

میراث ژنتیکی گسترش کوچ‌نشینان ترک‌زبان در سراسر اوراسیا^۱

پیشگفتار برگردان

مقاله «میراث ژنتیکی گسترش کوچ‌نشینان ترک‌زبان در سراسر اوراسیا»، نوشته بایازت یونس‌بایف و همکاران، در سال ۲۰۱۵ در مجله «PLOS Genetics» منتشر شده است. یافته‌های مهم آن را می‌توان در سه فقره خلاصه کرد:

۱. هم‌زبانی ترک‌زبانان به‌معنای هم‌تباری ژنتیکی آن‌ها نیست (به‌عبارت دیگر، «ترک‌زبان‌بودن» لزوماً به‌معنای «ترک‌تبار‌بودن» نیست). پژوهشگران با بررسی داده‌های ژنتیکی ۳۷۳ فرد از ۲۲ جمعیت ترک‌زبان دریافتند که بیشتر جمعیت‌های ترک‌زبان امروزی از نظر ژنتیکی به همسایگان جغرافیایی خود، اعم از ترک‌زبان و غیرترک‌زبان، شباهت دارند. این یافته از این فرضیه پشتیبانی می‌کند که گسترش زبان‌های ترکی در بخش بزرگی از اوراسیا، دست‌کم در بسیاری از مناطق، نه از طریق مهاجرت و جایگزینی گسترده جمعیت‌های بومی، بلکه عمدتاً از طریق گسترش زبان در میان جمعیت‌های بومی صورت گرفته است؛ فرایندی که ممکن است با سلطه سیاسی و نظامی گروه‌های کوچ‌نشین ترک‌زبان («سلطه نخبگان») تسهیل شده باشد.

۲. با وجود تفاوت‌های ژنتیکی میان جمعیت‌های ترک‌زبان، آثار مشترکی از خاستگاه آسیای درونی در ژنوم بسیاری از آن‌ها باقی مانده است. پژوهشگران دریافتند که جمعیت‌های ترک‌زبان در غرب اوراسیا، از جمله در قفقاز، آسیای مرکزی، شرق اروپا و آناتولی، در مقایسه با همسایگان غیرترک‌زبان خود، اشتراک ژنتیکی بیشتری با جمعیت‌های جنوب سیری و مغولستان دارند. این یافته از فرضیه «خاستگاه آسیای درونی» (عمدتاً جنوب سیری و مغولستان)، به‌عنوان نخستین حاملان زبان‌های ترکی پشتیبانی می‌کند؛ هرچند به‌تنهایی ثابت نمی‌کند که تمام جمعیت‌های ترک‌زبان امروزی از یک جمعیت واحد سرچشمه گرفته‌اند.

۳. آثار ژنتیکی مشترک با مهاجرت‌های تاریخی ترک‌زبانان هم‌خوانی دارد. برآوردهای ژنتیکی، دوره‌هایی از آمیزش جمعیت‌ها را تقریباً میان سده‌های ۹ و ۱۷ نشان می‌دهد؛ دوره‌ای که با مهاجرت‌های

تاریخی گروه‌های ترک‌زبان و گسترش امپراتوری مغول همپوشانی دارد. با این حال، پژوهشگران احتمال می‌دهند که بخشی از این اشتراک ژنتیکی ریشه در خاستگاه‌های جغرافیایی نزدیک و آمیزش‌های پیشین میان نیاکان جمعیت‌های ترک‌زبان و مغول‌زبان داشته باشد. بنابراین، یافته‌ها نشان می‌دهند که گسترش زبان‌های ترکی با جابه‌جایی‌ها و آمیزش‌های جمعیتی همراه بوده، اما به معنای گسترش یک تبار ژنتیکی یکسان در میان تمام ترک‌زبانان نیست.

