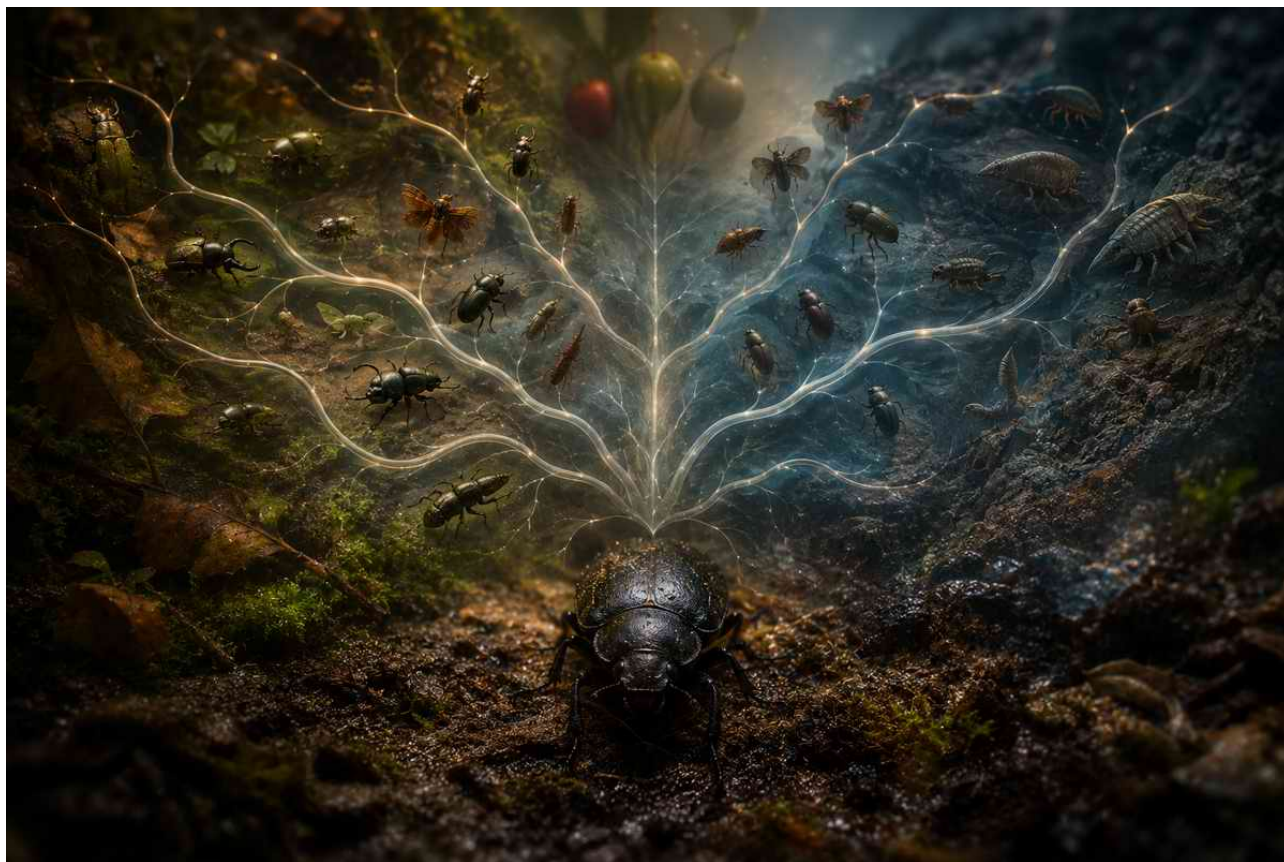


Тысячи исследований эволюции могли получать убедительные, но ложные ответы



Дата публикации: 03.09.2026

Почему одна группа животных порождает тысячи видов, а другая за миллионы лет остается сравнительно небольшой? Эволюционные биологи пытаются отвечать на такие вопросы с помощью математических моделей: сопоставляют родословные организмов, появление признаков, скорость образования новых видов и вымирания. На подобных расчетах основаны более тысячи научных работ. Но теперь выясняется, что внутри этого инструментария скрывается фундаментальная ловушка: совершенно разные сценарии эволюции иногда математически неотличимы друг от друга.

Проблема серьезнее обычной статистической погрешности. Представьте детектива, которому предъявили несколько совершенно разных историй преступления, но каждая из них оставляет абсолютно одинаковый набор улик. Даже идеальный анализ улик не позволит определить, какая история произошла на самом деле. Нечто похожее может происходить при реконструкции эволюции.

Исследование Сергея Тарасова из Финского музея естественной истории и

Йозефа Уеды из Виргинского политехнического университета опубликовано в Nature Communications. Любопытно, что путь к сложной математической проблеме начался вовсе не с уравнений, а с анатомии жуков и вопроса, который легко объяснить на яблоках.

Представим красное, светло-зеленое и темно-зеленое яблоки. Можно считать, что существует три состояния цвета. Но можно объединить два оттенка зеленого и оставить только два состояния – красное и зеленое. Когда такое упрощение допустимо? И главное, сохранит ли модель после объединения прежнее поведение?

