# Final report of the FK-127938 program (2018-2023)

**Program title: Human population genetic research of the present-day Carpathian basin**

**PI: Anna Szécsényi-Nagy**

**Host institution: Research Centre for the Humanities (Inst. of Archaeology, Inst. of Archaeogenomics)**

**Index:**



## Initial objectives

Five years ago, the project sought to fill the void in genetic research on modern Hungarian populations by establishing mitochondrial genome databases based on well-considered demographic sampling. We aimed to provide insights into recent genetic diversity and a reference for archaeogenetic studies. We hypothesized that by selectively sampling from certain villages and tracing the donors' ancestry, we could delineate the 19th-century Carpathian Basin's gene pool and discount the last century's migratory movements.

We explored the under-researched genetic events between the Conquest period and today, assuming our new modern dataset would more closely reflect the early modern era than previous collections, thus enhancing accuracy of historical genetic research.

Our inquiries addressed the genetic homogeneity or structure among present-day Hungarian-speaking groups and whether a regional genetic framework exists within the Carpathian Basin, using mtDNA, Y chromosome, and for the first time, ancestry informative autosomal markers.

Another crucial issue is the lack of study on population genetic events from the post-Conquest period (10-11th centuries) to the present, with comparisons between these times yielding only broad approximations. The project anticipated that the new modern dataset would align more closely with the early modern period than with prior findings on average Hungarian populations, offering a more accurate reference for future historical genetic research. We were especially interested in such

ancestral uniparental lineages of the current population that are informative to certain population historical events and that are more likely to survive in culturally isolated populations. We tested these hypotheses in rural populations of the Hungarian-speaking minorities in three distinct areas (Zobor region, Baranja and Transylvania) strengthened by the genetic analyses of a medieval Székely population in Transylvania.

A pivotal question was confirming previously detected Asian lineages in the present-day Hungarian-speaking Székely population in Transylvania, which we also approached by analyzing medieval DNA from sites near to Székelyudvarhely/Odorheiu Secuiesc. We were interested in detecting or rejecting the hypotheses of local continuity in the Székelyudvarhely area during the medieval and modern times. The outcomes of these research questions, which have now been largely addressed, will be detailed later in this report.

## Participants of the project

During the course of the project, there were minor alterations in the team compared to what was outlined in the original grant application; however, the core group of senior researchers remained constant. The hands-on laboratory work was undertaken by newcomers to the project, such as Orsolya Székely, Motahareh Amjadi, followed by Noémi Borbély, Koppány Kerestély, and Robert Selan. Orsolya Székely departed from the team upon the completion of her master's thesis, and Veronika Csáky took maternity leave partway through the project. The changes were all reported to the office.

The sampling was supported with suggestions by Károly Lábadi and the Democratic Union of Hungarians of Croatia in Baranja, Anna Sándor linguist in the Zobor region and by Elek Benkő archaeologist in Transylvania as planned in the original grant applications. Attila Paládi-Kovács and David Reich have not contributed as consultant or supervisor.

From the Research Centre of the Humanities, Institute of Archaeology several team members helped the project, who were not listed in the NKFIH-EPR system: Bea Szeifert, Dániel Gerber, and the laboratory assistants.

## Milestones of the project and applied modifications on the work plan

The original host institute of the project was the Institute of Archaeology in the Research Centre for the Humanities (RCH). Since the Institute of Archaeogenomics was established in March of 2021, and the PI became its director, it took over the project from the 3rd project year onward. In the following we summarize the major events of the project.

### 1. project year

This year, we successfully executed the preparatory administrative tasks required for DNA sampling and collected samples from the Odorheiu Secuiesc/Székelyudvarhely area. Leveraging the

support of local mayors and priests, we gathered samples from 115 individuals with ancestral ties to the microregion.

Orsolya Székely, our MSc student, along with the laboratory team and the Principal Investigator (PI), optimized laboratory protocols and methods, and then processed the mitochondrial genomes of these contemporary Székely individuals at the RCH laboratory.

Concurrently, the team also conducted Y-chromosomal short tandem repeat (STR) analysis on the Székely samples, advancing a task scheduled for the project's final year. We processed 92 male samples at the Laboratory of Reference Samples Analysis of the Department of Genetics, Directorate of Forensic Expertise, Hungarian Institute for Forensic Sciences in Budapest, directed by Horolma Pamjav.

In September 2019, the PI, Bea Szeifert, and Balázs G. Mende visited the Croatian Baranja to collect DNA samples from the Hungarian minority, adhering to the initial plan.

The PI secured a New National Excellence Program (Bolyai +) grant at ELTE University, which complements project Nr. 127938 through mitochondrial DNA analysis of the Székely population. The Genetics Department at ELTE University, in conjunction with the Bolyai + program, has appointed the PI to voluntarily supervise and instruct new biology students, who also contribute to the 127938 research project.

**In summary, our achievements in the first year have surpassed the expectations set in the original work plan.**

**2. project year**

**The project's progress this year was significantly impacted by the COVID-19 pandemic.**

The planned sampling in the Zobor region was abruptly canceled in March 2020 due to the epidemic and resulting lockdown. A limited remote DNA sampling effort took place in the summer in the Zobor region of Slovakia, organized by Veronika Csáky with local mayors' help, though fewer samples were collected than originally intended.

The collection of ancient samples from the Székelyudvarhely area remained incomplete amidst ongoing uncertainty caused by the pandemic.

Orsolya Székely and Noémi Borbély conducted mitochondrial genetic analyses on the modern samples. Orsolya Székely completed her MSc thesis on the full mitochondrial DNA series in Szeklerland, which was submitted in spring 2020 at Eötvös Loránd University's Department of Genetics and presented at an international conference online (EAA). Alongside mtDNA analyses, we initiated quality testing for whole genome shotgun sequencing, which proved more informative and cost-effective than the initially planned Affymetrix Human Origin Panel. Inclusion of private SNP databases (as planned in the original project proposal) became unnecessary in the light of the new genome sequencing strategy.

The PI was awarded the ÚNKP Bolyai Plus grant in 2019, which correlates with the NKFI program through mitochondrial DNA studies of the contemporary Szekler population. She presented the initial findings at the ELTE-ÚNKP conference.

