

Материалы по истории развития *Populus alba*, произрастающего в европейской части России и на Кавказе, основанные на результатах полногеномного секвенирования

Попченко М.^{1*}, Гладыш Н.¹, Володин В.^{1, 2}, Богданова А.¹, Краснов Г.^{1, 2}, Кудрявцева А.^{1, 2}

¹ Институт молекулярной биологии им. В.А. Энгельгардта РАН, Москва, Россия

² Центр высокоточного редактирования генома и генетических технологий для биомедицины, Институт молекулярной биологии им. В.А. Энгельгардта РАН, Москва, Россия

* popchenko_m@inbox.ru

Ключевые слова: *Populus alba*; полногеномное секвенирование; филогенетика; популяционная генетика; эволюция

Мотивация и цель: Белый тополь (*Populus alba* L.), эволюционно молодой вид двудомных древесных растений, обычно встречается в умеренных и субтропических регионах Евразии и Северной Африки. В России его можно встретить на юге европейской части, юге Западной Сибири и на Кавказе. В ареале между европейской и кавказской частями существует разрыв. Несмотря на экономическую значимость этого вида и его большой потенциал для использования, молекулярные исследования филогеографии популяций белого тополя в России не проводились. Цель данного исследования – представить гипотезу о формировании природных и культурных ареалов белых тополей в европейской части России.

Методы и алгоритмы: Мы секвенировали полные геномы 36 растений белого тополя, в том числе 12 из Москвы и прилегающих к ней регионов, 2 из Нижегородской, 6 из Волгоградской областей, 9 с Северного Кавказа и 2 из Дагестана. Мы также добавили 3 растения из коллекции Центрального сибирского ботанического сада (местная популяция белых тополей, гибриды между белым тополем и осиной – *Populus canescens* (Aiton) Sm. и var. ‘Хоперский’) и 2 растения, собранные близ Алма-Аты в Казахстане. Секвенирование проводили с использованием платформы Illumina NextSeq 500 (в режиме 2×150 нукл.) с покрытием в диапазоне от 7х до 20х. Данные секвенирования опубликованы в NCBI BioProject (PRJNA1109753). Филогенетический анализ был проведен на основе сравнения профилей замен (SNP), включая точечные SNV и вставки/делеции (indels). Такой анализ позволяет учесть возможность полиплоидии. Прочтения были картированы на референсный геном белого тополя (GCF_005239225) с использованием BWA-MEM (-k 14) с удаленными вторичными выравниваниями [1]. Для поиска замен использовался freeBayes. Далее для векторов частот вариантов аллелей (VAF) было рассчитано попарное евклидово расстояние между образцами. Затем образцы были кластеризованы с использованием алгоритма UPGMA с бутстреппингом. Оценка смешения популяций проводилась с использованием NGSAdmix на основе оценки вероятности генотипирования для всех SNP (-K 3) [2]. Этот метод позволяет оценить наличие генетических примесей в популяциях даже на данных с небольшой глубиной

секвенирования. Для повышения точности в анализ были включены дополнительные образцы деревьев белого тополя, произрастающие в китайской провинции Синьцзян и в Италии (данные получены из NCBI SRA).

Результаты: На основе анализа схожести профилей SNP-профилей можно выделить по крайней мере три генетически сходные группы: образцы из Нижегородской, Волгоградской областей и Новосибирска образуют кластер (C1), в то время как образцы с Северного Кавказа и Южного Дагестана образуют другой кластер (C2). Образцы из Москвы и прилегающих районов образуют третий кластер (C3). Хотя дендрограммы генетического сходства для хлоропластных, митохондриальных и ядерных геномом различаются, на всех дендрограммах наблюдаются три крупных кластера, причем два кластера в основном состоят из деревьев центрального региона, а еще одна однородная группа включает деревья Сочи, Дагестана и Северо-Кавказского федерального округа (рис. 1).

