

П.Е. БУЛАХ

Национальный ботанический сад им. Н.Н.Гришко НАН Украины  
Украина, 01014 г. Киев, ул. Тимирязевская, 1

## АЛГОРИТМЫ ТЕОРИИ СХОДСТВА В ИНТРОДУКЦИИ И СЕЛЕКЦИИ РАСТЕНИЙ

*Обсуждается необходимость создания теории сходства в биологических науках. Показана перспектива использования её алгоритмов в интродукции и селекции растений. Рассматриваются отдельные составные элементы теории, их преимущества, недостатки и возможность модификации с целью повышения эффективности исследований.*

В биологических дисциплинах решение многих задач связано со сравнением между собой объектов исследования, с выделением в пределах изучаемого множества родственных между собой групп, выявлением в них типичных, ординарных и экстраординарных (исключительных) объектов. Эти операции проводятся преимущественно на основании использования субъективных представлений о сходстве объектов по одному или многим признакам. Отсутствие объективных взглядов на эту проблему в различных областях биологии является существенным тормозом их развития и препятствует эффективному применению математических методов и вычислительной техники.

Проблема сходства объектов в биологии является не только одной из важнейших, но пока ещё и слабоосознанной, прежде всего

из-за своей кажущейся простоты. Известны лишь отдельные работы, где эта проблема затрагивается либо в самом общем виде, либо при решении узких частных задач. Каких-либо концептуальных исследований на эту тему в настоящее время не существует. Возникает потребность в создании теории сходства со специальным методическим аппаратом для решения конкретных практических задач. Вероятно, формализация и совершенствование представлений о сходстве, обобщение и анализ методов классификации объектов в биологических дисциплинах возможны только на основе этой теории. В настоящее время только в геологических науках выкристаллизовались предпосылки для построения теории сходства, определены её назначение, содержание и программа построения [6].

Логично предположить, что теория сходства может базироваться на анализе уже ис-



пользуемых представлений о сходстве в систематике растений [24], животных [20] и микроорганизмов [15]. Именно в данных областях эти представления изложены наиболее чётко. Сформировалось даже самостоятельное научное направление — числовая (нумерическая) таксономия, где используются многочисленные методические варианты решения задач классификации объектов исследования по совокупности диагностических признаков. Это учение основывается на использовании большого арсенала математических средств, обобщённо называемых методами теории распознавания образов.

Практическое воплощение основных положений теории сходства в интродукции растений тесно связано с широким использованием эволюционного арсенала систематики, осознанием её приоритетной роли в развитии науки о переселении растений [9, 21] и реализацией в ботанических садах и дендропарках метода филогенетических комплексов, разработанного Ф.Н. Русановым [16, 17]. Идеи данного метода пронизаны филогенетическим или эволюционным подходом к классификации и систематике растений, в настоящее время он является одним из основных в интродукционной работе и противопоставляется бессистемной хаотической интродукции. Сущность метода состоит в мобилизации как можно большего количества видов определённого рода (или других таксономических единиц) и сравнительного изучения их приспособительных возможностей в новых условиях. По этому поводу Н.В. Цицин [23] отмечает, что только при сравнении родственных форм можно надёжно определить амплитуду и направление изменчивости признаков, в том числе признака экологической приуроченности и пластичности, а это весьма важно для прогнозирования в интродукции.

В силу различных обстоятельств интродукция растений в ботанических садах различных регионов проводится в основном

фрагментами родовых комплексов. При этом группа видов данного рода выращивается и изучается на одном экспериментальном участке, в однородных условиях, а основным методом их исследования является сравнительный анализ.

Именно в результате анализа (с греч. — разложение), одного из средств человеческого познания, устанавливается степень целостности изучаемого комплекса, т.е. ставится задача определения степени сходства его компонентов. Отсутствие стройной теории сходства не позволяет эффективно решить подобную задачу, что значительно снижает результативность интродукционной работы.

К изучению фрагмента родового комплекса можно подойти и с позиций другого метода познания действительности — синтеза (с греч. — соединение, составление), являющегося органическим дополнением анализа. Только в единстве эти методы дают наилучший результат.

Перефразируя известное выражение Ф. Энгельса можно сказать, что изучение родового комплекса (или его фрагмента) стоит столько же в разложении его на составляющие элементы, сколько в объединении связанных друг с другом элементов в единство. Без анализа нет и синтеза.