**3. project year**

We acquired 30 medieval (12th-15th century) DNA samples from the István Molnár Museum collection in Cristuru Secuiesc/Székelykeresztúr. These samples underwent processing to enrich mitochondrial DNA (mtDNA). We planned further ancient sampling in Transylvania for autumn 2021, aiming to access the Haáz Rezső Museum's collection in Odorheiu Secuiesc/Székelyudvarhely.

Noémi Borbély and the PI evaluated the data and analyzed uniparental genetic markers. Noémi B. also processed additional samples from the Drávaszög population in Croatian Baranja and prepared DNA libraries for whole-genome sequencing. We made significant advancements in optimizing ancient Y-STR (Y-chromosomal short tandem repeat) analyses this year.

The findings were presented at an international conference, and due to delays caused by the pandemic, we requested an extension by the NRDI into a fifth project year.

Compared to the original work plan, due to technical advancements, and the rapid development of our research filed, **from this time on we planned one mtDNA+Y-chromosomal paper about the Székely population, one mtDNA+Y-chromosomal paper about the Zobor region and Baranja groups and one study with Székely ancient DNA focus**, including the full genome analyses as well. We reported to the NRDI office about these modifications on the work plan.

**4. project year**

In the fourth year of the project, we adhered to the adjusted work plan and pursued the scientific objectives outlined in the third year's report. We collected samples from 65 additional medieval-early modern era individuals from cemeteries, housed at the Haáz Rezső Museum in Székelyudvarhely/Odorheiu Secuiesc, which the Institute of Archaeogenomics' laboratory team then processed. Following quality screening, the highest quality samples underwent deep sequencing, resulting in 18 modern and 24 ancient genomes.

Y-STR data were generated from ancient samples, and mitochondrial capture was conducted on those not subjected to whole-genome sequencing. Analysis of whole-genome data from contemporary Székelys and a selection of medieval Székelys took place in the latter half of the year. **With those points we outperformed the original project plan that considered only the mtDNA analyses of the medieval Székelys**.

Results were disseminated through five oral and poster presentations at international and national conferences. In addition to laboratory efforts, the fourth year was heavily dedicated to manuscript preparation.

**5. (extended) project year**

Our first paper detailing uniparental data of the current Székely population was published by the MDPI journal Genes in new year of 2023, titled as “High Coverage Mitogenomes and Y-Chromosomal Typing Reveal Ancient Lineages in the Modern Day Székely Population in Romania”.

After the publication, both the PI and Noémi Brobély advertised the results in television and journal interviews and short video communications (see the list of public outreach activities).

This extended project year was used to present the results at three national and international conferences, and work on further manuscripts.

The second manuscript, focusing on the uniparental results of the populations from Drávaszög and the Zobor region, was underway. Horolma Pamjav's team, in collaboration with Noémi Borbély, was responsible for the Y-chromosomal evaluations, while Noémi Borbély also undertook the analysis of the mitochondrial DNA findings with the PI. This paper has been submitted to the journal 'Human Molecular Genetics' published by Oxford Academics in the last quarter of the fifth project year.

Koppány Kerestély prepared an MSc thesis on the ancient Székely paternal genetic results that will be submitted at the ELTE University Genetic Department in December 2023.

The PI and Noémi Borbély worked further on the full genome analyses of ancient and modern samples and prepared a third manuscript from the results.

The team organizes a conference on the 12$^{th}$ of December 2023 with external guests, where the key achievements of the projects are going to be presented (see reference to the conference in the Dissemination chapter).

**With those milestones the modified work plan of the project has been accomplished.**

## Results and outcome

By the end of the third project year, a unique human DNA sample set had been established in Hungary that had not existed before. These DNA sample sets formed the foundation of this research and with the publications of the team became available for later analyses beyond the current financial framework's scope.

Through this project, the domestic and international research community gained access to a total of **283 new, well-documented mitogenomic sequences, genome-wide SNP genotypes, and Y chromosomal profiles of approximately 214 individuals** for further comparative phylogenetic analyses.

In the early fourth year, the Székely uniparental dataset was **published in a peer-reviewed scientific journal** (Borbély et al. 2023a). The data became the subject of two MSc (Orsolya Székely and Koppány Kerestély) theses and one PhD thesis (Noémi Borbély) with the supervision of the PI and Balázs Egyed senior participant of the project.

By the fifth year, the existing public Y chromosomal database of the present-day population in the Carpathian basin was extended and reevaluated. The findings together with the mitochondrial data were also presented at several conferences and **submitted to a Q1 journal in genetics** (Oxford Journals: Human Molecular Genetics, the Borbély et al. 2023b manuscript is available as preprint).

**We generated 18 shotgun genomes** from present-day individuals, whose dataset were not numerous enough to publish them separately. Therefore, we included them in the ancient DNA study focusing on the Székely medieval groups.

The pilot ancient DNA project concentrated on the Medieval period of Transylvania (Romania). The 13-17th century human DNA data from the Székelyudvarhely area were compiled into a third manuscript (Borbély et al. 2024 in prep.) by the end of the fifth year and incorporated into the doctoral thesis (expected submission year of the dissertation is 2026) of the project's PhD student Noémi Borbély. We originally planned only ca. 60 mtDNA analyses from the ancient Székelys, but we managed to obtain **89 ancient mitogenomes**.

Furthermore, we could afford **Y-STR analyses and shotgun genomes of 58 and 24 samples** respectively from the Tranylvanian ancient sample set. Out of the 58 Y-chromosomal STR typing, 47 were successful and resulted in a minimum of 8 detected Y-STR loci per sample, which were already sufficient for a Y-chromosomal macrohaplogroup definition. We performed detailed phylogenetic analyses of several paternal lineages, comparing ancient and modern-day diversities.

The modern-day uniparental **DNA sequences have been deposited in ENA** (European Nucleotide Archive, https://www.ebi.ac.uk/ena/browser/home) **and the Y-chromosomal haplotypes in YHRD database** (https://yhrd.org/) after careful ethical considerations of data privacy. The uniparental data presented in the modern-day Székely study are openly and anonymously available in the European Nucleotide Archive (ENA), with accession number [PRJEB52529] and at Y-STR Haplotype Reference Database (YHRD), with accession number: [YA00612]. Y-STR Haplotype Reference Database (YHRD) accession numbers for samples from the Baranja region is YA006013 and for samples from the Zobor region: YA006014. The ENA accession number regarding the mitogenome sequences from Baranja and Zobor region is PRJEB64294 (the database will withhold release of data until publication). ENA accession number for the ancient Székely mitogenomes and shotgun genomes is PRJEB70309, the data will be available upon publication.

