

ДНК-анализ в генеалогических исследованиях



Павел Морозик,
директор Института
генетики и цитологии
НАН Беларуси, кандидат
биологических наук,
доцент



Олег Левданский,
заведующий сектором
биоинформатики Института
генетики и цитологии
НАН Беларуси, кандидат
биологических наук

Развитие генетических технологий произвело революцию во многих сферах деятельности человека, включая медицину, фармакологию, сельское хозяйство, криминалистику, спорт и, разумеется, генеалогию.

Генетическая генеалогия – новое направление, возникшее на стыке популяционной генетики, истории и этнографии, позволяющее делать предположения о степени родства и происхождении человека. Мировая и отечественная практика применения современных методов ДНК-анализа в популяционных исследованиях получила широкое распространение в XXI в., предпосылкой для чего стало не только развитие и стремительное удешевление себестоимости молекулярно-генетических тестов, но и успешная реализация проекта «Геном человека», создание баз данных о генетической структуре популяций и появление эффективных биоинформатических алгоритмов индексации и обработки огромных массивов данных. Во многом внимание к ДНК-тестам связано с ростом интереса людей

к своим корням (происхождению), а также возможностью найти «генетических» родственников.

Использование метода составления родословных начали применять еще в XX в. для медико-генетического консультирования при определении типа наследования признака (болезни) в семье с указанием родственных связей. Этот подход основывался на клинических обследованиях: сборе сведений о близких больного, составлении и анализе родословной. Исследования в области ДНК-генеалогии стали возможны только после появления способов генотипирования.

Впервые метод выявления индивидуализирующих особенностей ДНК (ДНК-профилирование) для целей криминалистики был описан в 1985 г. английским биологом Алеком Джеффрисом и в этом же году успешно применен в Великобритании для разрешения иммиграционного спора. В 1986 г. ДНК-идентификацию впервые использовали в уголовном деле об убийстве двух молодых женщин в Лестершире (Великобритания): оправдали невиновного и осудили преступника [1]. С тех пор за относительно короткий срок ДНК-анализ получил широкое распространение по всему миру и теперь играет важную роль не только для ДНК-идентификации, но и при установлении биологического родства, что совпадает с задачами генетической генеалогии. В дальнейшем методы полимеразной цепной реакции (ПЦР) и секвенирования существенно расширили возможности подхода, включая анализ фрагментов генома, представляющих интерес в первую очередь для ДНК-генеалогии – однородительских маркеров в Y-хромосоме и митохондриальной ДНК (мтДНК). Один из наиболее впечатляющих примеров – идентификация останков членов царской семьи Романовых [2].

Геномное профилирование в ДНК-генеалогии успешно применяется для решения целого ряда фундаментальных и прикладных задач, включая изучение демографической истории отдельных популяций человека, установление отсутствующих в родословных данных, оценку степени родства, предположение о геногеографическом и этническом происхождении предков и др. [3]. Методы ДНК-профилирования позволяют проанализировать любой участок ядерной или митохондриальной ДНК человека и имеют множество преимуществ, включая высокую достоверность, чувствительность (анализ минимальных количеств биологических образцов), возможность как исключить, так и подтвердить идентичность образцов. Процедура состоит из нескольких последовательных этапов, включающих сбор генетического материала, выделение ДНК, проведение молекулярно-генетического исследования, анализ и интерпретацию результатов, подготовку заключения.

В криминалистике золотым стандартом ДНК-идентификации являются короткие tandemные повторы (микросателлиты, STR-маркеры), которые расположены в некодирующей последовательности генома, являющейся высокополиморфной. STR-маркеры распространены по всему геному, при этом длина локуса составляет от 2 до 7 оснований, а число повторов – от 3 до 20. На разных хромосомах находятся тысячи STR-маркеров, которые потенциально могут быть идентифицированы.