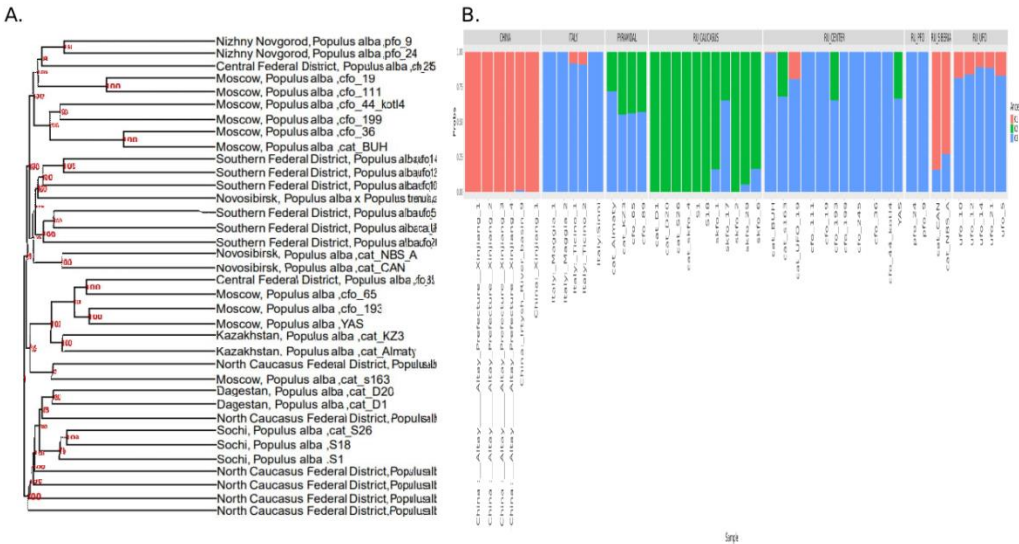


Рис. 1. А – филогенетический анализ деревьев белого тополя, произрастающего на территории России и Казахстана. «Удлиненные ветви» на дендрограмме являются следствием неточностей в определении значений VAF, которые рассчитываются по числу прочтений, подкрепляющих референсную и альтернативную аллель. В – оценка смеси условных популяций для тополей из разных регионов. Некоторые образцы опущены

Проведенный анализ смешения условных популяций белого тополя (NGSAdmix) был основан на гипотезе о существовании трех основных центров, которые могли послужить основой для формирования современной популяции белого тополя на юге европейской части России в послеледниковый период. К таким центрам относятся: Южная Европа, Кавказский регион, а также южные районы Западной Сибири и Южного Урала. В период максимального оледенения на этих территориях сохранялась лесная растительность. Генетический материал из Южной Европы преобладает в геномах популяций Центральной Европы и в лесных районах России. В степях Волгоградской области до четверти генома исследованных образцов содержит генетический материал из Южной Сибири. Следует отметить, что небольшое количество генетического материала из

Западной Сибири присутствует в популяциях Центральной Европы. Также генетический материал из Южной Европы был обнаружен в популяциях Западной Сибири, что вполне естественно, учитывая непрерывное распространение видов в доледниковый период. Кавказский генетический материал полностью формирует геном растений в популяциях Закавказья и Северного Кавказа. Большинство образцов содержат в своих геномах южноевропейский генетический материал. Наличие этого генетического материала можно объяснить историческими связями между Европой и Кавказом через Малую Азию, где белый тополь растет и по сей день, а также миграцией его носителей с севера на юг через долины рек Дон и Кубань. Интересной особенностью является полное отсутствие южносибирского генетического материала у европеоидных популяций, и наоборот. По-видимому, существует разрыв в распространении белого тополя, поскольку он произрастает в восточном Туркменистане, северо-восточном Иране, равнинном Узбекистане и северо-западном Афганистане, что служит препятствием для обмена генетическим материалом между этими регионами. Точное распространение белого тополя в Центральной Азии остается спорным из-за недостатка информации о цветоводстве и долгой истории его выращивания в регионе.

Генетический состав популяций растений в центрально-европейской части России требует отдельного рассмотрения. При анализе образцов растений здесь можно обнаружить следы двух из трех очагов. Генетический материал Южной Европы получен как из южнорусской, так и из центральноевропейской популяций. Роль последних, по-видимому, возрастает вблизи крупных городов и в северо-западных регионах, где это связано с периодом появления парков в конце XIX – начале XX в. Кавказский генетический материал, с другой стороны, связан с интродукцией в середине XX в., когда ботанические сады и селекционные центры начали пополнять свои коллекции материалом со всей страны. Особый интерес представляют пирамидальные формы белого тополя в центральной части Европы. Считается, что эти экземпляры произошли от среднеазиатских пирамидальных типов. Однако, как показал анализ, в их составе присутствует только южно-европейский и кавказский генетический материал. Из-за ограниченного объема доступной информации невозможно дать однозначное объяснение этому открытию. Следовательно, необходим более детальный анализ геномных последовательностей образцов пирамидальных форм белого тополя, произрастающих в Южной Европе и Центральной Азии.

