Генетики построили детальную модель истории американских арктических народов

Статья, на днях опубликованная в журнале Nature (Flegontov et al. 2019), призвана поставить точку в многолетнем споре об истории народов Северной Америки: палеоэскимосов и народов языковой семьи на-дене.

Это исследование явилось результатом сотрудничества трех групп генетиков: под руководством Павла Флегонтова (Остравский университет, Чехия), Штефана Шиффельса (Институт Макса Планка в Йене, Германия) и Дэвида Райха (Гарвардская медицинская школа, Бостон, США), а также ряда генетиков, археологов и лингвиста Эдварда Вайды, специалиста по кетскому языку.

Еще в 2012 году Райхом и соавторами было показано (Reich et al. 2012), что частоты аутосомных SNP аллелей в выборке народа чиппева из Канады нельзя объяснить происхождением из одного азиатского источника, общего для всех американских индейцев, и что существовал второй азиатский источник, близкий недавно секвенированному геному индивида археологической культуры Саккак из Гренландии (Rasmussen et al. 2010). Этот же предковый источник, по данным группы Райха (Reich et al. 2012), внес примерно 50% в генофонд народов эскимосско-алеутской языковой семьи (алеутов и эскимосов, к последним относятся инуиты и юпики).

Чиппева говорят на языке атабаскской группы семьи на-дене, а индивид Саккак возрастом около 3900 лет является типичным представителем так называемых палеоэскимосов – охотников, господствовавших в американской Арктике и субарктике с 5000 до примерно 700 лет назад, и затем вытесненных предками современных инуитов и юпиков. Еще Джозеф Гринберг на основании в основном своей спорной классификации языковых семей (Greenberg et al. 1986) высказал гипотезу о трех волнах заселения Америки: 1) «первые американцы», давшие начало большинству американских индейцев, 2) народы, говорящие на языках семьи на-дене, 3) инуиты, юпики и алеуты. По данным археологии вторая волна заселения Америки представлена палеоэскимосами, проникшими на Аляску с Чукотки примерно 5000 лет назад, а третья волна – опять же юпиками и инуитами (Potter 2010). Ученик Гринберга Мерритт Рулен развил идеи более ранних авторов о родстве кетского языка и языков семьи на-дене (Ruhlen 1998), а Эдвард Вайда в 2010 году опубликовал масштабный сравнительный анализ енисейской семьи языков (кетского и ряда вымерших языков) и семьи на-дене, также придя к выводу об их “вертикальном” родстве. Работа Вайды была опубликована в рамках мультидисциплинарного тома обзоров и исследований о возможной связи енисейско-язычных и дене-язычных народов (Vajda 2010).

В том же 2010 году на сцену вышла палеогенетика. Примечательно, что первым древним секвенированным геномом человека современного вида был именно геном индивида культуры Саккак из Гренландии. В этой пионерской работе группы Эске Виллерслева (Копенгаген, Дания) ДНК была выделена из волос, найденных в вечной мерзлоте, и с помощью методов анализа главных компонент (PCA) и ADMIXTURE было показано родство палеоэскимоса с современными чукчами и коряками. На стоянках палеоэскимосов по каким-то причинам человеческие останки очень редки, и до сих пор геном «Саккак» с покрытием около 14x остается наиболее глубоко секвенированным представителем паолеоэскимосов. Недостатком широко используемых в исторической генетике методов PCA и ADMIXTURE является то, что они ничего не говорят о филогении, т.е. порядке разделения и смешения популяций. В 2012 году группа Райха опубликовала метод admixture graph (qpGraph), который позволяет строить дерево популяций и наносить на него «события смешения», и тут же применила его для анализа возможного вклада палеоэскимосов в генофонд американских индейцев (см. выше). В той же статье был впервые применен модный сейчас метод qpWave, который позволяет установить, какие предковые группы внесли вклад в исследуемую популяцию и каково было их число, однако при этом не опирается на филогенетическое дерево или граф смешения.

В 2014 и 2015 годах группа Эске Виллерслева вступила в дискуссию с группой Райха, опубликовав две работы в журнале

Science (Raghavan et al. 2014, 2015; обзор последней можно прочитать на сайте), в которых с использованием в первую очередь D-статистик, f-статистик, и филогенетического метода TreeMix был сделан вывод о том, что палеоэскимосы и народы на-дене не связаны генетически, что с предками на-дене смешивались инуиты, и что они же возможно смешались с палеоэскимосами на Аляске или на Чукотке. В 2018 году группа Эске Виллерслева продолжила исследовать историю палеоэскимосов, на этот раз в первую очередь методами TreeMix и qpGraph (Moreno-Mayar et al. 2018; см. обзор на сайте). Авторы построили модель, объясняющую геном атабасков как смесь северной клады американских индейцев и группы, родственной корякам (коренной народ Камчатки), а инуитов – как смесь палеоэскимосов и американских индейцев. Т.е., по последним результатам группы Виллерслева, в палеоэскимосской кладе коряки (народ чукотско-камчатской языковой семьи) группируются с предками на-дене, а палеоэскимосы — с предками инуитов.

