

Филогения и филогеография Y-хромосомной гаплогруппы N1a1a1a-F1419

Харьков В.^{1*}, Валихова Л.¹, Хитринская И.¹, Зарубин А.¹, Адамов Д.², Степанов В.¹

¹ Научно-исследовательский институт медицинской генетики, Томский НИМЦ РАН, Томск, Россия

² Северо-Восточный федеральный университет им. М.К. Аммосова, Якутск, Россия

* vladimir-kharkov@medgenetics.ru

Ключевые слова: Y-хромосома; гаплогруппа N1a1; филогения; филогеография

Мотивация и цель: Подробная детализация филогении гаплогрупп Y-хромосомы и оценки возраста различных линий за счет оценки результатов полногеномного секвенирования вносит существенный вклад в определение их территориального происхождения, историю демографических процессов и миграций. Обнаружение новых терминальных SNP для различных сублиний Y-хромосомы позволяет уточнить специфичность всех исследуемых мужских линий, детализировать их этнические различия и оценить время расхождения всех ветвей любой гаплогруппы. Целью работы явилась подробная филогения северо-евразийской гаплогруппы N1a1a1a-F1419 с помощью полногеномного секвенирования Y-хромосом из различных филогенетических ветвей основных линий и отдельных редких образцов со специфичными YSTR-гаплотипами. Для более детальной филогении этноспецифичных линий было увеличено количество образцов мужчин из новых популяционных выборок. Это позволяет максимально детально охарактеризовать структуру этой гаплогруппы на основании извлечения данных по новым YSNP из полных геномов, а также сопоставлении их с данными гаплотипов по YSTR-маркерам.

Методы и алгоритмы: Были извлечены специфичные для гаплогруппы N1a1a1a-F1419 SNP из файлов с полногеномным секвенированием методом NGS у 202 образцов ДНК из различных популяций Сибири, Восточной Европы, Дальнего Востока, Средней Азии и Северного Кавказа. Проведен расчет возраста отдельных ветвей по соответствующим выборкам образцов [1]. Рассчитан общий размер прочитанных нуклеотидных последовательностей образцов выборки внутри границ области combBED [2]. Отбор производных вариантов осуществлялся по координате референсной последовательности hg38, попадающей в пределы регионов combBED. Расчет погрешности оценки возраста основывается на предположении о пуассоновском характере процесса SNP мутаций [3]. Анализ STR-гаплотипов проводили с применением 45 YSTR (DYS19, 385a, 385b, 388, 389I, 389II, 390, 391, 392, 393, 426, 434, 435, 436, 437, 438, 439, 442, 444, 445, 448, 449, 456, 458, 460, 462, 481, 504, 505, 518, 525, 531, 533, 537, 552, 570, 576, 635, 643, YCAIIa, YCAIIb, GATA H4.1, Y-GATA-A10, GGAAT1B07). Полученные оценки возраста сравнивались с данными работы [4] и расчетами YFull.