Выводы: Современное распространение белого тополя на юге европейской части России сформировалось в послеледниковый период в результате миграции популяций с Балканского полуострова, Восточной и Западной Сибири и Южного Урала. Кавказские популяции белого тополя не способствовали расселению на этой территории в исторические времена. Современные культивируемые белые тополя в европейской части России включают растения различного происхождения (Центральная Европа, Юг России и Кавказ).

Финансирование: Исследование поддержано Российским научным фондом (грант № 22-14-004).

Materials on the history of the development of *Populus alba*, which grows in the European part of Russia and the Caucasus, are based on whole-genome sequencing

Popchenko M.^{1*}, Gladyshev N.¹, Volodin V.^{1, 2}, Bogdanova A.¹, Krasnov G.¹,
Kudryavtseva A.^{1, 2}

¹ Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia

² Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia

* popchenko_m@inbox.ru

Key words: *Populus alba*; whole-genome sequencing; phylogenetics; population genetics; evolution

Motivation and Aim: The white poplar (*Populus alba* L.), an evolutionarily young species of dioecious woody plant, is commonly found in temperate and subtropical regions of Eurasia and North Africa. In Russia, it can be found in the southern European part, southern Western Siberia, and the Caucasus. There is a gap in the range between the European and Caucasian parts. Despite the economic importance of this species and its great potential for use, no molecular studies have been conducted on the phylogeography of white poplar populations in Russia. This study seeks to present a hypothesis on the formation of natural and cultivated areas of white poplars in the European part of Russia.

Methods and Algorithms: We sequenced the complete genomes of 36 white poplar plants, including 12 from Moscow and its surrounding regions, 2 from Nizhny Novgorod, 6 from Volgograd regions, 9 from the North Caucasus and 2 from Dagestan. We also added 3 plants from the collection of the Central Siberian Botanical Garden (a local population of white poplars, hybrids between white poplar and aspen – *Populus canescens* (Aiton) Sm. and var. 'Khopersky'), and 2 plants collected near Alma-Ata in Kazakhstan. Sequencing was performed using the Illumina NextSeq 500 platform (2x150 bp), with coverage ranging from 7x to 20x for each sample. The sequencing results are published in NCBI BioProject (PRJNA1109753).

A phylogenetic analysis was conducted based on SNPs profiling, including point SNVs and insertions/deletions (indels). Reads were mapped to the white poplar reference genome (GCF_005239225), using BWA-MEM (-k 14) with removed secondary alignments [1]. FreeBayes was used for variant calling. The Euclidean distance between samples was calculated for variant allele frequency (VAF) vectors, and samples were clustered using UPGMA algorithm with bootstrapping, taking into account the possibility of polyploidy. The assessment of population mixing was performed using NGSAdmix software, based on the assessment of the probability of genotypes for all SNPs (-K 3) [2]. This method allows for the evaluation of admixture at a low sequencing depth. For the accuracy of the analysis, additional samples of white poplar trees were used, growing in the Xinjiang province of China and in Italy. The raw genomic data for these samples were obtained from the NCBI SRA.

Results: Based on the analysis of SNP profiles, at least three genetically similar groups can be identified: samples from the Nizhny Novgorod, Volgograd regions and Novosibirsk form a cluster (C1), while samples from North Caucasus and southern Dagestan form another cluster (C2). Samples from Moscow and surrounding areas form a third cluster (C3). Although the patterns of genetic similarity between chloroplast, mitochondrial and nuclear genomes vary, three large clusters are observed in all dendrograms, with two clusters mainly composed of central region trees and one more homogeneous group containing Sochi, Dagestan and the North Caucasus Federal District trees (Fig. 1).