С чем же связаны противоречивые результаты исследований? Флегонтов и соавторы считают, что здесь сошлось несколько факторов. Во-первых, все кандидаты на роль источника второго потока азиатских генов в популяции на-дене (коряки, палеоэскимосы, инуиты) относительно мало различаются генетически и обособились друг от друга, по-видимому, порядка 5000 тыс. лет назад. Во-вторых, возможно имели место многократные миграции этих групп туда и обратно через Берингов пролив и сопутствующие смешения популяций. В таких условиях – малые генетические различия между альтернативными источниками и сложная топология графа смешения с участием этих источников — стандартные в настоящее время методы палеогенетического анализа (f- и D-статистики, PCA, ADMIXTURE, TreeMix, qpWave и qpAdm) дают ненадежные результаты.

Для преодоления этой проблемы Флегонтов и соавторы собрали широкий арсенал методов, как стандартных (f-статистики, PCA, ADMIXTURE, qpWave и qpAdm, GlobeTrotter, ALDER), так и новых, а также впервые получили геномные данные методом capture для древних алеутов (11 индивидов со средними радиоуглеродными датировками от 280 до 2050 лет до н.в.), атабасков (3 индивида, датировка в среднем 710 лет до н.в.) и древних эскимосов Чукотки (могильники Эквен и Уэлен, 23 индивида со средними радиоуглеродными датировками от 620 до 1770 лет до н.в.). Также были получены геномные данные для 8 индивидов с могильника Усть-Белая на Ангаре (7 индивидов бронзового века с радиоуглеродными датировками от 4490 до 7020 лет до н.в. и один средневековый индивид возрастом около 610 лет до н.в.) и для палеоэскимоса культуры Дорсет (1760 лет до н.в.).

В статье также представлены данные генотипирования для современных популяций: инуитов с Аляски, кетов, нганасанов, энцев и селькупов (всего 93 индивида). См. результаты PCA на **рис. 1**, где представлены как новые, так и опубликованные данные для Евразии и Америки. Кроме многочисленных клин, свидетельствующих о смешении азиатов и европейцев (слева направо), четко заметна клина, протянувшаяся снизу вверх между палеоэскимосами и «первыми американцами». Примечательно, что все исследованные арктические и субарктические народы расположены на этой клине в следующем порядке: коряки и ительмены, чукчи, чукотские юпики, инуиты Аляски и древние чукотские эскимосы, древние алеуты, древние атабаски, современные атабаски, прочие индейцы севера Северной Америки (алгонкины и др.). Метод PCA лишь указывает на возможное смешение популяций, но не доказывает его. С помощью метода qpAdm авторы доказали, что геномы всех перечисленных групп успешно моделируются как смешение «первых американцев» и палеоэскимосов.