Data management policy of the project: for sampling, handling and storage of personal data and genetic samples of living persons, we considered the Hungarian 2008/XXI. law as guidelines. The Hungarian 2011/CXII. law provided us with rules about the information and self-determination rights of the sample providers. The modern DNA sequences generated and the results of this project are stored anonymously in the archived collection of the IAG RCH. The custodian ensures the lawfulness and security of the processing of personal data in accordance with the requirements of the applicable data protection legislation (General Data Protection Regulation 2016/679/EU - General Data Protection Regulation 2016/679/EU) and the Ethical Codex of the RCH. Only ancestry informative SNP panel genotypes (e.g. Human Origin Panel) gained from the modern samples will be shared with other researchers, upon request (by submitting a dated, signed letter containing specific commitments on data handling).

## Discussion of the results

We outperformed our original research plan with 119 modern-day Y-chromosomal analyses, and we produced additional 47 Y-STR (short tandem repeat) haplotypes and 24 shotgun genomes from the ancient Székely samples. Furthermore, we have surpassed our initial target by producing 29 additional ancient mitogenomes from the Székelyudvarhely area.

Reflecting to our original research questions, we can make the following statements about the populations of the modern-day and ancient Carpathian Basin:

In the studied three regions we collected samples from local people in remote villages (Figure 1). These samples resulted occasionally in archaic uniparental lineages that could be followed back to early Hungarians or to the precursor autochthon people of the Carpathian Basin. We presented and proved this through detailed phylogenetic analyses of both the mtDNA and Y-chromosome data (Borbély et al. 2023a, 2023b). In that sense the carefully planned sampling was successful and met our expectations.

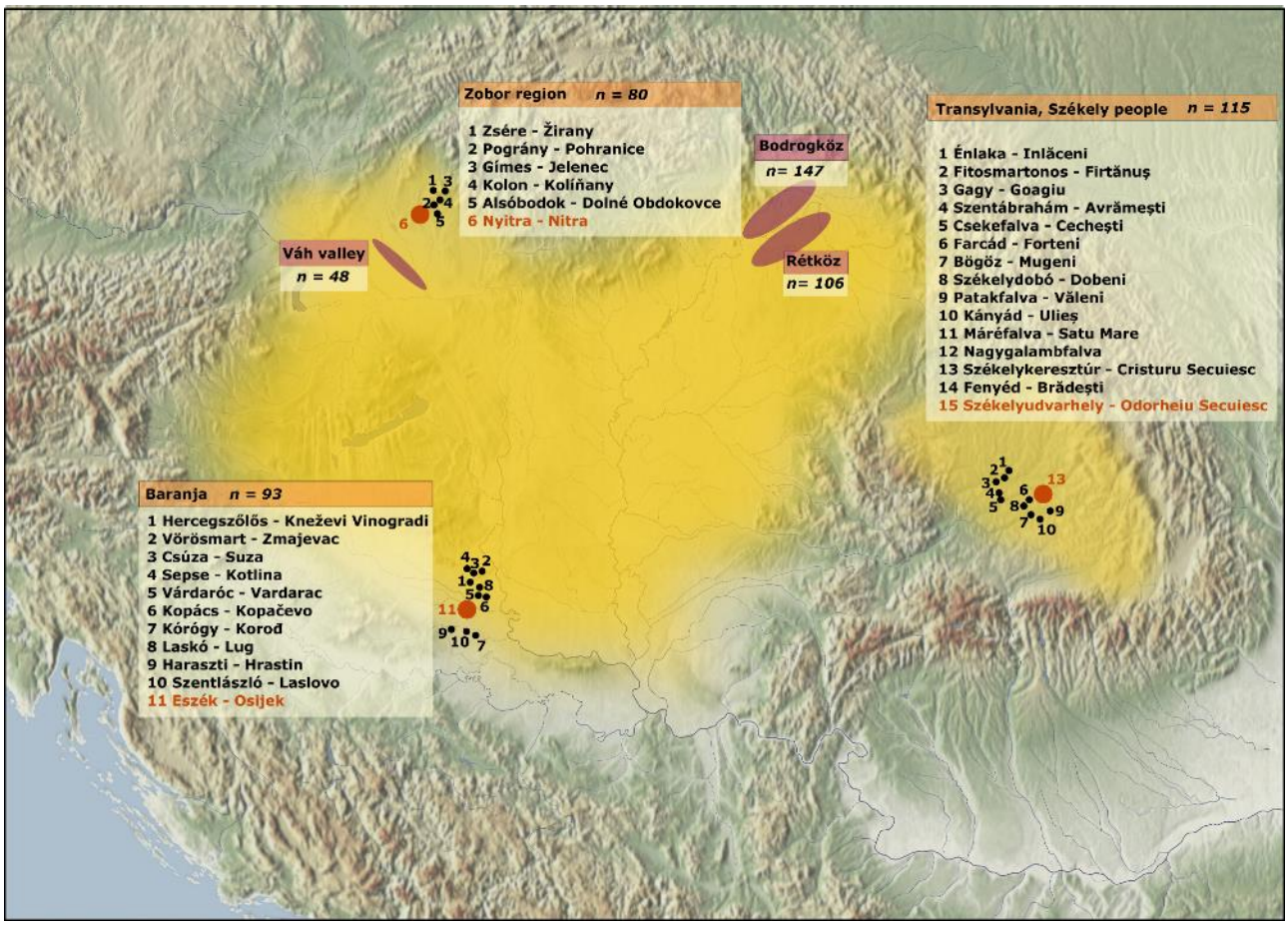


**Figure 1.** Map of modern-day sample collection in the frame of this project.
The geographical origin (villages as places of residence) of non-related Hungarian sample donors from Transylvania (Romania), Baranja (Croatia) and Zobor region (Slovakia). Red town names are indicated as reference points, no samples were collected from them. From Y-chromosomal research aspects, the previously

published Váh valley, Rétköz (Pamjav et al. 2022) and Bodrogköz areas (Pamjav et al. 2017) are also indicated on the map. Yellow shades roughly indicate the regions where the Hungarian language is spoken today.