Используя богатый методический арсенал средств познания структуры рода или его фрагмента целесообразно в условиях культуры выделить в его пределах родственные группы близких видов на основании изучения морфологических признаков, особенностей жизненного цикла и их экологической приуроченности. Такие группы видов, относящиеся к одному роду и характеризующиеся сходными приспособительными особенностями к определённому типу местообитаний, мы называем биоморфологическими типами и рассматриваем в качестве самостоятельной единицы в интродукционном исследовании. Подобный под-



ход позволяет глубже изучить механизм приспособления каждого биоморфотипа к изменившимся условиям. Один из видов данного типа может характеризовать поведение в культуре остальных видов этого же типа [4].

К отдельным составным элементам теории сходства с полным правом можно отнести методы многомерной статистики (кластерный анализ и метод главных компонент), метод корреляционных плед П.В. Терентьева [22] и таксономический анализ Е.С. Смирнова [19]. Каждый из этих методических подходов имеет своих сторонников и можно привести многочисленные примеры их использования в ботанике.

Апробация их с целью классификации объектов исследования по комплексу диагностических признаков осуществлена нами на тестовом примере (интродуцированные виды рода *Allium* L., хорошо изученные в систематическом отношении: *A. altissimum* Regel, *A. nutans* L., *A. ramosum* L., *A. caesium* Schrenk, *A. christophii* Trautv.).

Сравнительный анализ результатов использования каждого метода позволил оценить их особенности, преимущества и недостатки. Алгоритмы таксономического и кластерного анализов существенно модифицированы. Признано целесообразным:

1) широкое использование в интродукционной работе идей строгого анализа систематических категорий (таксономический анализ), предложенных Е.С. Смирновым [18, 19] с учётом наших дополнений и изменений к ним в связи с автоматизацией вычислений [2–4];

2) заимствование правил кластерного анализа [11] с соответствующими поправками к нему, позволяющими избежать обесценивания признаков объектов классификации.

Предлагаем наши дополнения к алгоритмам каждого из методов с указанием их преимуществ.

*Таксономический анализ* в его первоначальном виде используется лишь при условии, когда в исследование включено небольшое количество объектов, сравниваемых по ограниченному набору признаков. В противном случае достижение цели представляет собой весьма трудоёмкий процесс. Использование вычислительной техники помогает значительно повысить производительность метода и избежать возможных ошибок при расчётах по алгоритмам Е.С. Смирнова. Однако эти алгоритмы не приспособлены для автоматизации вычислений. Удобной формой для программирования является запись таксономического анализа в терминах булевых функций. Новый алгоритм позволяет легко составить программу вычислений матрицы таксономических отношений (программа TAXON) с последующим автоматическим анализом её (программа BIOTIP). Первая программа направлена на количественное определение степени связи между сравниваемыми по заданным признакам объектами. Вторая — осуществляет автоматизированный анализ полученных результатов.

Работа сводится к вычислительному эксперименту, в результате которого выбирается оптимальное значение коэффициента связи, при котором объекты отчётливо классифицируются на группы родства. Степень расчленения определённой совокупности объектов на родственные группы можно варьировать, что определяется конкретной задачей исследования.

Программы написаны на языке БЕЙСИК и имеют необходимые сервисные функции: проверка результатов расчёта, аварийная остановка и преобразование матрицы таксономических отношений в матрицу "расстояний", что необходимо для графической интерпретации результатов расчёта.

*Кластерный анализ* по своей природе отражает адансоновский подход к выбору

признаков, на основании которых производится сравнение объектов исследования, т.е. ценность всех признаков признается равной. Однако существуют признаки, использование которых сразу приводит к естественной классификации исследуемой совокупности единиц, а также такие, которые не несут существенной информации при классификации объектов исследования. В связи с этим сформировался неадансоновский подход к классификации, базирующийся на признании неодинаковой информативной ценности признаков. Он объективно отражает меру сходства объектов исследования и представляется более удачным. Поэтому мы предлагаем алгоритм пошаговой иерархической процедуры, позволяющей устранить основной недостаток кластеризации (признание равноценности признаков). В самом общем виде он может быть представлен следующим образом. По каждому из рассматриваемых признаков проводится одновременная кластеризация. На последних ее этапах вычисляются значения функционала и в качестве "веса" признака используются его величины. Большие значения функционала должны соответствовать информационно-ценным признакам.

Этот алгоритм нуждается в соответствующей апробации и может служить основой компьютерной программы классификации объектов.