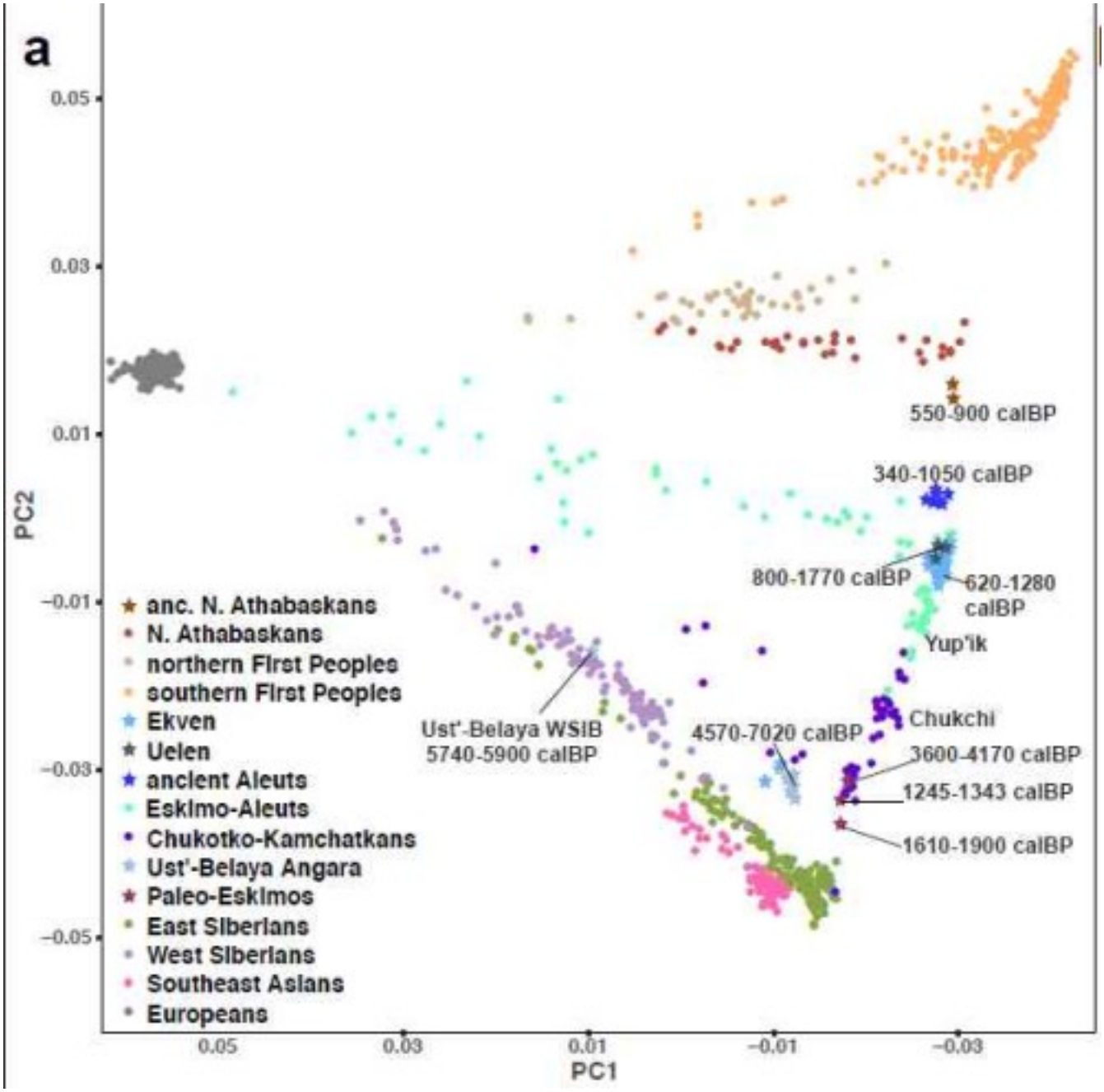


Рис. 1. Анализ главных компонент. На рисунке 940 индивидов, генотипированных на чипе Affymetrix HumanOrigins или секвенированных по протоколу 1240K capture, расположены в пространстве первой и второй главных компонент HumanOrigins dataset are plotted. Генетически нетипичные индивиды не были исключены из популяций для этого анализа, основанного на 586487 локусах. Для древних образцов приведены калиброванные радиоуглеродные датировки в годах до настоящего времени (95% доверительные интервалы для индивидов, минимальные и максимальные средние значения для групп). Древние индивиды обозначены звездочками. Популяции и группы популяций обозначены согласно легенде: N. Athabaskans – северные атабаски; northern First Peoples – «первые американцы» северной клады; southern First Peoples – «первые американцы» южной клады; Eskimo-Aleuts – народы эскимосско-алеутской языковой семьи, Chukotko-Kamchatkans – народы чукотско-камчатской языковой семьи.

Редкие аллели, достигающие глобальной частоты не более 5%, являются в основном результатом недавних мутаций, не успевших распространиться в популяциях, и поэтому особенно информативны относительно событий в последние 5 тыс. лет. Аналогичным свойством обладают аутосомные гаплотипы, анализируемые программой ChromoPainter. Штефан Шиффельс из Института Макса Планка в Йене разработал аналог f3-статистки, основанный на редких аллелях (rare allele sharing statistic, RASS), а также статистику, измеряющую суммарную длину гаплотипов, имеющих общее происхождение в паре индивидов (haplotype sharing statistic, HSS). На **рис. 2** показан пример применения двух RAS статистик на чукотских и американских полных геномах. И RASS, и HSS указывают на вклад палеоэскимосов в генофонд атабасков и алеутов, в то время как в геномах современных юпиков и инуитов есть сигнал еще одного потока генов — от чукотско-камчатских народов.