چکیده

مردمان ترک‌زبان مجموعه‌ای متنوع از گروه‌های قومی را تشکیل می‌دهند که وجه اشتراک آن‌ها زبان‌های ترکی است. این گروه‌ها در منطقه‌ای وسیع، شامل سیبری، شمال غربی چین، آسیای مرکزی، اروپای شرقی، قفقاز، آناتولی، خاورمیانه و افغانستان پراکنده شده‌اند. در مورد خاستگاه و تاریخ اولیه پراکندگی مردمان ترک‌زبان اختلاف نظر وجود دارد و گمانه‌زنی‌ها در باره سرزمین مادری باستانی آن‌ها، دامنه‌ای از استپ‌های ماورای خزر/کسپین تا منچوریا در شمال شرقی آسیا را در بر می‌گیرد. مطالعات ژنتیکی پیشین، نشانگر ژنتیکی واحد و مشخصی را برای مردمان ترک‌زبان شناسایی نکرده‌اند؛ امری که از مدل «جایگزینی زبانی» (language replacement) در برابر «گسترش جمعیتی» (demic diffusion) به عنوان الگوی بسط و گسترش زبان‌های ترکی پشتیبانی می‌کند. ما در این پژوهش، با تحلیل داده‌های ژنوتیپ با تراکم بالا در سطح کل ژنوم، به بررسی خاستگاه ژنتیکی ۳۷۳ فرد از ۲۲ جمعیت ترک‌زبان پرداختیم که نماینده گستره جغرافیایی کنونی این مردمان هستند. در راستای مدل «تسلط نخبگان» (elite dominance) در گسترش زبان، اکثر مردمان ترک‌زبان مورد مطالعه از نظر ژنتیکی شباهت زیادی به همسایگان جغرافیایی خود دارند. با این حال، مردمان ترک‌زبان غربی (که در سراسر اوراسیای غربی نمونه‌برداری شده‌اند) دارای قطعات کروموزومی طویل زیادی هستند که از نظر نسب/تبار (identical by descent یا IBD) با جمعیت‌های «سیبری جنوبی و مغولستان امروزی» (SSM) یکسان هستند؛ منطقه‌ای که مورخان آن را کانون شکل‌گیری مجموعه‌ای از حکومت‌های اولیه استپی (اعم از ترک‌زبان و غیرترک‌زبان) می‌دانند. اگرچه وجود قطعات IBD مشترک با منطقه SSM ($1cM$) در جمعیت‌های غیرترک‌زبان نیز مشاهده می‌شود، اما مردمان ترک‌زبان درصد بالاتری از این قطعات را نسبت به همسایگان غیرترک‌زبان خود نشان می‌دهند ($p\text{-values} < 0.01$). در نهایت، با استفاده از روش الدر، زمان‌های اختلاط ژنتیکی (تقریباً سده‌های ۹ تا ۱۷) برآورد شد که با دوره مهاجرت‌های مردمان ترک‌زبان در سده‌های ۵ تا ۱۶ همپوشانی دارد. بنابراین، نتایج ما حاکی از اختلاط ژنتیکی تاریخی در میان مردمان ترک‌زبان است و وجود

نسب مشترک متأخر با جمعیت‌های امروزی در منطقه SSM، یکی از خاستگاه‌های احتمالی برای نیاکان کوچ‌نشین ترک‌زبان و مغولان خویشاوند آن‌ها را تأیید می‌کند.

خلاصه نویسنده

سده‌ها مهاجرت کوچ‌نشینی در نهایت منجر به توزیع زبان‌های ترکی در منطقه‌ای وسیع از سیبری، سراسر آسیای مرکزی تا اروپای شرقی و خاورمیانه شده است. با وجود تأثیر عمیق فرهنگی که این مردمان کوچ‌نشین بر جای گذاشته‌اند، اطلاعات کمی در مورد ریشه‌های پیشاتاریخی آن‌ها وجود دارد. علاوه بر این، از آنجا که گویشوران ترک‌زبان معاصر از نظر ژنتیکی شبیه همسایگان جغرافیایی خود هستند، مشخص نیست که آیا نیاکان کوچ‌نشین آن‌ها رد ژنتیکی قابل شناسایی از خود به جا گذاشته‌اند یا خیر. در این مطالعه، ما نشان می‌دهیم که مردمان ترک‌زبان نمونه‌برداری شده در سراسر خاورمیانه، قفقاز، اروپای شرقی و آسیای مرکزی نسبت‌های متفاوتی از اجداد آسیایی را که از یک منطقه واحد، جنوب سیبری و مغولستان سرچشمه می‌گیرند، به اشتراک می‌گذارند. جمعیت‌های مغول‌زبان و ترک‌زبان این منطقه تعداد غیرمعمول زیادی از قطعات کروموزومی طولانی دارند که از نظر نسب با مردمان ترک‌زبان سراسر غرب اوراسیا یکسان هستند. تجزیه عدم تعادل پیوستگی ناشی از اختلاط در کروموزوم‌ها در این جمعیت‌ها نشان می‌دهد که اختلاط در طول سده‌های ۹ تا ۱۷ رخ داده است، که با مهاجرت‌های کوچ‌نشینان ترک‌زبان که از نظر تاریخی ثبت شده‌اند و گسترش مغول‌ها در آینده مطابقت دارد. بنابراین، یافته‌های ما ردپای ژنتیکی مهاجرت‌های کوچ‌نشینی اخیر در مقیاس بزرگ را آشکار می‌کند و منشای آنها را به منطقه‌ای که قبلاً فرض شده بود، یعنی مغولستان و جنوب سیبری، نشان می‌دهد.