Результаты: Детализированы филогения и территориальное распределение специфичных сублиний гаплогруппы N1a1a1a1a-F1419 Y-хромосомы за счет расширения спектра SNP. У многих образцов мужчин, относящихся к сибирским и европейским популяциям, обнаружено большое количество новых, не описанных ранее SNP. В общем массиве маркеров они составляют около 20 %, но для различных образцов из этноспецифичных ветвей их доля в терминальных SNP порой доходит до 100 %. Эти результаты позволяют более точно выявлять и дифференцировать ее варианты у мужчин различных родов и субэтнических групп. По результатам выявления новых YSNP и привязке образцов к детализированным сублиниям в различных гаплогруппах был проведен расчет возраста по стандартной формуле с учетом рассчитанного ранее темпа мутирования SNP, указанного в различных статьях и на сайте YFull. Проведено генотипирование широкого набора информативных для детализации этой гаплогруппы YSNP, извлеченных из просеквенированных образцов на имеющихся популяционных выборках. Почти все SNP полностью подтвердили свои аллели на уже просеквенированных образцах и на всех остальных образцах из этих гаплогрупп. Показано, что большая часть новых терминальных YSNP, в сочетании с YSTR-гаплотипами, обладает большой этнопопуляционной специфичностью. В большинстве случаев сравнительный анализ случайно выбранных образцов позволяет провести этнотерриториальную привязку образца с точностью до этноса, субэтноса или рода, а в отдельных случаях и до конкретного населенного пункта и носителей гаплотипов YSTR. Выполнено построение филогенетического древа этой гаплогруппы по 202 образцам и медианных сетей YSTR для ее различных линий от общего предка по этой гаплогруппе до отдельных образцов мужчин со специфичными терминальными YSNP. В качестве примера можно привести оценку возраста специфичной якутско-эвенкийской линии N-M1984, который составил 960 лет. Возраст более ранней линии N-M2016, к которой вместе якутами и эвенками принадлежат отдельные образцы узбеков и хакасов, определен в 2520 лет. Общий возраст гаплогруппы N1a1a по всем секвенированным образцам сибирских, среднеазиатских и европейских популяций составил 14 020 лет. Возраст специфичной линии для бурятов составил 980 лет, у сибирских татар – 1400 лет, у хакасов и шорцев – 740 лет. Практически для всех генотипированных образцов получена точная привязка к конкретным специфичным сублиниям, которые полностью соответствуют их этнической, субэтнической и родовой принадлежности. В качестве примера можно привести данные о различиях по Y-хромосомным гаплогруппам у тундровых и лесных ненцев. У лесных ненцев полностью доминирует гаплогруппа N1a1a1a1a2a1c1~ (Y13852), к которой принадлежат все мужчины рода Пяк. Возраст рода Пяк лесных ненцев составил 451 год, род тундровых ненцев Салиндер из фратрии Вануйто – 331 год, рода Тибичи – 216 лет, что хорошо согласуется с данными антропологов и этнографов. Предковая для коми, марийцев и удмуртов линия N-Z35179 имеет возраст 2800 лет, общая для марийцев и удмуртов европейская линия N-Z35198 – 1660 лет. Линия N-Y40117, к которой принадлежат три образца коми, имеет возраст 200 лет. Возраст линии N-Z35193 у трех удмуртов составляет 300 лет. В целом результаты расчетов возраста ветвей гаплогруппы N1a1a1a1a-F1419 согласуются с данными сравниваемых работ, но в некоторых случаях имеются различия. Часть наших образцов формирует отдельные сублинии, которые не были описаны ранее, и позволяет более подробно охарактеризовать филогению

этих линий. Была проведена картографическая привязка основных линий гаплогруппы N1a1a1a1a-F1419 по их частоте в исследованных популяциях.

Выводы: Филогения гаплогруппы N1a1a1a1a-F1419 подтверждает, что местом ее исходного распространения по территории Сибири и Европы была зона Южной Сибири или прилегающих территорий современной Монголии. Все образцы характеризуются специфичным спектром гаплотипов, подчеркивающим недавний эффект основателя у различных этноспецифичных линий этой гаплогруппы. Оценка возраста по YSNP и YSTR показывает, что представители родов у хакасов, шорцев, ненцев, северных и южных алтайцев являются родственниками по мужской линии и имеют родоначальника, жившего в относительно недалеком прошлом. Эти результаты позволяют гораздо точнее разделить филогенетическую структуру этой гаплогруппы и точнее определить возраст их расхождения, детализировать их происхождение и связь с миграциями их носителей в прошлом.

Финансирование: Исследование поддержано за счет гранта Российского научного фонда (№ 22-64-00060, <https://rscf.ru/project/22-64-00060/>).