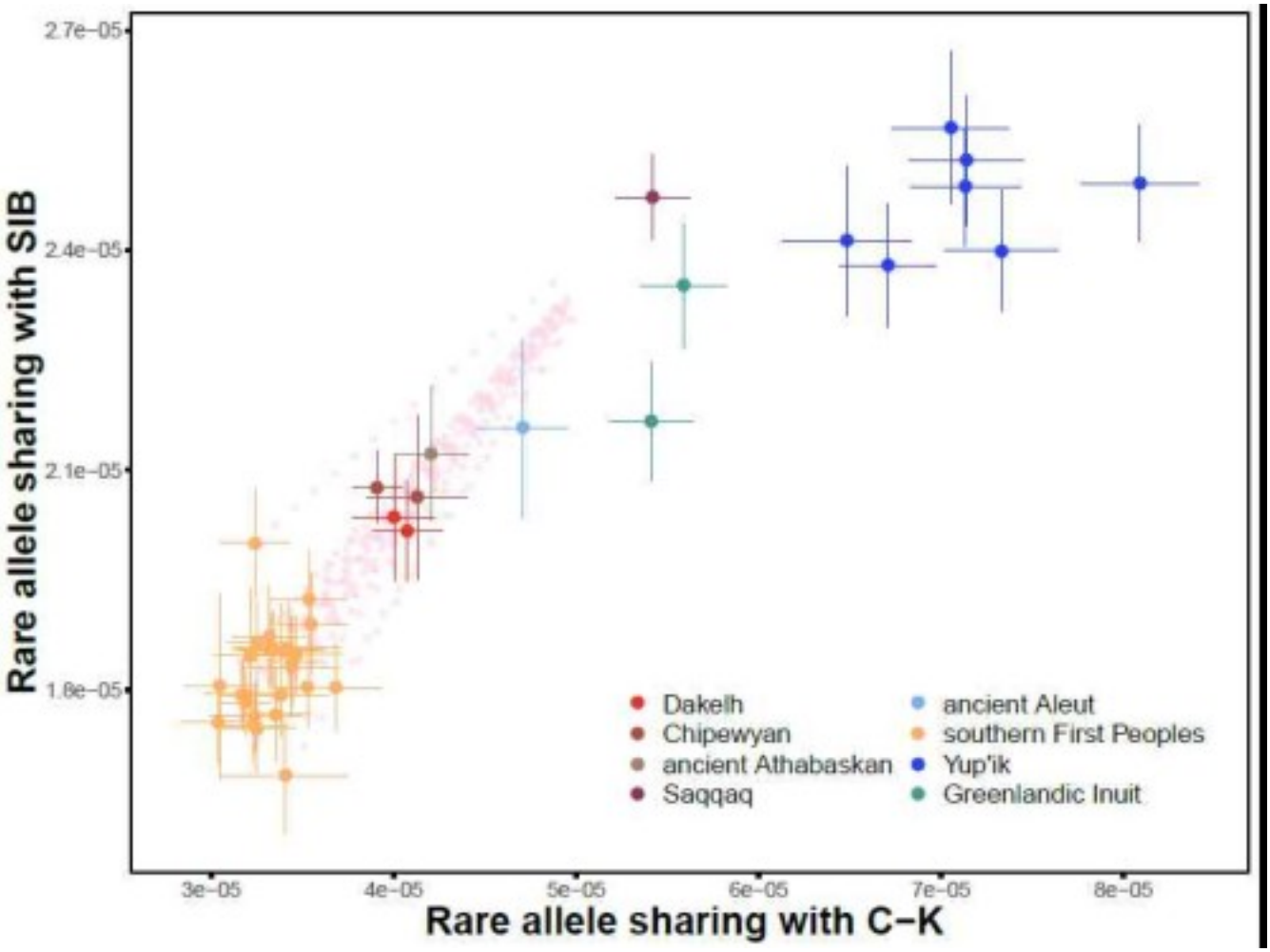


Рис. 2. RASS. Чукотско-камчатская (C-K) и сибирская (SIB) RASS для «первых американцев» (First Peoples), атабасков, народов эскимосско-алеутской языковой семьи и индивида Саккак. Каждая точка соответствует полному геному. Для вычисления статистик использованы аллели, встречающиеся от 2 до 5 раз в наборе 238 гаплоидных геномов Евразии и Африки (что соответствует частотам 0.8-2.1%), причем транзиции были удалены из анализа. Таким образом, данный анализ основан на 918474 локусов. Стандартные отклонения вычислены с помощью подхода jackknife, причем удалялись целые хромосомы. Авторы также симулировали RASS для геномов, образованных смешением в различной пропорции геномов «первых американцев» и Саккак (от 5% до 75% вклада Саккак, с шагом в 5%). Симулированные данные показаны розовыми кружками.