Тарасов столкнулся с аналогичной задачей при классификации анатомических признаков жуков. Исследователи обратились к математической концепции, известной как lumpability – условно ее можно назвать допустимым объединением состояний. Она появилась еще в 1960-х годах и описывает условия, при которых несколько состояний марковского процесса можно объединить так, чтобы существенная динамика системы не изменилась.

Марковские модели используются далеко за пределами эволюционной биологии. В простейшем случае они описывают систему, переходящую между различными состояниями с определенными вероятностями. Такие модели применяют в генетике, эпидемиологии, физике, экономике, информатике и анализе поведения. Их фундаментальное свойство состоит в том, что вероятность следующего состояния определяется текущим состоянием системы.

Работая с объединением состояний, Тарасов и Уеда обнаружили более общий математический принцип. Они показали, что любую дискретную марковскую модель можно представить в другой форме – как эквивалентную конструкцию со скрытыми состояниями. Авторы назвали этот подход скрытым расширением. Новая модель выглядит крупнее исходной, зато состоит из более простых повторяющихся элементов.

Именно такое преобразование проявило симметрии, которые раньше оставались спрятанными внутри сложных уравнений. Выяснилось, что несколько существенно различных моделей способны создавать одну и ту же наблюдаемую картину. Для исследователя, располагающего только конечными данными, эти сценарии могут выглядеть одинаково правдоподобными.

В эволюционной биологии последствия особенно интересны. Ученые часто хотят понять, способствует ли определенный признак появлению новых видов. Например, увеличивает ли яркая окраска скорость диверсификации птиц, влияет ли способ питания на разнообразие рыб или способствует ли появление нового оружия у самцов разделению эволюционных линий. Для проверки таких гипотез

признаки накладывают на филогенетическое дерево и сравнивают математические модели.

Но если модели обладают скрытой симметрией, возникает опасность красивого ложного объяснения. Исследователь может получить статистически убедительный результат, согласно которому определенный признак изменил скорость образования видов, хотя другая эволюционная история без такой причинной связи способна породить те же наблюдаемые данные.

Авторы продемонстрировали проблему на опубликованных данных о палочниках. В этих насекомых у самцов некоторых видов имеются выступы на ногах, используемые как оружие в борьбе за самок. Исследователей интересовало, связано ли появление такого оружия с темпами диверсификации.

Первоначальная работа не обнаружила убедительных доказательств такой связи, и новый анализ этого вывода не опроверг. Но когда ученые расширили набор рассматриваемых моделей, произошло нечто показательное: стандартные статистические методы начали предпочитать сценарии, в которых мужское оружие все-таки влияло на диверсификацию.

Получается неприятный парадокс. Данные не изменились, организмы не изменились, их эволюционное дерево осталось прежним – изменился лишь набор допустимых математических объяснений. И этого оказалось достаточно, чтобы статистически предпочтительной стала совсем другая биологическая история.

Это не означает, что тысяча предыдущих исследований автоматически неверна. Новая работа не проверяет заново каждый опубликованный результат и не дает оснований отбросить все модели диверсификации. Она показывает более фундаментальное ограничение: в некоторых случаях имеющихся данных недостаточно, чтобы однозначно различить несколько скрытых эволюционных механизмов.

Такую проблему математики называют неидентифицируемостью. Если две различные модели принципиально дают одинаковое распределение наблюдаемых результатов, увеличение вычислительной мощности само по себе не спасет ситуацию. Можно запустить более быстрый компьютер и собрать больше наблюдений того же типа, но машина будет лишь точнее вычислять ответ на вопрос, который не имеет единственного решения.

Практический выход может заключаться в поиске дополнительных типов данных, введении обоснованных ограничений и предварительной проверке моделей на идентифицируемость. Главное достижение новой работы именно в том, что скрытая математическая структура стала видимой. Теперь исследователь может спрашивать не только какая модель лучше соответствует

данным, но и способны ли данные вообще отличить эту модель от ее математических двойников.

Особенно красиво происхождение самого открытия. Биолог пытался понять, как правильно классифицировать признаки жуков, столкнулся с проблемой группировки состояний, обратился к теории марковских процессов и в результате обнаружил математическую структуру, которая вернулась обратно в биологию и объяснила давнюю загадку эволюционных моделей.

Наука нередко представляется дорогой от фундаментальной математики к ее практическим приложениям. Здесь движение произошло в обе стороны. Жук поставил биологический вопрос, вопрос привел к новой математике, а математика обнаружила слабое место методов, которыми биологи пытаются восстановить историю жизни. Иногда для понимания миллионов лет эволюции действительно достаточно начать с вопроса о том, считать ли два зеленых яблока одинаковыми.

Ссылка: «Неидентифицируемость и ложноположительные выводы в моделях диверсификации, зависящих от состояния» DOI: [10.1038/s41467-026-75829-5](https://doi.org/10.1038/s41467-026-75829-5).