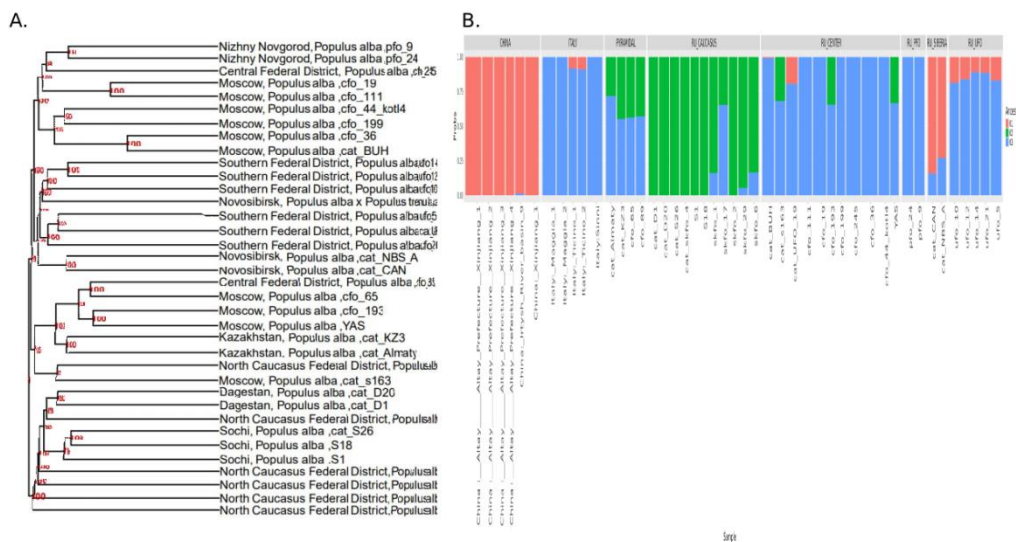


Fig. 1. A, Phylogenetic analysis of white poplar trees growing in Russia and Kazakhstan, with some samples omitted. “Long branches” on the dendrogram are a consequence of inaccuracies in the determination of VAF values, which are calculated by the number of reads supporting the reference and alternative allele. B, Evaluation of a mix of conditional populations for poplars from different regions, with some additional samples also omitted

The population mixing analysis (NGSAdmix) was based on the hypothesis that there were three main centers that could have served as the basis for the formation of the modern white poplar population in the south of the European part of Russia during the post-glacial period. These centers include: Southern Europe, the Caucasus region, and the southern parts of Western Siberia and the Southern Urals. During the period of maximum glaciation, forest vegetation was preserved on these territories. The genetic material from Southern Europe predominates in the genomes of Central European populations and in forest areas of Russia. In steppes of the Volgograd region, up to one-quarter of the genome of the studied samples contains genetic material from South Siberia. It should be noted that a small amount of genetic material from Western Siberia is present in Central European populations as well as genetic material from southern Europe found in populations in Western Siberia, which is quite natural, given the continuous distribution of species during the pre-glacial period. The Caucasian genetic material completely forms the genome of plants in populations of Transcaucasia and the North Caucasus. Most samples contain Southern European genetic material in their genomes. The presence of this genetic material can be explained by historical links between Europe and the Caucasus through Asia Minor, where white poplar still grows today, as well as migration of its carriers from the north to the south via the Don and Kuban River valleys. An interesting feature is the complete absence of South Siberian genetic material from Caucasian populations, and vice versa. There appears to be a gap in the distribution of white poplar since it originated in eastern Turkmenistan, north-eastern Iran, lowland Uzbekistan and north-western Afghanistan, which serves as a barrier to exchange of genetic material between those regions. The exact distribution of white poplar in Central Asia remains debatable due to lack of floral information and the long history of its cultivation in the region.

The genetic composition of plant populations in the central European part of Russia requires separate consideration. When analyzing plant samples, traces of two of the three centers can be found here. Southern European genetic material comes from both Southern Russian and Central European populations. The role of the latter seems to increase near large cities and in Northwestern regions, where it is associated with the period of park introduction in the late 19th–early 20th centuries. Caucasian genetic material, on the other hand, is associated with mid-20th-century introduction, when botanical gardens and breeding centers began expanding their collections with material from all over the country. Pyramidal forms of white poplar in the central part of Europe are of particular interest. It is believed that these specimens were derived from Central Asian pyramidal types. However, as the analysis has shown, only southern European and Caucasian genetic material is present in their composition. Due to the limited amount of information available, it is not possible to provide a definitive explanation for this finding. Therefore, a more detailed analysis of the genome sequences of samples from pyramidal white poplar forms in southern Europe and central Asia would be necessary.

Conclusion: The modern distribution of white poplar in southern European Russia was shaped during the postglacial period through migration from populations in the Balkan Peninsula, eastern and western Siberia, and southern Urals. Caucasus populations of white poplars did not contribute to settlement in this area during historical times. Modern cultivated white poplars in European Russia include plants of diverse origins, such as Central Europe, South Russia, and the Caucasus.

Funding: The study is supported by the Russian Science Foundation (Grant No. 22-14-004).

Список литературы/References

1. Li H., Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760. doi 10.1093/bioinformatics/btp324
2. Skotte L., Korneliussen T.S. et al. Estimating individual admixture proportions from next generation sequencing data. *Genetics*. 2013;195(3):693-702. doi 10.1534/genetics.113.154138