By integrating the newly acquired genetic data with pre-existing Eurasian and ancient DNA datasets, our goal was to enrich the understanding of the genetic history trajectories of Carpathian Basin populations. We proved that the uniparental genetic structure of the three analyzed regions largely corresponded to the surrounding areas' populations and even without available close comparative datasets from the surrounding population we testified that the Hungarians and the Hungarian-speaking minorities in the surrounding countries have a similar uniparental makeup to other East-Central European populations, due to long-time admixture events and shared common history.

Phylogenetic analyses however shed light on specific paternal and maternal lineages in the gene pool of the three regions that can be traced back through ancient reference data to specific origins. Analysis of Y-STR data and mitogenome sequences did uncover multiple lineage ties to far-flung regions and eras. While the predominant portions of both paternal and maternal DNA align with the East-Central European spectrum, rarer subhaplogroups and lineages have unveiled ancient ties to both prehistoric and historic populations spanning Europe and Eastern Eurasia (Borbély et al. 2023a, 2023b). An example of such is the U5a maternal haplogroup, which is prevalent across Western Eurasia, and also well-represented in the modern Carpathian Basin. Notably, its U5a2 subclade establishes a clear link with ancient samples from the closer and wider region, with important examples from the 9-11th century cemeteries of ancient Hungarians as well (see Fig. 9, Fig. S16 in Borbély et al. 2023b). The Y-chromosomal phylogenetic analyses showed that Hungarian-speaking males from Zobor region and Baranja share certain common haplotypes with ancient Xiongnu, ancient Avar, and Caucasian males within haplogroups R1a-Z93, G2-L156, and R1b-L23, suggesting a minor common genetic footprint. On the other hand, subgroups like R1a-Z280 and N1a1-Tat connects Hungarians from Hungary and the newly investigated populations to the Ural region (see Fig. 4, FigS6 in Borbély et al. 2023b).

Combined Y-STR haplotype and mitogenome sequence analyses respectively (using median joining method) showed the mixed distributions of the uniparental lineages among the sampled villages, as a natural consequence of marriage alliances among village communities. Comparing the three regions, we discovered subtle genetic differences (like some haplogroups were missing or more frequent at one than at the other areas), but in general terms we could not detect marking genetic structure of the uniparental makeup within the Carpathian Basin (see Figure 2 as an illustration for the Y-STR pattern).

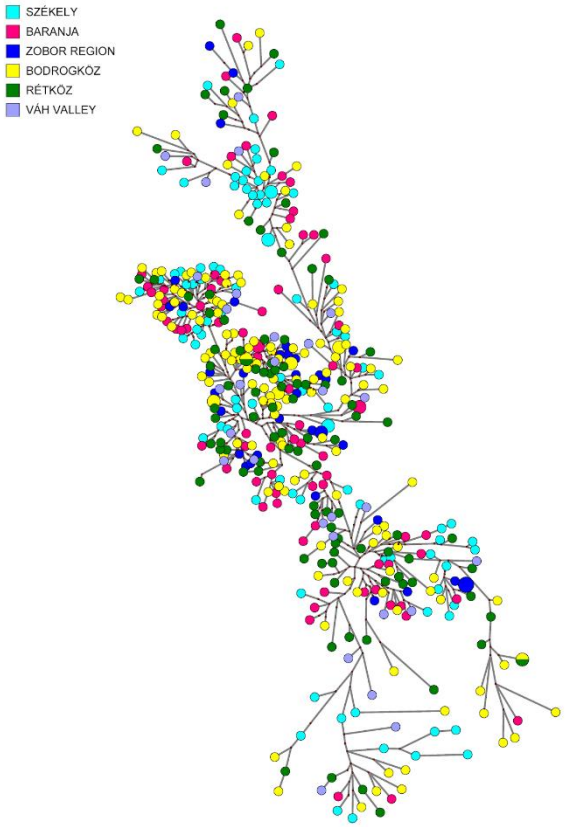


**Figure 2.** Summarized Median-Joining networks of Hungarian-speaking male populations. Median-Joining Networks of all Hungarian-speaking male populations analyzed from Baranja, Zobor region, Transylvania, Bodrogköz, Rétköz and Váh valley based on 21 Y-STR data (Pamjav et al. 2017; 2022; Borbély et al. 2023a, 2023b), colored by regions. The haplotypes arranged naturally (by their Y-STR pattern) into Y-chromosomal haplogroups, as it is seen on Fig.S4 in Borbély et al. 2023b.

Our fourth research question was the verification of the Asian lineages in the Székely population. Whereas the typical but rare haplogroups (such as Y-N1a) expected from previous studies of the population from Csíkszereda/Mercurea Ciuc were not reproduced in the Székelyudvarhely area, other signs of ancient Eastern origin emerged from both maternal and paternal sides in the new dataset. However the frequency of these remained low in the dataset (ca. 6-7% for the paternal lines (such as Q and R1a-Z93) and 4% for the maternal lines (such as A, C, D haplogroups)). These novelties are analyzed in detail in the manuscripts and paper of the team. While medieval Székelys and previously analyzed Hungarians from the 10th-11th centuries (Neparáczki et al. 2019, Maróti et al. 2022) share common haplogroups, the tendencies in their frequencies diverge, which can be a consequence of sampling strategy of the previous studies with focus on the Great Hungarian Plain.

The study of the medieval Székely population is still in manuscript form (Borbély et al. 2024 in prep), therefore we discuss these results in detail.

The ancient Székelys showed a strong connection system to the present-day inhabitants of the area, and these were not only confirmed (and explained in detail in the manuscript) by uniparental analyses, but supplemented by full-genome analyses as well.

We present here the genome-wide SNP data called from the shotgun genomes on principal component analysis (PCA) first, that shows the modern-day European and Near-Eastern genetic variation (Figure 3). The historic and modern Székely datasets are closely aligned in the principal components 1 and 2 space, slightly extending towards the southeastern populations, which are representative of the present-day Caucasian groups, in comparison to the Hungarian sample set. Modern-day Székelys exhibit less genetic variability than their ancient counterparts at this level.