مقدمه

ارتباط زبانی اغلب برای اطلاع‌رسانی در مطالعات ژنتیکی استفاده می‌شود (۱) و در اینجا ما این مسیر را برای بازسازی جنبه‌هایی از یک رویداد جمعیتی بزرگ و نسبتاً جدید، یعنی گسترش مردمان ترک‌زبان کوچ‌نشین، که بخش عمده‌ای از چشم‌انداز قومی-زبانی اوراسیای غربی را در دو هزاره گذشته تغییر شکل دادند، در پیش می‌گیریم. جمعیت‌های ترک‌زبان مدرن عمدتاً مردمی ساکن هستند؛ تعداد آنها در سراسر اوراسیا بیش از ۱۷۰ میلیون نفر است و پس از یک دوره مهاجرت – که تقریباً از سده ۵ تا ۱۶ م را در بر می‌گیرد – پراکندگی جغرافیایی گسترده‌ای دارند که شامل اروپای شرقی،

خاورمیانه، قفقاز شمالی، آسیای مرکزی، سیبری جنوبی، چین شمالی و سیبری شمال شرقی می‌شود (۲-۴).

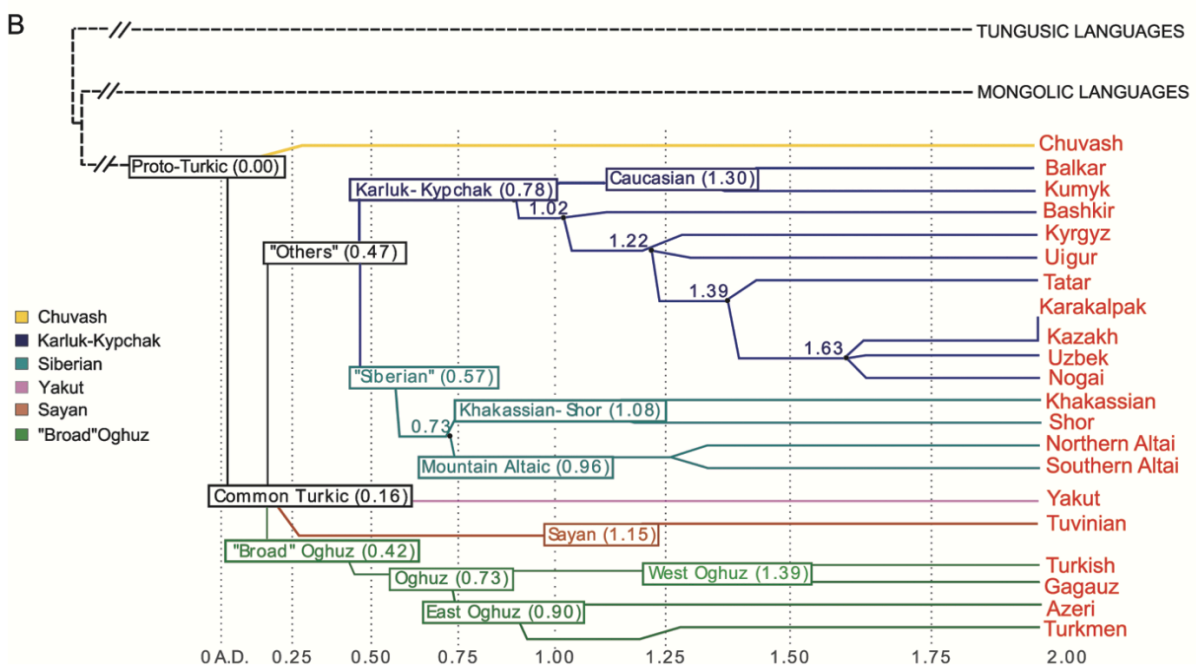
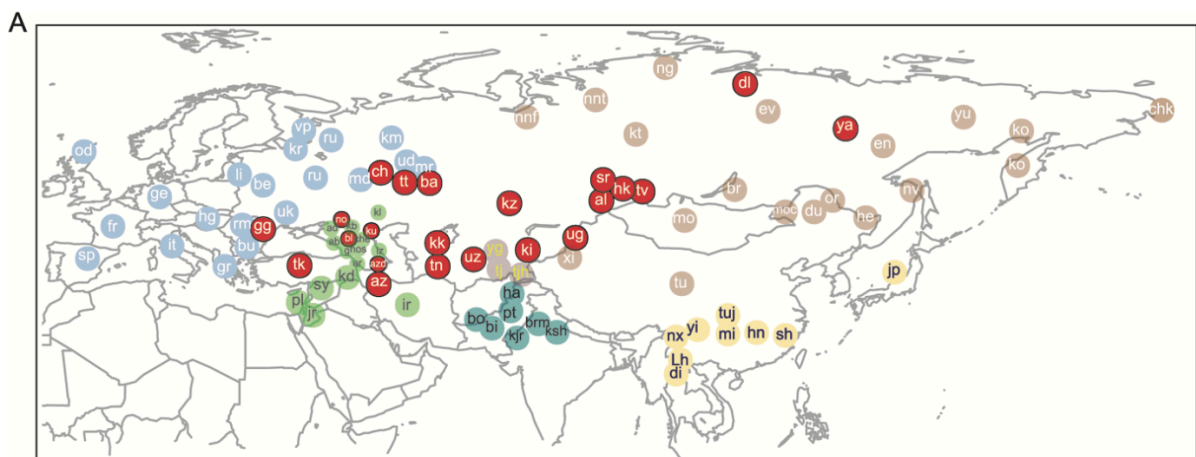
تنوع کنونی زبان‌های ترکی که در این گستره جغرافیایی وسیع رواج دارند، تنها بازتاب‌دهنده تاریخچه اخیر (۲۱۰۰ تا ۲۳۰۰ ساله) و اگرایبی آن‌هاست؛ تاریخی که شامل انشعابی عمده به دو شاخه اوغوری (Oghuric) (یا بلغاری) و ترکی مشترک می‌شود (۵ و ۶). پیش از این دوره، مرحله ترکی باستانی اولیه (که هیچ‌گونه داده تاریخی در باره آن در دست نیست) و هم مرحله‌ای درازمدت به نام ترکی-نیا قرار داشته است؛ مشروط بر آنکه در حدود سال‌های ۴۵۰۰ تا ۴۰۰۰ پ م، نوعی وحدت زبانی (یا زبانی-نیا) میان ترکی و مغولی وجود داشته باشد (۷ و ۸).