Однонуклеотидные полиморфные локусы (SNP-маркеры) – переменные участки, в которых встречается как минимум 2 разных нуклеотида (максимум 4 – А, Т, Г, Ц). Они удобны для использования в ДНК-генеалогии и становятся все более востребованными, так как имеют ряд ключевых преимуществ перед STR-маркерами: низкую частоту мутаций, высокую распространенность в геноме, короткую длину ампликонов (важно при анализе деградированной ДНК), универсальность (подходят как для целей идентификации, так и для определения родства и отцовства, этнического и биогеографического происхождения). Стоит отметить, что биаллельные SNP менее информативны по сравнению с мультяллельными STR-маркерами, и для достижения высокой достоверности необходимо включать не менее 40–60 таких маркеров в профиль [4]. В то же время три- и тетра-аллельные SNP могут значительно повысить разрешающую способность метода.

При изучении истории миграции и расселения человека по миру с помощью молекулярно-

генетических методов пользуются понятиями гаплогрупп (комбинации уникальных наборов генетических маркеров (гаплотипов), произошедших от общего предка). Составлены гаплогруппы, включающие маркеры Y-хромосомы и мтДНК, являющиеся универсальными для различных этнических групп и популяций [5], по которым составляют индивидуальный ДНК-профиль. Все члены одной гаплогруппы – потомки одного мужчины или женщины из очень далекого прошлого. Применение аутосомных маркеров в ДНК-генеалогии ограничено, поскольку они позволяют с необходимой точностью получить сведения только о близких родственниках (не более 6–7 поколений), так как для них характерны более высокое генетическое разнообразие и степень рекомбинации.

Передается только по отцовской линии, и, следовательно, позволяет исследовать мужскую родословную индивида Y-хромосома. На ней локализовано большое количество маркеров, для которых проведен филогенетический анализ и известно распределение частот в различных популяциях [6]. Генотипирование по 20–30 таким маркерам позволяет с высокой достоверностью определить вероятное геногеографическое происхождение индивида с географической привязкой по всему миру. Из-за отсутствия рекомбинации на Y-хромосоме по ней можно проследить происхождение на сотни и тысячи поколений в прошлое.

Для анализа истории популяций по женской линии аналогичным образом исследуется митохондриальная ДНК, поскольку ее все дети, независимо от пола, получают от матери. Благодаря тому, что между молекулами мтДНК не осуществляется рекомбинация, анализ значительно упрощается. В филогенетических исследованиях показано, что некоторые гаплогруппы мтДНК имеют ограниченное распространение на различных континентах, и для выявления различий между географическими регионами в панель потребуется включить 20–30 SNP-маркеров [7].

Основная концепция генетической генеалогии основана на том, что у любых двух индивидов был общий предок, и они содержат определенный процент одинаковых фрагментов ДНК. При этом чем больше таких участков, тем ближе родство (то есть меньше поколений до ближайшего родственника). Принцип ДНК-генеалогии в некоторой степени основан на вероятностной оценке. Случайное совпадение большого количества ДНК-маркеров гаплогрупп у разных лиц между собой крайне маловероятно.

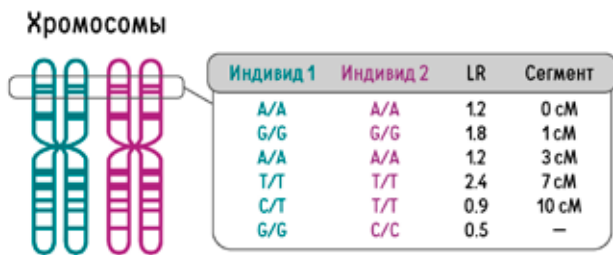


Рис. 1. Схематическое изображение подходов к определению степени родства между двумя индивидами. По результатам тестирования сравниваются генотипы по исследуемым маркерам, рассчитывается коэффициент родства по методу оценки правдоподобия (LR) и размеры идентичных между индивидами сегментов (в сантиморганах, сМ)

Создание баз данных, включающих частоты встречаемости в популяциях аллелей различных локусов, гаплотипов и генотипов, а также выявление генетических различий между этими популяциями, стало возможным благодаря их общему происхождению, скрещиванию (вступлению в брак предпочтительно с представителями той же популяции) и общностью территории проживания. В противном случае было бы невозможно с помощью генетических методов определить принадлежность человека к тому или иному этносу. Обычно при этом разрабатывают строгие критерии включения: только лица, у которых все родственники на протяжении 3 поколений родились в данной популяции и относились к данной этнической группе [8]. В больших городах с непрерывными миграционными потоками проведение подобных изысканий невозможно.