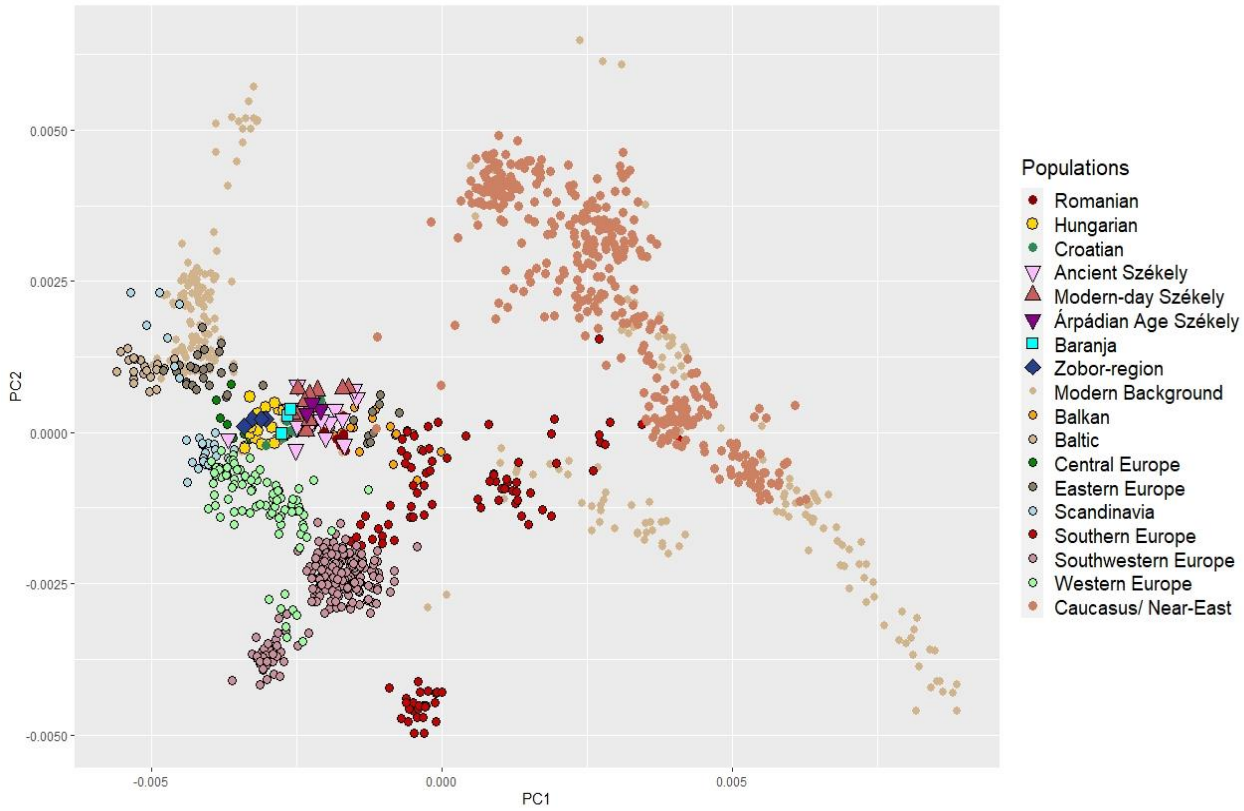


**Figure 3.** PCA plot with modern-day genotypes from the European Human Origin panel. Cca. 560k SNPs were used for the calculations, performed in the smartpca program (https://github.com/DReichLab/EIG). The scatterplot was generated in R. New Hungarian genotypes are highlighted in the plot.

The Székelys are also positioned at the confluence of the genetic currents identified from the 9th to 12th centuries in the Carpathian Basin (Csáky&Gerber et al. in prep). Their archaic genetic composition was preserved through the centuries. Within the Székelys, there is a 2-7% detectable Eastern Eurasian genetic element, which also reflects the conservative, up to present-day survival of some ancient uniparental markers. This eastern element has almost vanished in other modern Hungarians and also diminished in the Székely local population through the last 3-6 centuries to as low as 1-3%. Since the component is very low, and based on uniparental phylogenetic analyses it most probably come from multiple sources (e.g. Avars, Conquest period Hungarians among others), its origin is not traceable with traditional population genetic methods (such as f statistics).

A new bioinformatics development of the Institute of Archaeogenomics, using the infrastructure of the Wigner DC (ELKH and HUN-REN Cloud), enabled us to analyze long shared haplotype segments of the autosomes (Ringbauer et al. 2023). Analyses of such identical by descent (IBD) segments could be performed on the best quality 15 ancient Székely genomes, where at least 380k

SNP could be called from the 1240k panel. We observed, that they were not close relatives of each other, having the largest sum IBD only 57 cM and the largest chunk of 47 cM shared. The IBD analyses revealed that the ancient and present-day Székely communities are strongly interconnected within the historic space, represented by a graph (Figure 4). Most of the connections are among the ancient and modern-day Székely, forming one cluster. They are both connected to the 10-12th century population of the Carpathian Basin. Only some of them are detached from the Székely cluster, and one of such individuals is from a 15-16th century grave from the Jesus chapel of Odorheiu Secuiesc, who interestingly connects an Inner Asian genome (DA41, Damgaard et al. 2018) with other Székelys. Other than that, the Asian connections are rather sparse. On the presented graph the Xinjiang region and the Central Steppe are clearly separated and the Avar period population of the Carpathian Basin forms a second cluster with East-Asian and Central Asian Iron Age and early Medieval individuals. In general, the archaeologically well characterized Hun period samples (like Budapest Vezér street, Árpás Dombiföld) are rather connected to the Central and Inner Asian Iron Age, and some of them to the Early-Middle Avar period.