На **рис. 2** показаны RASS для отдельных индивидов. Для каждого индивида вычислены RASS относительно двух референсных популяций: чукотско-камчатских народов и прочих сибиряков. Локусы были выбраны согласно частоте минорного аллеля в референсном наборе геномов Евразии и Африки. Видно, что обе статистики, и «сибирская» и «чукотско-камчатская» повышены у современных и древних атабасков, Саккак, древнего алеута, юпиков и инуитов по сравнению с «первыми американцами». Причем и атабаски, и алеут лежат на прямой линии между Саккак и «первыми американцами», т.е. геномы первых групп в данной системе координат представляют собой смесь последних. Однако геномы чукотских юпиков и гренландских инуитов не лежат на этой прямой, так как имеют повышенную «чукотско-камчатскую» RASS, что говорит о дополнительном потоке генов из этого источника в истории инуитов и юпиков.

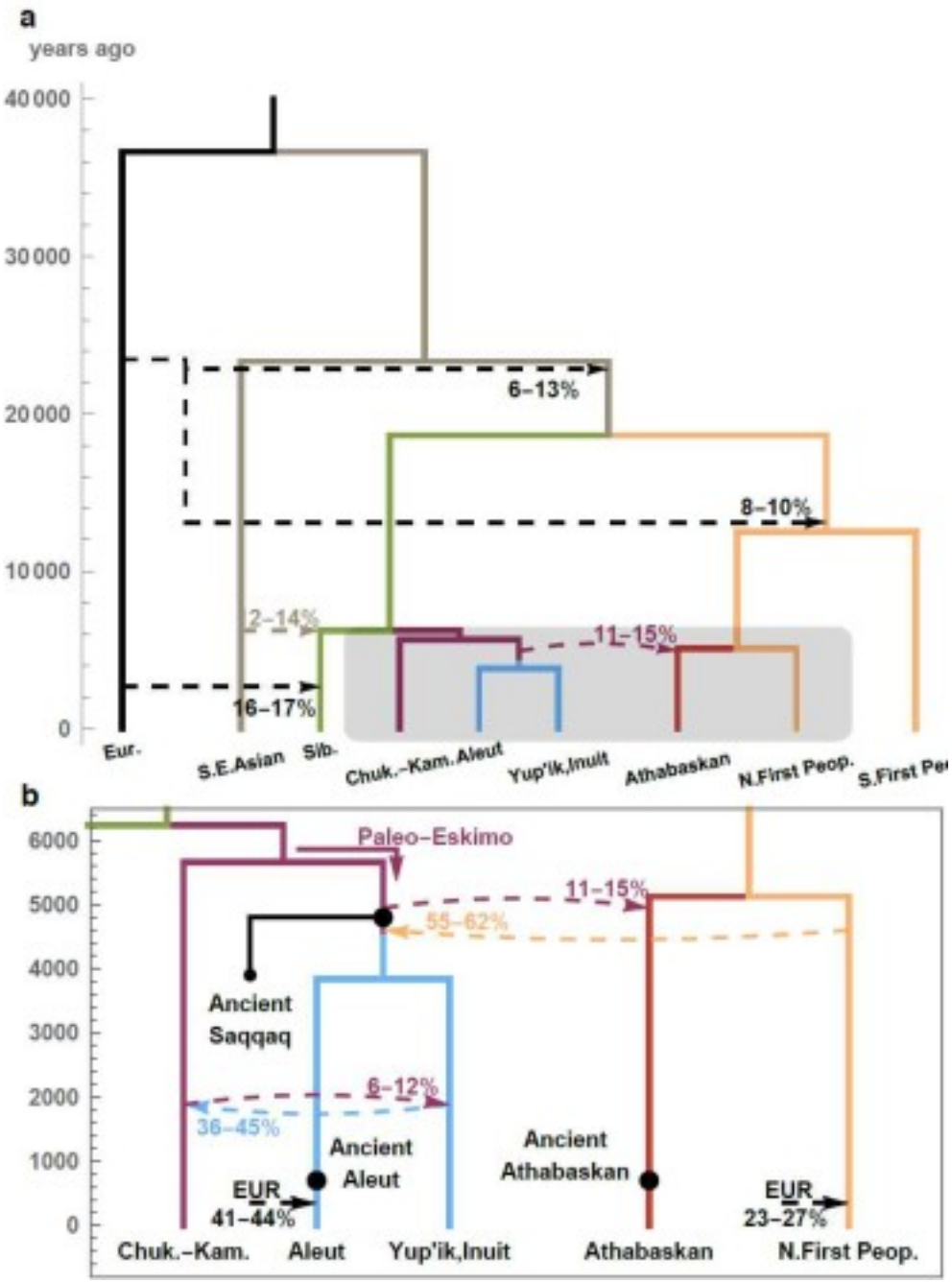


Рис. 3. Демографическая модель, основанная на 114 современных индивидах 9 мета-популяций. а) Граф смешения для европейцев (Eur.), жителей Юго-Восточной Азии (S.E. Asian), сибиряков (Sib.), чукотско-камчатских народов (Chuk.-Kam.), алеутов, юпиков и инуитов, атабасков и «первых американцев» северной и южной групп (N. First Peop., S. First Peop.), построенный с помощью комбинации методов qpGraph и Rarecoal. Потоки генов показаны штриховыми линиями. б) Увеличенный участок дерева (6000 лет и 5 мета-популяций), на котором более подробно показаны потоки генов между Азией и Америкой. Древние геномы (Саккак, алеут, атабаск) были картированы на граф, и их наиболее вероятные позиции показаны черными кружками. Также черными стрелками показаны потоки генов от европейцев (EUR) в современные популяции алеутов и северных американских индейцев. Датировки были вычислены, используя частоту мутаций на нуклеотид на поколение и длительность поколения 29 лет.

Однако прорывные результаты были получены методом admixture graph. Вначале были протестированы все возможные топологии (порядки ветвления) в палеоэскимосской кладе, состоящей из чукотско-камчатских народов, Саккак, атабасков, древних алеутов, и современных или древних эскимосов или инуитов. Три топологии «выиграли», т.е. их вероятности не отличались статистически значимо. Выяснилось, что двунаправленный поток генов между чукотско-камчатскими народами и эскимосами/инуитами существенно улучшает соответствие модели генетическим данным. На фоне двунаправленного потока были опять протестированы все возможные порядки ветвей, и на этот раз «выиграла» только одна модель (**рис. 3б**). На данном рисунке стоит обратить внимание на порядок ветвей: первыми отделяются чукотско-камчатские народы, затем палеоэскимосские предки атабасков, затем Саккак, затем происходит приток генов от «первых американцев» и формируется эскимосско-алеутская ветвь, разделяющая на алеутов и юпиков/инуитов.