По результатам таких работ устанавливаются не только степень родства с другим индивидом (совпадения по маркерам), но и к какой гаплогруппе принадлежит исследуемая ДНК. Все люди происходят по прямой женской линии от Митохондриальной Евы – древней африканской женщины, которая

жила, вероятно, около 150 тыс. лет назад. Разные ветви ее потомков могут нести одну из трех основных гаплогрупп мтДНК – L, M и N. Первая и представляет собой Митохондриальную Еву и разделена на 7 подгрупп: L0, L1, L2, L3, L4, L5, L6. Гаплогруппа L3 и возникшие от нее гаплогруппы M и N появились в момент миграции человека за пределы Африки. Все остальные гаплогруппы мтДНК, встречающиеся за пределами континента (A, B, C и др.), произошли от гаплогрупп M либо N.

Исследование же Y-хромосомы кроме родства позволяет идентифицировать гаплогруппы, с помощью которых можно определить происхождение от одного человека, которого назвали Y-хромосомным Адамом (гаплогруппа A), который жил, вероятно, около 200–300 тыс. лет назад. Затем гаплогруппа A разделяется на 2 основные гаплогруппы: B и CT. От последней произошли остальные африканские и неафриканские гаплогруппы Y-ДНК (DE, F и т.д.) [9]. В некоторых случаях ДНК-тест Y-хромосомы может даже помочь установить фамилию индивида, поскольку она наследуется по тому же принципу [10].

Основной этап генеалогического ДНК-анализа – сопоставление результатов исследования с другими лицами (рис. 1) или базами данных, филогенетический анализ. Методологический арсенал молекулярной генеалогии использует способы оценки генетического расстояния между полиморфными маркерами, времени до ближайшего общего предка (TMRCA). Генеалогическое изыскание может включать в себя построение генеалогического дерева, поиск вероятных родственников. Предположения о родстве можно сделать, если уровень совпадения генетических маркеров превышает определенный порог. Нередки случаи, когда дальние, а порой и близкие родственники находят друг друга в базах данных компаний, проводящих тестирование. Хотя результаты таких тестов необходимо под-

тверждать, поскольку даже родные братья и сестры теоретически могут иметь 100% одинаковых генов, а могут не иметь их вовсе, если получают от родителей разные комбинации хромосом.

Аналогичным образом устанавливается принадлежность к той или иной популяции, если выявлены характерные для нее маркеры. Стоит отметить, что обычно разработкой алгоритмов и указанием пороговых значений степени родства занимаются раз-



Рис. 2. Программно-информационный комплекс референсных баз данных, реализующий технологию ДНК-идентификации и выявления популяционной принадлежности неизвестного индивида по характеристике его ДНК

ные исследователи или компании самостоятельно, результаты сильно зависят от объема баз данных, в связи с чем могут возникать несоответствия для одного и того же образца по этногеографическому/популяционному происхождению.

Использование генеалогических баз данных ДНК может оказаться полезным для сопоставления ДНК-профилей разных людей. Именно так в 2018 г. в США был задержан серийный убийца Джозеф Деанджело, который скрывался от правоохранителей более 40 лет [11].

ДНК-генеалогические исследования проводились и в белорусской популяции. Анализ этнических представителей показал, что у них те же Y-хромосомные линии, что и у русских, с теми же общими древними предками. Следовательно, оба восточнославянских народа близки по своему происхождению, различаясь лишь в процентном соотношении. Вместе с тем в этногенезе белорусов выделяется вклад древних балтских этносов. Генетически наша нация гомогенная, жители разных регионов страны мало отличаются друг от друга [12].