نخستین واحدهای سیاسی زیر حاکمیت ترک‌زبانان (که در فاصله سده‌های ۶ تا ۹ شکل گرفتند) در مناطقی متمرکز بودند که امروزه شامل مغولستان، شمال چین و جنوب سیبری می‌شود. بر همین اساس، این منطقه به عنوان کانون آغازین پراکندگی کوچ‌نشینان دامپرور ترک‌زبان معرفی شده است (۳ و ۴). ما در اینجا از این منطقه با عنوان سرزمین مادری آسیای درونی (Inner Asian Homeland یا IAH) یاد می‌کنیم و به دست‌کم دو مسئله در خصوص این فرضیه اشاره خواهیم کرد.

نخست، همین محدوده جغرافیایی پیش‌تر زیر سلطه امپراتوری شیونگنو (Xiongnu) (۲۰۰ پ م تا ۱۰۰ م) و سپس کنفدراسیون کوتاه‌مدت شیانبی (Xianbei) (۱۰۰ تا ۲۰۰ م) و دولت روران (Rouran) که با نام‌های یوآن-یوآن یا آوارهای آسیایی نیز شناخته می‌شوند (۴۰۰ تا ۵۰۰ م) قرار داشت. این تشکیلات سیاسی استپ‌نشین احتمالاً توسط مردمان غیرترک‌زبان تأسیس شده و ظاهراً قبایل گوناگونی از نظر قومی را متحد کرده بودند. تنها در نیمه دوم سده ۶ م بود که مردمان ترک‌زبان کنترل منطقه را در دست گرفتند و خاقانات گوکتورک (Göktürk) را با سرعتی بالا گسترش دادند؛ حکومتی که دیری نپایید و جای خود را به خانان و خاقانات‌های متعدد دیگری داد که قلمروشان از شمال شرقی چین تا استپ‌های پونتیک-خزر (بحیره سیاه-کسپین) در اروپا امتداد داشت (۲-۴). دوم، گوکتورک‌ها نخستین واحد قومی شناخته‌شده‌ای هستند که در آن مردمان ترک‌زبان با نام تورک پدیدار می‌شوند. با این حال، رد پای مردمان ترک‌زبان در منابع تاریخی مکتوب به پیش از آن دوران بازمی‌گردد؛ به‌ویژه زمانی که قبایل ترک‌زبان اوغوری در سده ۵ در استپ‌های شمال دریای سیاه (پونتیک) ظاهر شدند، یعنی بسیار پیش از ظهور خاقانات گوک تورک در IAH (۹). بدین ترتیب، مراحل اولیه پراکندگی مردمان ترک‌زبان همچنان به‌درستی شناخته‌نشده باقی مانده و دانش ما در باره زیستگاه باستانی آنان، همچنان در حد یک فرضیه است.