Для осуществления любых классификационных построений на научной основе необходимо сформулировать, что именно подразумевается под классификацией и классом. Общепринятых определений в настоящее время не существует. Нет и достаточно общих конструктивных идей, которые могли бы лечь в основу будущей теории сходства.

Рассмотренные выше методы числовой (нумерической) таксономии имеют один общий и существенный недостаток. Он заключается в том, что приоритет в этой методологии принадлежит математическим сред-

вам классификации, а не тем признакам, по которым объекты классификации объединяются в родственные группы. При таком подходе выделенные группы оказываются вторичным результатом использования того или иного математического метода группирования объектов. Вероятно, в любом классификационном учении первичным должен быть класс, а не математические средства его формирования.

Это положение в теории систематики уже осознанно используется в связи с разработкой понятия "биологическое расстояние" [8], оно должно быть приоритетным и в интродукционной работе в связи с использованием метода филогенетических комплексов и решением задач классификации объектов по совокупности признаков.

Практическое использование элементов теории сходства в интродукции растений не ограничивается решением задач классификации и выделением родственных групп в пределах филогенетических комплексов. Например, одно из основных понятий теории — мера сходства может быть использовано при прогнозировании адаптационной способности растений. Проведенный нами анализ количественных методов сравнительной флористики и многочисленных показателей сходства флор в контексте проблемы интродукционного районирования позволил оценить преимущество показателя "мера включения" по сравнению с другими традиционно используемыми критериями [1].

Отдельные алгоритмы теории сходства в последние годы находят свое применение и в селекционной работе, которую можно рассматривать как логическое завершение интродукционного процесса или его заключительный этап. В связи с этим Г.И. Родионенко [14, с. 245] образно утверждает, что "интродукция растений, не подкрепляемая селекцией, напоминает одностороннюю инвалида. Там, где с годами накапливается огром-



ный генофонд, представленный растениями дикой флоры, а этим и характерны ботанические сады, — там обязательно сам интродуктор или в контакте с селекционером, должен заниматься оригинаторской деятельностью. К сожалению, эта традиция после распада СССР в России и странах СНГ пошла на резкое снижение".

В селекции растений обычно выделяют три основных этапа: формулирование цели селекции и моделирование будущего сорта; подбор родительских пар для гибридизации; отбор перспективных гибридов и их испытание [5]. Каждый из этих этапов имеет свои особенности и соответствующее методическое обеспечение.

Для эффективного управления селекционным процессом наряду с традиционными методами селекции возникает необходимость широкого использования математических методов, в частности, некоторых алгоритмов теории сходства и вычислительной техники. Это не дань моде, а необходимость, вызванная изучением сложной системы "организм — среда", являющейся объектом исследования селекционера. Подобного рода системы можно познать только с позиций системного анализа, используя присущие ему методические приемы исследования.

Рассмотрим возможность применения некоторых элементов теории сходства на отдельных этапах селекционного процесса.

*Формулирование цели селекции и моделирование будущего сорта.* Любой селекционный процесс начинается с разработки требований к сорту. Эти требования должны быть реальными и отражать крайние пределы признаков сорта. Использование на этом этапе методов многомерного статистического анализа позволяет выделить устойчивые группы (кластеры) тесно взаимосвязанных признаков, а также разбить на кластеры набор исходных форм. У сортов, входящих в один кластер, должно наблюдаться сходное проявление некоторых комплексов признаков.

*Подбор пар для гибридизации.* Подбор родительских форм, способных дать требуемое потомство — следующий этап селекционного процесса. Его осуществление связано с умением прогнозировать результаты скрещиваний. Один из подходов к этой проблеме основан на оценке степени сходства или различия между формами. Существует большое разнообразие методов этой оценки: евклидово расстояние, мера Махаланобиса, регрессионный анализ, вероятностный способ и др. [5]. С этой целью можно использовать основные идеи таксономического анализа, разработанного Е.С. Смирновым [19], и описанную выше нашу автоматизированную модификацию его с программным обеспечением.

*Отбор перспективных гибридов и их испытание.* В сортоиспытании широко известны математические методы В.Е. Перфильева [13] и Э.Х. Гинзбурга [7], позволяющие провести сравнение новых сортов с контрольными по урожайности и моделировать возможный ареал возделывания испытуемого сорта. Для оценки сортов растений интерес представляет и применение теории распознавания образов, релаксационный алгоритм которой разработал В.Г. Нестеров [12].

В рассмотренном перечне математических технологий сравнения объектов исследования в одних случаях уже имеется программное обеспечение, в других — только алгоритмы, на основе которых возможна автоматизация исследований. Перечисленные методы теории сходства дают некоторое представление о возможностях интенсификации селекционной работы.

