

Указания для решения задач: исправленная версия

1. Методика кладистического анализа

- 1) Кладистический анализ позволяет построить дерево эволюционной истории таксона (так называемую кладограмму), а затем систему таксона.
- 2) Вначале все заданные таксоны (любые таксономические группы одного ранга) выписываются в столбик. Например:

Alphaceae
Betaceae
Gammaceae

- 3) Затем выписываются признаки, которые будут использованы при анализе. Каждому признаку присваивается порядковый номер. Надо постараться, чтобы число признаков хотя бы на один превышало число таксонов (а лучше всего, чтобы признаков было $2N + 1$, где N — число таксонов). Например:

Цвет бутявок	(1)
Размер курсанов	(2)
Рассечение пойлов	(3)
Наличие ропариков	(4)

4) Затем у каждого признака выделяются два состояния: одно — примитивное (плезиоморфное), другое — продвинутое (апоморфное). Первое обозначается «0», второе — «1». Например:

- (1) Бутявки красные - 0; бутявки зеленые - 1
- (2) Курсаны маленькие - 0; курсаны большие - 1
- (3) Пойлы рассеченные - 0; пойлы цельные - 1
- (4) Ропарики есть - 0; ропариков нет - 1

Не стоит выделять более двух состояний на признак. Если же иначе не получается, надо разделить один многозначный признак на несколько двузначных.

5) Нули и единицы присваиваются не просто так, а на основании сведений о:

- а) истории таксонов в палеонтологической летописи (*палеонтологический* метод);
 - б) сведений об индивидуальном развитии тех или иных признаков в процессе жизни особи (*эмбриологический* метод);
 - в) сравнения признаков у взятых для анализа, а также у всех прочих таксонов (*сравнительно-морфологический* метод).
- 6) Если эту так называемую «триаду» применить не удастся, используется метод внешней группы: находится таксон (если только он не задан заранее), который, как нам кажется, ближе всего к предкам данной группы. Все его признаки a priori (заранее) считаются плезиоморфными. Допустим, в данном случае это семейство *Omegaseae*, у которого бутявки красные, курсаны маленькие, пойлы рассеченные и имеются ропарики.

- 7) Теперь строим таблицу, в которой отражены признаки и таксоны (внешняя группа всегда ставится первой!):

	1	2	3	4
Omegaseae	0	0	0	0
Alphaseae	1	0	0	0
Betaseae	1	1	0	0
Gammaseae	0	0	1	1

Важные замечания:

- а) бывает так, что у таксона встречается сразу оба состояния (полиморфизм), тогда принято либо игнорировать редкие проявления, либо, если оба состояния встречаются сравнительно часто — записывать состояние как «1»;
- б) бывает, что признак не применим к данному таксону (например, бессмысленно у бескрылых насекомых искать признаки крыльев), тогда признак обозначается «М» («missing») и используется (обычно) как «0».
- 8) После этого нормальные кладисты передают полученные данные компьютеру для анализа в одной из кладистических программ (например, PAUP, PHYLIP¹, Hennig86, SYNAP); а мы будем строить кладограмму вручную.
- 9) Для этого начинают с внешней группы:

¹Эта программа распространяется свободно. Последнюю версию можно найти в Internet по адресу: <http://evolution.genetics.washington.edu/phylip.html>.

+----- Omegaseae

- 10) Затем к внешней группе присоединяют наиболее примитивный таксон и отмечают филогенетические события, то есть появление, исчезновение (*реверсии*) или параллельное появление (параллелизм, гомоплазия) признаков. Сходство в плезиоморфных признаках (*симплезиоморфия*) в расчет не принимается; наоборот, сходство в апоморфных признаках (*синапоморфия*) считается очень важным. Например:

```

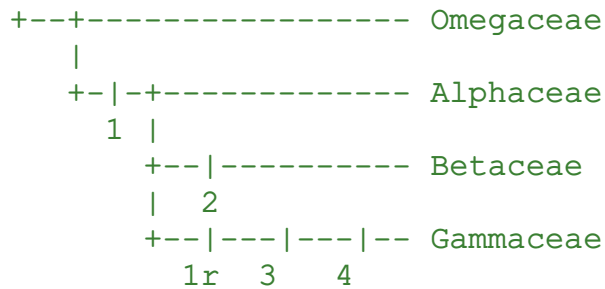
+--+----- Omegaseae
  |
  +----- Alphaseae
  1

```

Появился признак 1 (синапоморфия).