Затем на массиве из 133 тысяч комбинаций популяций были протестированы две наиболее вероятные топологии: «топология Виллерслева» (Moreno-Mayar et al. 2018) и топология, полученная в обсуждаемой работе (**рис. 3б**). Наконец, несколько наиболее вероятных топологий были протестированы независимым методом Rarecoal, основанным на редких аллелях (**рис. 3**). В отличие от метода admixture graph, Rarecoal позволяет строить граф смешения в реальном времени (в поколениях). Для обсуждаемой статьи этот метод был существенно доработан Штефаном Шиффельсом. Таким образом, два независимых метода, основанные на графах, не подтвердили «топологию Виллерслева», и этот результат нашел также подтверждение с помощью RASS, HSS и иных методов. Датировки событий смешения были получены также независимыми методами GlobeTrotter и ALDER.

Так какова же предложенная авторами историческая модель и как она согласуется с лингвистикой и археологией? На **рис. 4а** представлена лучшая топология палеоэскимосской клады, наложенная на карту. Более 5000 лет назад произошло разделение палеоэскимосской и чукотско-камчатской популяций вследствие миграции первой через Берингов пролив. Около 4800 лет назад, т.е. вскоре после миграции (**рис. 3**) палеоэскимосы смешались с двумя группами «первых американцев». Одно из этих смешений (30-40% палеоэскимосской примеси) дало начало народам языковой семьи на-дене. Другое (около 50% палеоэскимосской примеси) – народам эскимосско-алеутской языковой семьи. Примечательно, что палеоэскимосская примесь обнаружена во всех ветвях семьи на-дене (тлингит, северные и южные атабаски), а у соседних народов других языковых семей ее уровень заметно ниже. Хотя генетические результаты не позволяют сделать однозначный вывод о родине праязыка на-дене (в Америке или в Сибири), мы можем предположить, что палеоэскимосы смешались именно с популяцией американцев, которая стояла у истока всех народов на-дене. Учитывая достаточно близкое родство палеоэскимосов с современными кетами (Flegontov et al. 2016 и неопубликованные результаты группы Флегонтова), гипотеза Вайды о вертикальном родстве енисейских языков и языков на-дене теперь не выглядит невероятной с точки зрения генетики.