В рамках реализации научно-технической программы Союзного государства «Разработка инновационных геногеографических и геномных технологий идентификации личности и индивидуальных особенностей человека на основе изучения генофондов регионов Союзного государства» проведены работы по определению вероятного этногеографического и популяционного происхождения и созданию баз данных генетических характеристик населения указанного региона. По целевым маркерам было исследовано свыше 30 тыс. биологических образцов представителей более чем 200 этнорегиональных групп. Впервые получены генетические характеристики 15 локальных популяций 8 народов Восточной Европы по расширенной панели, включающей 21 аутосомный STR-маркер, применяемый для ДНК-идентификации и установления родства. Проведена оценка полиморфизма маркеров в российских, белорусских и молдавских популяциях, охарактеризованы уровни внутри- и межпопуляционной генетической дифференциации данных групп населения. Выборка из популяций Беларуси включала 6 историко-этнографических регионов республики: Центральное, Восточное и Западное Полесье, Поднепровье, Поозерье и Понеманье. Созданные базы данных частот аллелей аутосомных STR-локусов подходят для вероятностно-статистических расчетов при оценке уровня достоверности экспертного исследования в странах Союзного государства [13]. Полученные результаты позволили испол-

нителю с белорусской стороны – НПЦ Государственного комитета судебных экспертиз – разработать программно-информационные комплексы для определения наиболее вероятного этногеографического происхождения и популяционной принадлежности неизвестного индивида на основе генетических характеристик по маркерам мужских и женских генетических линий (рис. 2).

Российским ученым по результатам изучения по маркерам Y-хромосомы генетической структуры ненецких родов удалось выявить основную гаплогруппу, представителей которой в исследованной выборке абсолютное большинство. Генотипирование позволило с высокой степенью достоверности определить не только этническую и субпопуляционную принадлежность образцов, но в большей части случаев также их отношение к конкретным роду и даже фамилии [14].

Разработанные базы данных и установленные гаплогруппы будут дополняться по мере появления новых данных и расширения панелей ДНК-маркеров. ■

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Saad R. Discovery, development, and current applications of DNA identity testing. Proceedings (Baylor University. Medical Center) // <https://pubmed.ncbi.nlm.nih.gov/16200161/>.
2. Rogaev E.I., Grigorenko A.P., Moliaka Y.K. [et al.]. Genomic identification in the historical case of the Nicholas II royal family // PNAS. 2009. V. 106(13). P. 5258–5263.
3. Е.Я. Тетушкин. Генетическая генеалогия: история и методология // Генетика. 2011. Т. 47. №5. С. 581–596.
4. Pakstis A.J. et al. SNPs for a universal individual identification panel // Human Genetics. 2010. V. 127. P. 315–324.
5. Underhill P.A., Kivisild T. Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations // Annual Review of Genetics. 2007. V. 41. P. 539–564.
6. Kayser M. Forensic use of Y-chromosome DNA: a general overview // Human genetics. 2017. Vol. 136(5). P. 621–635.
7. Desalle R., Schierwater B., Hadrys H. MtDNA: The small workhorse of evolutionary studies // Frontiers in bioscience (Landmark edition). 2017. V. 22. P. 873–887.
8. Балановский О.П. Генофонд Европы / О.П. Балановский. – М., 2015.
9. R.J. Herrera, R. Garcia-Bertrand. The Nature of Evolution // Ancestral DNA, Human Origins, and Migrations. Academic Press. 2018. P. 1–31.
10. King T.E., Jobling M.A. What's in a name? Y chromosomes, surnames and the genetic genealogy revolution // Trends in Genetics. 2009. Aug. 25(8).
11. Phillips C. The Golden state killer investigation and the nascent field of forensic genealogy // Forensic Science International: Genetics. 2018. V. 36. P. 186–188.
12. И. Рожанский, И. Цыбовский, А. Богачева, С. Котова, Т. Забавская, Н. Шахнюк, А. Клесов. Белорусы: этногенез и связь с другими славянскими народами с позиции ДНК-генеалогии // Наука и инновации. 2013. №3. С. 55–62.
13. Н.В. Харьков, С.А. Котова, Н.А. Колесников и др. Генетическое разнообразие 21 аутосомного STR-маркера системы CODIS в популяциях Восточной Европы // Генетика. 2021. Т. 57. №12. С. 1396–1402.
14. В.Н. Харьков, Л.В. Валихова, Е.Л. Яковлева, В.Н. Сереброва, Н.А. Колесников, Т.И. Петелина, И.Ю. Хитринская, В.А. Степанов. Реконструкция происхождения гыданских ненцев на основе генетического анализа их родовой структуры с помощью нового набора YSTR-маркеров // Генетика. 2021. Т. 57. №12. С. 1403–1414.