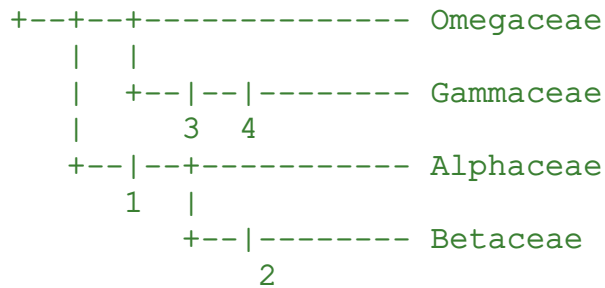
- 11) Затем можно присоединять все более и более продвинутые таксоны. Если есть выбор, то строится не одна, а две или больше кладограмм. Можно поступить и иначе — присоединять таксоны «как попало», а затем анализировать получившиеся кладограммы (см. следующий пункт). Например, мы можем присоединить сначала Betaseae, а потом к ним — Gamtaseae; а можем присоединить Gamtaseae прямо к Omegaseae:

а) Дерево (А)



«1r» означает, что у Gammaseae признак 1 исчез (реверсия). В так называемой *скобочной* записи это дерево выглядит так:
 (Omegaseae,(Alphaseae,(Betaseae,Gammaseae))).

б) Дерево (Б)



В скобочной записи дерево (Б) выглядит так:
 ((Omegaseae,Gammaseae),(Alphaseae,Betaseae)).

12) Теперь посчитаем количество филогенетических событий. В кладограмме (А) 4 синапоморфии + 1 реверсия = 5 событий; а в кладограмме (Б) 4 синапоморфии = 4 события. Теперь применим принцип парсимонии: **из двух кладограмм предпочтительнее более короткая.** В нашем случае это (Б).

13) Важные замечания к последним двум пунктам:

- а) бывают другие алгоритмы построения кладограмм, например, можно присоединять не «гребенкой», а поочередно справа и слева — главное, чтобы получилась наиболее короткая кладограмма;
 - б) бывает, что наиболее короткое дерево выбрать нельзя, тогда все получившиеся используются для построения некоей «средней» кладограммы;
 - в) бывает, что получается всего одна кладограмма.
- 14) Наконец, построим систему наших трех семейств — разделим их на два порядка:

Ordo 1. Gammales

Familia 1. *Gammaceae*

Ordo 2. Alphales

Familia 1. *Alphaceae*

2. *Betaceae*

- 15) Здесь применен **принцип голофилии**: правильно считаются только *голофилетические* таксоны (например, *Alphaceae* + *Betaceae*), то есть такие, которые включают всех потомков своего предка. *Парафилетические* таксоны (например, *Gammaceae* + *Alphaceae*) выделять нельзя; и тем более запрещены *полифилетические* (*Gammaceae* + *Betaceae*).

2. Методика нумерического анализа. Кластерный анализ

Следует иметь в виду, что если кладистический анализ — самостоятельный раздел систематики, то так называемый нумерический анализ представляет собой собрание методов многомерной математической статистики, применяемых к анализу таксономической информации. Результаты нумерического анализа могут использоваться как систематиками-типологами, так и кладистами для последующей обработки и построения систем.

Кластерный анализ — один из методов многомерной статистики, осуществляющий группировку исходных данных. Хотя его результаты могут походить на результаты кладистического анализа, но сами по себе они не дают никакого «руководства к действию». С другой стороны, кластерный анализ позволяет оценить возможные способы группировки таксонов.

Основные этапы кластерного анализа:

- 1) Построение матрицы. Этот этап соответствует пунктам 2–4 и 7 кладистического анализа, за тем исключением, что нули и единицы проставляются именно «просто так», лишь из соображений удобства. Отпадает и необходимость использовать внешнюю группу (хотя чем большее количество близких таксонов участвует в анализе, тем выше достоверность его результатов). Вот подобная матрица (для простоты она такая же, как и в предыдущем разделе):

	1	2	3	4
Alphaceae	1	0	0	0
Betaceae	1	1	0	0
Gammaceae	0	0	1	1

- 2) Дальнейшие вычисления удобно проводить при помощи какого-нибудь распространенного статистического пакета программ (например, SPSS, STATISTICA, R), но, если данных немного, можно делать анализ вручную.
- 3) Теперь надо построить матрицу коэффициентов сходства. Это — таблица, где и столбцы и строки соответствуют таксонам, а ячейки заполняются значениями коэффициента попарного сходства K :

	Alphaceae	Betaceae	Gammaceae
Alphaceae	1		
Betaceae	0.75	1	
Gammaceae	0.25	0	1

Здесь использован простейший коэффициент сходства:

$$K = \frac{\text{число сходных признаков}}{\text{число всех признаков}}$$

Можно использовать и другие коэффициенты сходства (их очень много).