Что касается истории эскимосов и алеутов, смешение индейцев и палеоэскимосов датируется в диапазоне 4900-4400 лет до н.в. (**рис 4б**). Хотя у нас нет образцов данного возраста с Аляски, совмещая генетические и археологические результаты, мы можем предположить, что это событие играло ключевую роль в этногенезе эскимосско-алеутских групп и произошло на юге штата Аляска: в районе архипелага Кадьяк и полуострова Аляска. Именно там обнаружены самые ранние сообщества, почти целиком полагавшиеся на морские ресурсы, а также некоторые особенности материальной культуры, типичные для позднейших эскимосов и алеутов (обозначены коричневым овалом на **рис. 4б**). Вскоре после смешения с индейцами предки алеутов вероятно мигрировали на Алеутские острова (тонкая голубая стрелка, направленная вниз на **рис. 4б**) и затем пребывали в относительной изоляции, что хорошо объясняет отсутствие чукотско-камчатской примеси именно у алеутов. Другая группа вероятно мигрировала на север в сторону Берингова пролива и там возможно оставила след в виде археологических культур Древнекитобойной, Чорис и Нортон (наиболее ранние датировки 3100-2500 лет до н.в.), сочетающих как палеоэскимосские, так и инуитские черты. Затем данная группа по каким-то причинам вернулась на Чукотку (тонкие голубые стрелки на **рис. 4б**), где около 2200 лет назад обнаруживается древнеберингоморская археологическая культура уже типично эскимосско-инуитского облика. Около 2000 лет назад, согласно модели Rarecoal и по результатам анализа индивидов из эквенского и уэленского могильников возрастом до 1800 лет до н.в. (Flegontov et al. 2019), произошло двунаправленное смешение эскимосов и предков чукчей и коряков (сиреневые стрелки на **рис. 4б**). Освоение охоты на китов в древнеберингоморской культуре привело к резкому росту численности и экспансии инуитов на Аляску примерно 1150 лет до н.в. В дальнейшем инуиты распространились по всей американской Арктике и вытеснили все еще живших там палеоэскимосов (толстая голубая стрелка на **рис. 4б**).