Существующие положения о сходстве в биологических науках сформировались, в основном, в результате восприятия действительности посредством построения искусственных или естественных классификаций и измерений (последние можно рассматривать как специфический частный случай процессов классификации). Вероятно, мето-

ды классификации должны играть ключевую роль в построении теоретических представлений о сходстве объектов исследования.

Можно выделить три различных типа задач классификации, нашедших свое место в различных областях науки и получивших определения: "классификация", "дискриминация" и "ранжирование" [10]. Каждый из них характеризуется своим специфическим методическим подходом, но всех их, вероятно, объединяет понятие расстояния между объектами исследования. В нашем случае речь идет об уже устоявшемся понятии "биологическое расстояние". К сожалению, в биологии термин "расстояние" определяется достаточно вольно, интуитивно. Например, расстояние между двумя популяциями представляет собой число, которое отражает степень их сходства или различия. Близкие между собой популяции характеризуются малым значением этого расстояния и наоборот. Это же утверждение справедливо и для видового или фитоценотического уровня организации организмов. Например, при выделении биоморфологических типов интродуцированных видов рода *Allium* L. мы, пользуясь модифицированным вариантом метода Е.С. Смирнова, получили количественные характеристики таксономических отношений видов друг с другом. Это позволяет судить о наличии родственных групп в пределах изучаемой выборки (положительные значения таксономических отношений свидетельствуют о сходстве сравниваемых видов, отрицательные — о различии, абсолютная величина — о степени сходства или различия). Биологическое расстояние между этими видами получено в виде матрицы расстояний, представляющей собой совокупность обратных положительных значений таксономических отношений. Она также характеризует степень родства отдельных видов, но при этом легко осуществляется графическая интерпретация таксономического

анализа. Чем меньше число, связывающее сравниваемые виды в матрице расстояний, тем ближе их родство и положение на графике [2, 4]. Это один из возможных способов измерения биологического расстояния, в целом же следует признать, что этому понятию в биологии недостает содержательного смысла в рамках какой-либо строгой логической системы.

Такой системой, вероятно, могла бы стать теория сходства, но в настоящее время она только формируется. Это обстоятельство сдерживает формализацию понятия "биологическое расстояние" и, соответственно, развитие способов его измерения. Вместе с тем существует достаточное количество методов измерения биологического расстояния и коэффициентов, его характеризующих. Однако совершенно неясно, в каких случаях пользоваться одними формулами, а в каких — другими. В связи с этим С. Смит [25] отмечает, что в этом вопросе наблюдается значительный произвол, при котором зачастую теряется всякая биологическая реальность.

Это обстоятельство лишний раз свидетельствует о необходимости создания теории сходства, в рамках которой возможны формализация и измерение биологического расстояния.

В некоторой степени представления о теории сходства могут основываться на прочно утвердившихся положениях теории вероятности, которая уже получила достаточное развитие в биологии. Эти теории по своей сути взаимосвязаны и в идеале должны дополнять друг друга. Однако степень их развития остается разной. Если теория вероятности имеет строгий математический вид, то теория сходства только находится на пути своей формализации. Ее становление возможно на основе анализа существующих алгоритмов сходства и поиска новых концептуальных положений, позволяющих эффективно решать задачи классификации объектов в различных облас-



тях биологии. Надеемся, что создание методического аппарата сравнения в интродукции и селекции растений поможет корректно ставить и решать задачи выделения, разделения, упорядочивания, прогнозирования, районирования и экономической оценки объектов разного уровня организации.

1. Булах П.Е. Использование методов сравнительной флористики в интродукционном прогнозе // Интродукция и акклиматизация растений. — 1994а. — Вып. 21. — С. 13–15.
2. Булах П.Е. Луки природной флоры Средней Азии и их культура в Украине. — К.: Наук. думка, 1994б. — 124 с.
3. Булах П.Е. Биоморфологический тип как самостоятельная интродукционная единица // Проблемы ботаники на рубеже XX–XXI веков: Тез. докл. II (X) съезда Русского ботанического общества. — Санкт-Петербург, 1998. — С. 276.
4. Булах П.Е. Биоморфологический тип как интродукционная единица и особенности его выделения // Доп. НАН України. — 2000. — № 3. — С. 171–173.
5. Бутенко А.И. Использование математических методов и ЭВМ в селекции плодовых и ягодных растений // Математическое моделирование в садоводстве. — Мичуринск, 1990. — С. 3–7.
6. Воронин Ю.А. Начала теории сходства. — Новосибирск: Наука, 1991. — 128 с.
7. Гинзбург Э.Х. Описание наследования количественных признаков. — Новосибирск: Наука, 1984. — 250 с.
8. Котов В.Н. Применение теории измерений в биологических исследованиях. — К.: Наук. думка, 1985. — 100 с.
9. Малеев В.П. Теоретические основы акклиматизации. — Л.: Сельхозгиз, 1933. — 160 с.
10. Малиновский Л.Г. Процессы классификации — основа построения наук о действительности // Алгоритмы обработки экспериментальных данных. — М.: Наука, 1986. — С. 155–181.
11. Мандель И.Д. Кластерный анализ. — М.: Финансы и статистика, 1988. — 273 с.