- 4) Теперь нужно построить дендрограмму (кластерное дерево). Для этого вначале соединяем таксоны, имеющие максимальный коэффициент сходства:

```

+----- Alphaceae
| 0.75
---+----- Betaceae

```


- 5) Следующий таксон нужно выбрать (и присоединить) так, чтобы он имел максимальный коэффициент сходства с присоединяемым таксоном. Если однозначно выбрать нельзя, возникает несколько дендрограмм. В нашем случае дендрограмма только одна:



Использованный метод называется методом UPGA (unweighted pair-group average) — метод невзвешенного попарного сравнения. Существуют десятки других методов кластеризации.

- 6) Таким образом, у нас получилось два кластера: (1) общий кластер Alphaceae + Betaceae + Gammaceae с силой связи 0.25 и (2) Alphaceae + Betaceae с силой связи 0.75. Получается следующая система:

Ordo 1. Gammales

Familia 1. *Gammaceae*

Ordo 2. Alphales

Familia 1. *Alphaceae*

2. *Betaceae*

Как видим, она такая же, как и система, полученная кладистическими методами. Это очень частый случай, хотя бывает и наоборот.

3. Методика построения определительных ключей

Вначале следует выписать таксоны и их признаки. Затем подобрать признак, позволяющий разделить все таксоны на две примерно равные группы. Затем найти признак, разделяющий одну из получившихся групп на две примерно равные части, и так далее. Удобнее всего эту процедуру проводить на отдельном листке бумаги, строя дихотомические деревья, где на развилках указаны признаки, а на ветвях — таксоны. Можно по мере того, как ветки будут заканчиваться одним таксоном, постепенно вычеркивать таксоны из первоначального списка. Затем получившуюся схему надо перевести на язык дихотомического ключа.

Почему делить надо на две примерно равные группы? Сравним определительный ключ с атласом, где на каждом развороте находится изображение одного вида растения. Допустим, у Вас в руках некое растение и Вы хотите узнать его название, просто перелистывая страницы атласа. При пролистывании атласа нужный рисунок с равной вероятностью может оказаться на любой странице от 1-ой до 500 000-ой. Положив время на перелистывание одной страницы равным 1 сек и усреднив время, затраченное в каждом из этих случаев (с помощью формулы для суммы арифметической прогрессии), получим величину, примерно равную 250 000 сек, или 69 часов.

Представим себе самый плохой определитель — в котором каждая пара теза-антитеза отделяет всего лишь один вид от остальной совокупности. Движение по такому определителю будет идти так же, как по альбому. Отсюда понятно, как должен быть организован хороший

определитель. В нем каждый пункт должен разбивать «подозреваемые» виды на две по возможности равные группы. Сколько времени потребует работа с таким определителем? После первого этапа в каждой группе окажется по 250 000 видов, второго — по 125 000, третьего — по 62 500 и так далее. Продолжив деления на 2 (или вычислив двоичный логарифм) и положив на чтение текста 1 мин, получим ответ: определение потребует примерно 19 этапов (19 минут). Однако на практике таких этапов будет больше (где-то около 25): ведь невозможно подобрать реальные признаки, которыми бы каждый раз обладала ровно половина «подозреваемых».

Вот другие требования к хорошему определителю:

- 1) Нужно использовать только такие признаки, которые в полевой сезон свойственны подавляющему большинству экземпляров данного вида;
- 2) Если указывается несколько признаков, то на первом месте надо ставить наиболее важные в диагностическом отношении, затем — менее важные и т.д.
- 3) В тезе и антитезе желательно использовать альтернативные описания признаков, не рекомендуются ступени типа «плоды колючие, цветки желтые — плоды не колючие», такие ступени рекомендуется заменить на: «плоды колючие, цветки желтые — плоды не колючие, цветки различной окраски»
- 4) Не рекомендуется использование антитезы типа: «сочетание признаков иное»
- 5) Не следует использовать ступени типа: «листья широкие — листья уже» или «листья жесткие — листья менее жесткие»

- 6) Для мерных признаков следует использовать конкретные числовые значения: не «листья длинные», а «листья 15–20 см дл.»
- 7) Не допускаются «перекрещивающиеся» ссылки, например, когда на «развилке» дихотомического ключа присутствуют ссылки на ступени 3 и 5, а на ступени 3 присутствует ссылка на ступень 9 (возможно, за исключением тех случаев, когда таксон упоминается в ключе более одного раза)
- 8) Разделение по нечетким признакам нужно по возможности продублировать (т. е. один таксон поместить в двух местах ключа) и «подстраховать» (давать в таких случаях не один, а несколько признаков);
- 9) Дополнительно к отличительным признакам следует привести подробное описание (включающее фенологические, экологические и географические признаки), которое позволит судить, правильно ли проведено определение;
- 10) Нужно, чтобы всегда была возможность пройти по ключу «обратно» — туда, где могла быть допущена ошибка. Для этого используются обратные отсылки (например, «11(1)», «7(5)», но не «7(6)»).