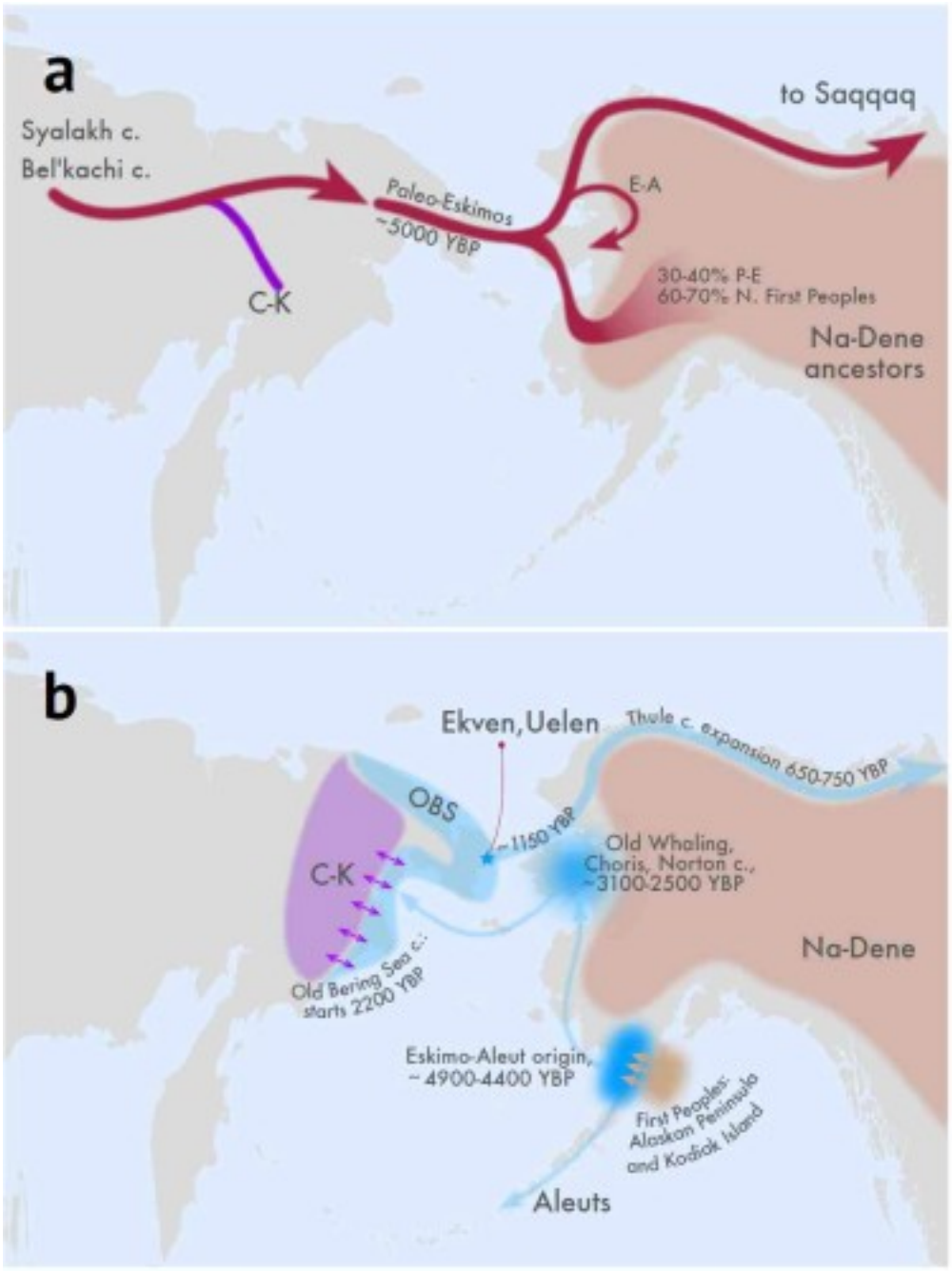


Рис. 4. Археологическая и географическая интерпретация модели, предложенной Флегонтовым и соавторами. а) Топология палеоэскимосской клады, наложенная на карту. Авторы предположили, что контакт между палеоэскимосами и предками на-дене происходил на границе между областями археологических культурх Arctic Small Tools и Northern Archaic на Аляске. Сокращения: E-A – предки народов эскимосско-алеутской языковой семьи, P-E – палеоэскимосы, N. First Peoples – американские индейцы северной клады. б) Модель истории эскимосско-алеутских народов, построенная на основании генетических и археологических результатов. Показаны движения предковых групп через Берингов пролив в Чукотку и обратно, а также двунаправленный поток генов на Чукотке между предками юпиков и инуитов (древнеберингоморская археологическая культура, OBS) и народами чукотско-камчатской языковой семьи (C-K). На обоих панелях для археологических культур и миграций указаны наиболее ранние даты в годах до настоящего времени (YBP). Некоторые пути миграций показаны лишь приблизительно.

текст Павла Флегонтова

Литература

Flegontov, P. *et al*. Genomic study of the Ket: A Paleo-Eskimo-related ethnic group with significant ancient North Eurasian ancestry. *Scientific Reports* **6**, 20768 (2016).

Flegontov, P. *et al*. Paleo-Eskimo genetic ancestry and the peopling of Chukotka and North America. *Nature* (2019).

Moreno-Mayar, J. V. *et al*. Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature* **553**, 203–207 (2018).

Potter, B. A. Archaeological patterning in Northeast Asia and Northwest North America: an examination of the Dene-Yeniseian hypothesis. *The Dene-Yeniseian Connection*, ed. Kari, J., Potter, B. A. *Anthropological Papers of the University of Alaska: New Series* **5**, 138–167 (2010).

Raghavan, M. *et al*. The genetic prehistory of the New World Arctic. *Science* **345**, 1255832 (2014).

Raghavan, M. *et al*. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* **349**, 1–20 (2015).

Rasmussen, M. *et al*. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757–762 (2010).

Reich, D. *et al*. Reconstructing Native American population history. *Nature* **488**, 370–374 (2012).

Ruhlen, M. The origin of the Na-Dene. *Proc. Natl. Acad. Sci. USA* **95**, 13994–13996 (1998).

Vajda, E. J. Siberian link with Na-Dene languages. *The Dene-Yeniseian Connection*, ed. Kari, J., Potter, B. A. *Anthropological Papers of the University of Alaska: New Series* **5**, 33–99 (2010).

Источник:

Pavel Flegontov, N. Ezgi Altınışık, Piya Changmai et al. Palaeo-Eskimo genetic ancestry and the peopling of

Chukotka and North America // Nature. 2019. https://www.nature.com/articles/s41586-019-1251-y