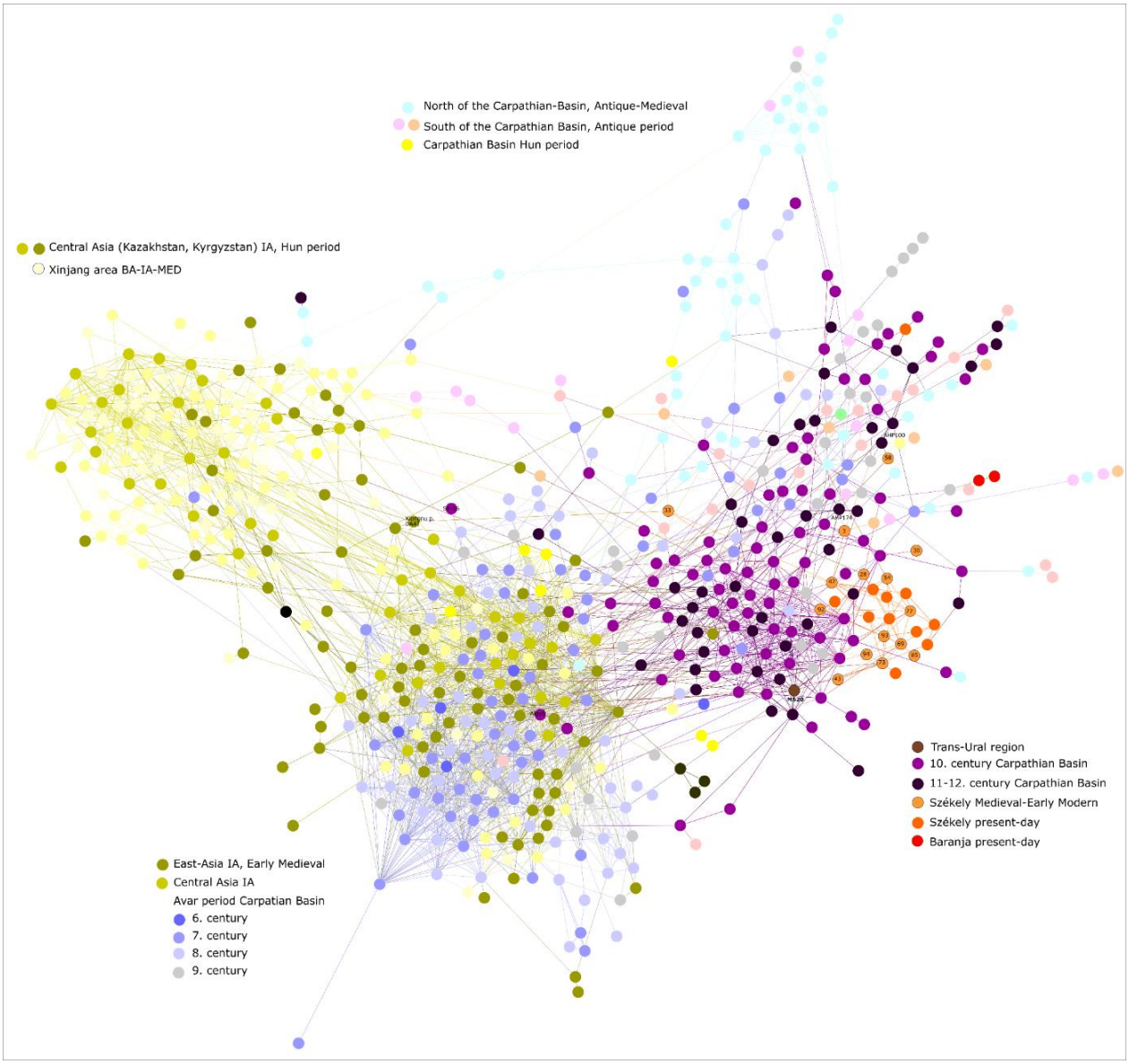


**Figure 4.** ancIBD analyses of the ancient and modern genomes produced for Borbély et al. 2024 (in prep). The graph was created in Gephi v.0.10.1, using ForceAtlas 2 layout and minimum 1x12cM long shared IBD segments. Colors indicate different regions and chronological phases. Numbers in the circles indicate laboratory numbers of the analyzed ancient Székely genomes.

However, these loose connections can also be understood as indirect connections between Avars and Huns via common Central Asian sources. Direct connections of the Székelys to the known Hun period genomes are not observable, however it should be noted that 12cM sharing get diluted relatively quickly over time due to recombination events in each generation, therefore it is quite unlikely to be found between pairs of individuals separated by more than 1000 years. An extreme example however for the survival of long stretches over a millennium is the case of MS20/MOT55 sample (kurgan 11/5) from the 8-10th century Trans-Uralic site Uyelgi near Chelyabinsk (Csáky et al. 2020b), who shows numerous connections on the graph toward Hungarian conquerors, but also one segment to each of two medieval Székelys (both from the 15th century phase of cemetery Patakfalva-Papdomb) and even to one modern-day Székely individual living in the rural area next to Székelyudvarhely.

Considering the geographic pattern of these IBD connections in an ancient dataframe, relations to both Transdanubia and Great Hungarian Plain are seen (Figure 5), but the Transdanubian connections are in the majority (archaeological sites like Szakony, Vörs-Papkert, Sárbogárd-Tringertanya, Zalavár, Székesfehérvár and Visegrád). The number of 10-12th century samples are in almost equal ratio from the two larger areas, but the 2nd half of the 9th century is better represented in Transdanubia (Csáky&Gerber et al. in prep).

