system multi-aspect estimation of characters is developed. It includes for each character: significance of establishment of polarity of conditions, phylogenetic importance of the character, evaluation of the possibility of convergent origination of apomorphies and a particular marking for characters whose apomorphic state is connected with the loss of some structure, reduction or simplification. On the basis of combination of these estimations and the data on the total number of duplicates on a particular configuration of a tree, on the number of interruptions of each character and on every knot of the tree nine indices of the optimum state of configuration analyzed are constructed, from which one or several best trees are chosen as a result of the work of the program.

The program is written in the FORTRAN-IV language and contains 700 operators.

Using it one can analyze groups containing not more than 10 taxa on a computer with middle internal performance