Phylogeny and phylogeography of the Y-chromosome haplogroup N1a1a1a1a-F1419

Kharkov V.^{1*}, Valikhova L.¹, Khitrinskaya I.¹, Zarubin A.¹, Adamov D.², Stepanov V.¹

¹ Research Institute of Medical Genetics, Tomsk National Research Medical Center, RAS, Tomsk, Russia

² M.K. Ammosov North-Eastern Federal University, Yakutsk, Russia

* vladimir-kharkov@medgenetics.ru

Key words: Y-chromosome; haplogroup N1a1; phylogeny; phylogeography

Motivation and Aim: Detailed analysis of the phylogeny of Y-chromosome haplogroups and estimation of the age of different lineages through the evaluation of whole-genome sequencing results contributes significantly to determining their geographical origin, history of demographic events, and migrations. The discovery of novel single nucleotide polymorphisms (SNPs) for various subclades of the Y chromosome allows us to clarify the specific features of all male lineages under study, detailing their ethnic differences and estimating the time of separation of all branches within any haplogroup. Our goal was to conduct a detailed phylogenetic analysis of the North Eurasian N1a1a1a1a-F1419 haplogroup using whole-genome sequences of Y chromosomes from different phylogenetic lineages and isolated samples with specific Y-STR haplotypes. To achieve a more comprehensive phylogenetic analysis of ethnically specific lineages, we increased the number of samples collected from new populations. This allows us to characterize the structure of this haplogroup as accurately as possible by extracting data on new single nucleotide polymorphisms (SNPs) from complete genomes and comparing them with haplotypes using Y-STR markers.

Methods and Algorithms: Haplogroup-specific N1a1a1a1a-F1419 single nucleotide polymorphisms (SNPs) were extracted from genome-wide next-generation sequencing (NGS) data in at least 202 DNA samples from diverse populations in Siberia, Eastern Europe, the Far East, Central Asia, and the North Caucasus. Age estimation was performed using the extracted SNP data. Separate branches were created according to the corresponding samples [1]. The total size of the nucleotide indices calculated for each sample within the combBED [2]. Another tool was created using the hg38

coordinate system, which is part of the combBED. The calculation of the probability of these changes occurring is based on the assumption that Poisson surgery is a mutant [3]. DNA haplotypes were analyzed using the following markers: DYS19, 385a, 385b, 388, 389I, 389II, 390, 391, 392, 393, 426, 434, 435, 436, 437, 438, 439, 442, 444, 445, 448, 449, 456, 458, 460, 462, 481, 504, 505, 518, 525, 531, 533, 537, 552, 570, 576, 635, 643, YCAIIa, YCAIIb, GATA H4.1, Y-GATA-A10, GGAAT1B07. The age estimates were compared with the data from [4] and YFull calculations.