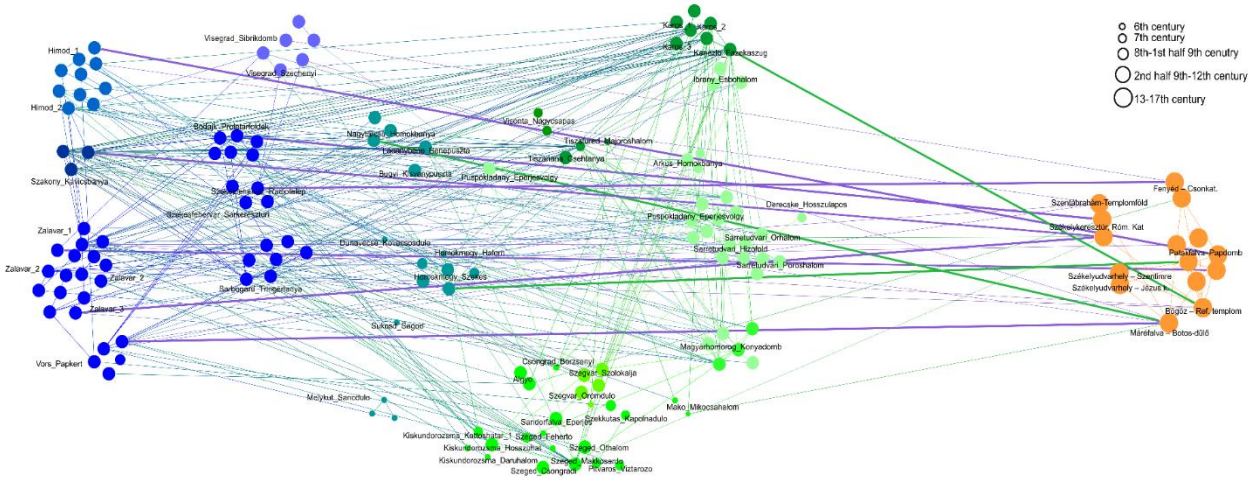


**Figure 5.** IBD analyses of the ancient Carpathian-Basin 7th-12th genomes published previously (Csáky&Gerber et al. in prep, Maróti et al. 2022, Gnecchi et al. 2022) and 13-17th century Székely genomes produced for Borbély et al. 2024 (in prep). The graph was created in Gephi v.0.10.1, with minimum 1x10cM long shared IBD segments and using an approximate geographic layout. Thicker lines represent at least 12 cM connections between the ancient Székelys and the 9th-12th century population of the Carpathian Basin. Colors indicate different regions and sites (bluish is Transdanubia, greenish is Great Hungarian Plain and orange is the historic Székely dataset).

This Borbély et al. 2024 (in prep.) manuscript presents the first dataset of ancient Székely genetics, spanning from the Árpádian Era to the Early Modern period. **The data are consistent, showing that the two Székely populations are remarkably similar in the known genetic landscape.** Nevertheless, the archaic analyses were necessary and proved to be useful. The comparison of the two time periods demonstrates which lineages have disappeared or been introduced into the population over time, and which have been consistently present over the past 300-700 years. **The whole-genome datasets proved to be extremely valuable, as they provided stronger evidence for the connection between the two time periods than the previous uniparental data, and shed light on the relationship between the medieval Székelys and the 10th-12th century Árpádian Era** data from the Carpathian Basin. Although the number of usable samples from the 10th-12th centuries of the Carpathian Basin has greatly increased in recent years, more complete cemeteries or series compiled through extensive random sampling are needed to definitively determine the sources of the ancient Székelys' settlement.

It will be equally important in the future to delve further back in time in Transylvania, examining periods prior to early Árpádian Era to uncover the genetic composition of the local population during the era of the great migrations. And amidst all these considerations, we must not forget that, according

to our current knowledge, modern Székelys exhibit a certain genetic heterogeneity, which possibly has antecedents in the past.

## Significance of the results

This research project has both narrower scientific and wider social impacts: it allows the interpretation of the genetic composition of today's Hungarians in a geographical and historical context. Thus, the results open up new opportunities not only for the research of present-day populations, but also for the interpretation of the genetic heritage of ancient populations.

This project, which focused on minorities whose numbers are dwindling in their native regions, was committed to the genetic conservation and documentation of these groups. The project significantly increased the number of sequenced and typed maternal and paternal lineages in the region. By obtaining samples from the eldest members of these communities and tracing their ancestral lineage, the team aimed to bridge the gap between the past and the present, providing a clearer understanding of these groups' historical journey to the wider public.

The research offered a wider social viewpoint by facilitating the understanding of the genetic makeup of contemporary Hungarians in a geographical and historical framework. It is important to note, as expressly mentioned in the proposal and upheld throughout the project, that defining or confirming a 'genetic profile' of Hungarian ethnicity was not an objective. The team deliberately steered clear of such determinations in their presentations and communications due to the potential harmful implications.

The origin of the Székelys has long been a matter of debate among historians and archaeologists, and it also attracts significant public interest. The formulating manuscript Borbély et al. 2024 (in prep) demonstrates that in the medieval Székely population, there are occasional eastern uniparental descent lines characteristic of the Hun and Avar eras. However, the majority of them have stronger local roots, and some are distinctly linked to the conquering Hungarian population.

For the first time, we examined the sharing of less conservative, autosomal haplotypes (IBD) in the history of genetic research on the Székelys. The findings clearly indicate continuity around the Odorheiu/Székelyudvarhely area, but they also extend beyond that. The study of the geographic arrangement of connections reveals strong Transdanubian links, aligning with some of the historical and linguistic hypotheses about the Székelys' origin, which genetic results however need further verification by additional data.

Over the course of the five-year project, the Laboratory of Archaeogenetics and the later-established Institute of Archaeogenomics underwent significant transformations. At the time the project was proposed, the Principal Investigator had not yet worked with whole genome data, presenting a challenge for the perspective of autosomal DNA analyses and necessitating plans for further methodological training. These challenges have been overcome, and by the time of the third study, the project has adopted and used the state-of-the-art methodologies of the research field. In this regard, the project successfully contributed to the methodological development of the population genetic and archaeogenetic research fields in Hungary.

# Dissemination

## Conference presentations:

Szécsényi-Nagy Anna: **Új uniparentális leszármazási vonalak a Székely populációból- Lehetséges genetikai kapcsolatok a székelyek és a korai magyarok között**, ÚNKP Konferencia ELTE, 2020

Székely Orsolya, Szeifert Bea, Gerber Dániel, Máthé István, Pamjav Horolma, Mende Balázs Gusztáv, Egyed Balázs, Szécsényi-Nagy Anna: **New uniparental lineages revealed in the modern day Székely population -possible genetic connections between Székelys and early Hungarians**, EAA 2020 virtual conference, talk, 26th Annual Meeting of European Association of Archaeologists, 26-30th of August, 2020

Noémi Borbély: **Székelyföld, Zoboralja és Drávaszög mai, magyar anyanyelvű lakosságának genetikai vizsgálata**, talk, Introduction to the researches of the Institute of Archaeogenomics, Conference at the RCH, ELKH, Budapest, Hungary, 30th November 2021

Noémi Borbély, Dániel Gerber, Bea Szeifert, Horolma Pamjav, Balázs Gusztáv Mende, Balázs Egyed, Anna Szécsényi-Nagy: **Towards building the genetic map of the Carpathian Basin**, poster, Hungarian Molecular Life Sciences, Eger, Hungary, 5-7 November 2021

Anna Szécsényi-Nagy, Veronika Csáky, Noémi Borbély, Dániel Gerber, Balázs Gyuris, Kristóf Jakab, Bea Szeifert, Balázs G. Mende: **Genetic formation and structure of the early Hungarian Kingdom's population based on whole genome data**, poster, EMBO/EMBL Symposium "Reconstructing the Human Past" Heidelberg, Germany, 13-16th September, 2022