12. Нестеров В.Г. Применение теории распознавания образов для оценки сортов растений и пород животных по принципу биоэкоза // Доклады ВАСХНИЛ. — 1971. — № 5. — С. 38–41.

13. Перфильев В.Е. О методике сортоиспытания плодовых культур // Плодоовощное хозяйство. — 1985. — № 4. — С. 28–32.

14. Родионенко Г.И. Некоторые частные и общие закономерности, свойственные интродукции растений // Биологическое разнообразие. Интродукция растений: Материалы Второй Междунар. конф. (20–23 апреля 1999 г., Санкт-Петербург). — Санкт-Петербург, 1999. — С. 245–248.

15. Романовская В.А., Столяр С.М., Малащенко Ю.Р. Систематика метилотрофных бактерий. — К.: Наук. думка, 1991. — 212 с.

16. Русанов Ф.Н. Новые методы интродукции растений // Бюл. Гл. ботан. сада АН СССР. — 1950. — Вып. 7. — С. 27–37.

17. Русанов Ф.Н. Метод родовых комплексов в интродукции растений и его дальнейшее развитие // Там же. — 1971. — Вып. 81. — С. 15–20.

18. Смирнов Е.С. Таксономический анализ рода // Журн. общ. биол. — 1960. — 21, № 2. — С. 89–103.

19. Смирнов Е.С. Таксономический анализ. — М.: Изд-во Моск. гос. ун-та, 1969. — 188 с.

20. Смирнов Е.С., Федосеева Л.И. Таксономическая структура рода *Meromyza* Meig. // Журн. общ. биол. — 1967. — 28, № 5. — С. 604–611.

21. Собко В.Г., Гапоненко М.Б. Интродукція рідкісних і зникаючих рослин флори України. — К.: Наук. думка, 1996. — 283 с.

22. Терентьев П.В. Метод корреляционных плеяд // Вестн. ЛГУ. Сер. Биол. — 1959. — № 9, вып. 2. — С. 137–141.

23. Цицин Н.В. О развитии поиска, испытаний и введение в культуру растений природной флоры // Бюл. Гл. ботан. сада АН СССР. — 1972. — Вып. 83. — С. 3–9.

24. Шмидт В.М. О методе таксономического анализа Е.С. Смирнова и некоторых возможностях его применения в ботанике // Ботан. журн. — 1962. — 47, № 11. — С. 1648–1654.

25. Smith C.A.B. Analyses of ecosystems // Ann. Hum. Genet. — 1972. — 36, N 2. — P. 241–245.



АЛГОРИТМИ ТЕОРІЇ СХОЖОСТІ  
В ІНТРОДУКЦІЇ ТА СЕЛЕКЦІЇ РОСЛИН

*П.Е. Булах*

Національний ботанічний сад ім. М.М.Гришка  
НАН України, Україна, м. Київ

Обговорюється необхідність побудови основ теорії схожості у біологічних науках. Показано перспективу використання її алгоритмів в інтродукції та селекції рослин. Розглянуто окремі складові елементи теорії, їхні переваги, недоліки та можливість модифікації з метою підвищення ефективності досліджень.

ALGORITHM OF SIMILARITY THEORY  
IN PLANT INTRODUCTION AND BREEDING

*P.E. Bulakh*

M.M. Grishko National Botanical Gardens,  
National Academy of Sciences of Ukraine, Ukraine,  
Kyiv

Necessity to form the backgrounds for similarity theory in biological sciences is discussed. A perspective of its algorithm usage in plant introduction and breeding is shown. Some elements of this theory, their advantage, shortage and ability of modification for increasing an effect of investigation are examined.