Results: The phylogeny and geographic distribution of specific subclades of the N1a1a1a1a-F1419 haplogroup of the Y chromosome are described by expanding the range of single nucleotide polymorphisms (SNPs). A large number of novel, previously unreported SNPs have been identified in many samples from men belonging to the Siberian and European populations. These SNPs account for approximately 20 % of the total set of markers, but in some samples from ethnic subgroups, their proportion among terminal SNPs can reach 100 %. These findings allow for more accurate identification and differentiation of its variants among men from different genders and ethnic backgrounds. Based on the identification of new YSNPs and the linkage of samples to specific subclusters within various haplogroups, an age estimate was calculated using a standard formula that takes into account the previously reported mutation rate of SNPs, as reported in various articles and on the YFull database. We performed genotyping of a wide range of YSNPs that were informative for detailed analysis of this haplogroup, extracted from sequenced samples using available population samples. Most of the SNPs had their alleles fully confirmed in the already sequenced samples and in all other samples from these haplogroups. It has been shown that most of the newly identified YSNPs in combination with YSTR haplotypes have a high level of ethno-population specificity. In most cases, a comparative analysis of randomly selected samples allows for an ethno-geographical reference of a sample with accuracy in terms of ethnicity, sub-ethnicity, or clan, and in some cases, even to a specific location and carriers of YSTR haplotypes. A phylogenetic tree of this haplogroup was constructed using 202 samples and median YSTR networks for its various branches from the common ancestor of this haplogroup to individual samples of men with specific terminal YSNPs. As an example, we can estimate the age of the specific Yakut-Evenki lineage, N-M1984, at 960 years old. The age of an earlier lineage, N-M2016, which includes individual samples from Uzbeks, Khakass, and Yakuts, is estimated at 2,520 years old. The total estimated age of haplogroup N1a1a for all sequenced samples from Siberian, Central Asian, and European populations is 14,020 years old. The age of the Buryat lineage is 980 years old, the Siberian Tatar lineage is 1,400 years old, and the Khakass and Shor lineage is 740 years old. For most genotyped samples, a specific link to a specific sublineage has been obtained, which corresponds to their ethnic and subethnic affiliation. As an example of this, we can refer to data on differences in Y-chromosome haplogroups between tundra and forest Nenets populations. Among forest Nenets, the haplogroup N1a1a1a1a2a1c1 (Y13852) is dominant, and all men from the Pyak clan belong to this group. The age of the Pyak clan of Forest Nenets is 451 years, while the tundra Nenets Salinder from Vanuito Phratry is 331 years old and the Tibich clan is 216 years old. These ages agree well with the data provided by anthropologists and ethnographers. The ancestral lineage N-Z35179 among the Komi, Mari, and Udmurt people is 2,800 years old, while the European lineage N-Z35198 common to the Mari and Udmurts is 1,660 years old. The line N-Y40117, which includes three Komi samples, is only 200 years old. Among the three Udmurt samples, the age of N-Z35193 is 300 years.

Overall, the results from calculating the ages of branches of haplogroup N1a1a1a1a-F1419 agree with the data from the compared studies, although there are some differences in the exact ages in some cases. Some of our samples have formed separate sublineages that were not previously described and allow us to better understand the phylogeny of these lineages. A study was conducted to map the main lines of the haplogroup N1a1a1a1a-F1419, based on their frequency in the populations under study. *Conclusion:* The phylogeny of N1a1a1a1a-F1419 haplogroup confirms that its initial distribution in Siberia and Europe occurred in the southern part of Siberia or neighboring regions of modern Mongolia. All the samples share a specific set of haplotypes, indicating a recent founder effect within various ethnically specific lineages of this group. Based on YSNP and YSTR analysis, the age estimates suggest that members of the Khakass, Shors, Nenets, Northern and Southern Altaians share a common ancestor who lived relatively recently in the past. This allows us to better understand the phylogenetic relationships within this haplogroup and determine the time of divergence more precisely. It also helps us to trace their origins and connections with past migrations.

Funding: The study is supported by a grant from the Russian Science Foundation (No. 22-64-00060, <https://rscf.ru/project/22-64-00060/>).

Список литературы/References

1. Adamov D.S. The Yakut branch of Y-chromosome as a part of the haplogroup N-M2016. *Sib Res.* 2022;2(8):29-36. doi 10.33384/26587270.2022.08.02.05e
2. Adamov D., Guryanov V., Karzhavin S., Tagankin V., Urasin V. Defining a new rate constant for Y-chromosome SNPs based on full sequencing data. *Russ J Genet Genealogy.* 2015;7(1):68-89
3. Poznik D., Henn B., Yee M.-C. et al. Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. *Science.* 2013;341(6145):562-565. doi 10.1126/science.1237619
4. Ilumäe A.M., Reidla M., Chukhryaeva M. et al. Human Y chromosome haplogroup N: A non-trivial time-resolved phylogeography that cuts across language families. *Am J Hum Genet.* 2016;99(1):163-173. doi 10.1016/j.ajhg.2016.05.025