Noémi Borbély, Bea Szeifert, Koppány Kerestély, Horolma Pamjav, Balázs Egyed, Zsolt Nyárádi, András Sófalvi, Szilárd Sándor Gál, Elek Benkő, Anna Szécsényi-Nagy: **Investigating the archaic and modern-day Székely gene pool around Székelyudvarhely**, talk, 28th Annual Meeting of the European Association of Archaeologists, Budapest, Hungary, 31st August-3rd September 2022

Borbély Noémi, Csáky Veronika, Szeifert Bea, Gerber Dániel, Gyuris Balázs, Jakab Kristóf, Mende Balázs Gusztáv, Szécsényi-Nagy Anna: **Genetikai képek a Kárpát-medencéből: teljes genom elemzések a 9-12. századi Dunántúlról és az Árpád-kori székelység köréből**, talk, XXI. "Genetic Workshops in Hungary", Szeged, 7th September, 2022

Noémi Borbély, Koppány Kerestély, Elek Benkő, András Sófalvi, Zsolt Nyárádi, Horolma Pamjav, Anna Szécsényi-Nagy: **Genetic study of modern and medieval Székely Land,** poster + flash talk, ELTE Doctoral School of Biology Conference, Budapest, Hungary, 1st December, 2022

Noémi Borbély, Bea Szeifert, Elek Benkő, Dániel Gerber, Kristóf Jakab, Koppány Kerestély, András Sófalvi, Zsolt Nyárádi, Balázs Gusztáv Mende, Anna Szécsényi-Nagy: **Whole-genome data from three Hungarian-speaking minorities of the Carpathian Basin**, Hungarian Molecular Life Sciences, Eger, Hungary, 24th-26th March, 2023

Noémi Borbély, Horolma Pamjav, Balázs Egyed, István Máthé, Anna Szécsényi-Nagy: **Uniparental genetic diversity of three Hungarian-speaking isolated communities in the Carpathian Basin**, Haploid Markers 12, Budapest, 17-20. 15. 2023

Horolma Pamjav, Noémi Borbély, Dániel Dudás, Attila Tapasztó, Eszter Dudás-Boda, Veronika Csáky, Balázs Mende and Anna Szécsényi-Nagy: **Uniparental study of the Hungarian-speaking populations of Drávaszög and Zoboralja**, XXII. "Genetikai Műhelyek Magyarországon" Minikonferencia, Szeged, 2023. 09. 15.

Szécsényi-Nagy Anna, Szeifert Bea, Gyuris Balázs, Mende Balázs Gusztáv, Csáky Veronika, Langó Péter, Türk Attila: **Új genetikai adatok a honfoglalás kori népesség eredetéhez és biológiai kapcsolat rendszeréhez**, Árpád népe. A magyar honfoglalás kor kutatásának legújabb eredményei, Szentendre, 2023.04.13–15.

### MSc Theses:

Székely Orsolya: **A mai székelység genetikai vizsgálata teljes mitogenom szekvenálással.** MSc diplomamunka, biológus mesterszak, Molekuláris genetika, sejt- és fejlődésbiológia szakirány, ELTE TTK Biológiai Intézet, 2020. Témavezetők: Szécsényi-Nagy Anna és Egyed Balázs.

Kerestély Koppány: **Archaikus és recens humán DNS adatsorok összevetése a székely populációban.** MSc diplomamunka, biológus mesterszak, Molekuláris genetika, sejt- és fejlődésbiológia szakirány, ELTE TTK Biológiai Intézet, 2023. Témavezetők: Szécsényi-Nagy Anna és Egyed Balázs.

## Papers:

Borbély et al. 2023a: Noémi Borbély, Orsolya Székely, Bea Szeifert, Dániel Gerber, István Máthé, Elek Benkő, Balázs Gusztáv Mende, Balázs Egyed, Horolma Pamjav, and Anna Szécsényi-Nagy. 2023. "**High Coverage Mitogenomes and Y-Chromosomal Typing Reveal Ancient Lineages in the Modern-Day Székely Population in Romania**" *Genes* 14, no. 1: 133. https://doi.org/10.3390/genes14010133

Borbély et al. 2023b: Noémi Borbély, Dániel Dudás, Attila Tapasztó, Eszter Dudás-Boda, Veronika Csáky, Bea Szeifert, Balázs Gusztáv Mende, Balázs Egyed, Anna Szécsényi-Nagy, Horolma Pamjav. 2023. "**Phylogenetic study of the Hungarian-speaking Baranja (Croatia) and Zobor region (Slovakia) populations**" Preprint available at Research Square [https://doi.org/10.21203/rs.3.rs-3604738/v1], submitted to Human Molecular Genetics (Oxford Journals)

### Manuscript in preparation:

Borbély et al. 2024: Noémi Borbély, Elek Benkő, András Sófalvi, Zsolt Nyárádi, Koppány Kerestély, Kristóf Jakab, Botond Heltai, Szilárd Gál, Balázs Gusztáv Mende, Anna Szécsényi-Nagy: **Genomic analyses of ancient Székely communities in the Odorheiu area**. In prep.

## Conference organization:

Project closing conference entitled as "Kárpát-medence középkori és újkori népességeinek genetikai kutatásai" held on the 12.12.2023, in the HTK Research Centre for the Humanities, Budapest. Detailed program is available on the link.

**Public outreach:**

The Borbély et al. 2023a paper about the Székely population had a significant media impact, that we collected on our webpage.

Interview with Anna Szécsényi-Nagy on M5 TV channel about the ancient Székely lineages.

Interview with Noémi Borbély and Anna Szécsényi-Nagy in Népszava journal: "A népvándorlás korából fennmaradt apai és anyai leszármazási vonalak kutatása a székelyek között" reporter journalist: András Kenessei

Interview of József Makkay with Anna Szécsényi-Nagy Anna in Magyar Nemzet journal: "Székelyföldi „génkoktél": bizonyítható-e a hun–székely rokonság?"

AGI BTK youtube channel: short video summary of the Székely paper

Presentation by Noémi Borbély on the "Apáczai napok", which is a scientific program for students of the Apáczai Csere János Gymnasium.

BBC History, 2023 December: Interview about the genetic analyses of the Székely population (András Kenessei)

**Budapest, 27.11.2023** **Anna Szécsényi-Nagy**