مطالعات پیشین مبتنی بر کروموزوم-وای، دی‌ان‌ای میتوکندریایی (mtDNA) و نشانگرهای اتوزومی نشان می‌دهند که اگرچه مردمان ترک‌زبان غرب آسیا (ترک‌های آناتولی و آذری‌ها) و اروپای شرقی (گاکائوزها، تاتارها، چوواش‌ها و باشقیرها) عموماً از نظر ژنتیکی به همسایگان جغرافیایی خود شباهت دارند، اما سهم اندکی از هاپلوگروه‌های mtDNA و وای را نیز نشان می‌دهند که در حالت عادی مختص آسیای شرقی هستند (۱۰-۱۵). همان‌طور که انتظار می‌رفت، جمعیت‌های ترک‌زبان آسیای مرکزی (قرغیزها، قزاق‌ها، اوزبک‌ها و ترکمن‌ها) بخش بزرگ‌تری از خزانه ژنی تک‌والدی خود (۹ تا ۷۶٪ از کروموزوم-وای و بیش از ۳۰٪ از دودمان‌های mtDNA) را با جمعیت‌های آسیای شرقی و سیبری مشترک هستند (۱۶ و ۱۷). از این نظر، آن‌ها با همسایگان جنوبی غیرترک‌زبان خود – از جمله تاجیک‌ها، ایرانیان و گروه‌های قومی مختلف در پاکستان (به‌جز هزاره‌ها) – تفاوت دارند. با این حال، هدف این مطالعات تعیین دقیق منشای جغرافیایی و زمان ورود یا آمیختگی ژن‌های اوراسیای شرقی در میان مردمان ترک‌زبان امروزی نیست. دودمان‌های mtDNA و به‌ویژه دودمان‌های کروموزوم-وای با منشای «شرقی» (با توجه به سطح دقت روش‌های موجود در زمان انجام آن مطالعات)، فاقد ویژگی‌های جغرافیایی لازم برای تمایز صریح میان مناطق مختلف در شمال شرقی آسیا و سیبری و/یا میان جمعیت‌های ترک‌زبان و غیرترک‌زبان آن منطقه هستند (۱۸ و ۱۹).

مطالعات متعددی با استفاده از داده‌های پنل SNP در سطح کل ژنوم، ساختار ژنتیکی جمعیت‌های اوراسیا را توصیف کرده‌اند؛ اگرچه برخی از آن‌ها شامل جمعیت‌های مختلف ترک‌زبان نیز می‌شوند (۱۵ و ۲۰-۲۴)، اما تمرکز آن‌ها بر روشن ساختن تاریخچه جمعیتی پیوستار ترک‌زبانان نیست. در مواردی که امکان مقایسه با بیش از یک همسایه جغرافیایی وجود دارد، مردمان ترک‌زبان از نظر ژنتیکی به همسایگان جغرافیایی غیرترک‌زبان خود در آناتولی (۲۲ و ۲۵)، قفقاز (۱۵) و سیبری (۲۱ و ۲۳) نزدیک هستند. یک بررسی اخیر بر روی جمعیت‌های سراسر جهان، نشان‌دهنده وجود نشانه‌ای از آمیختگی در دوران متأخر (سده‌های ۱۳ و ۱۴ م) میان سه جمعیت ترک‌زبان (ترک‌ها، اوزبک‌ها و اویغورها) و یک جمعیت غیرترک‌زبان (لزگی‌ها) با مغول‌ها (اهل شمال چین)، داورها (که به زبانی مغولی سخن می‌گویند) و هزاره‌ها (که منشأ مغولی دارند) بوده است (۲۶). این مطالعه همچنین شواهدی از اختلاط (مربوط به دوره پیش از مغول ۴۴۰-۱۰۸۰ م) در میان جمعیت‌های غیرترک‌زبان (به جز چوواش‌ها) اروپای شرقی و بالکان با گروه منبع مرتبط با اوروک‌ها، مغول‌ها و یاکوت‌های مدرن را نشان داد. این اولین شواهد ژنتیکی از جریان ژن تاریخی از یک منبع چینی شمالی و سیبری به برخی از جمعیت‌های اوراسیای شمالی و مرکزی است، اما مشخص نیست که آیا این سیگنال اختلاط در مورد سایر جمعیت‌های ترک‌زبان در سراسر اوراسیای غربی نیز صدق می‌کند یا خیر.



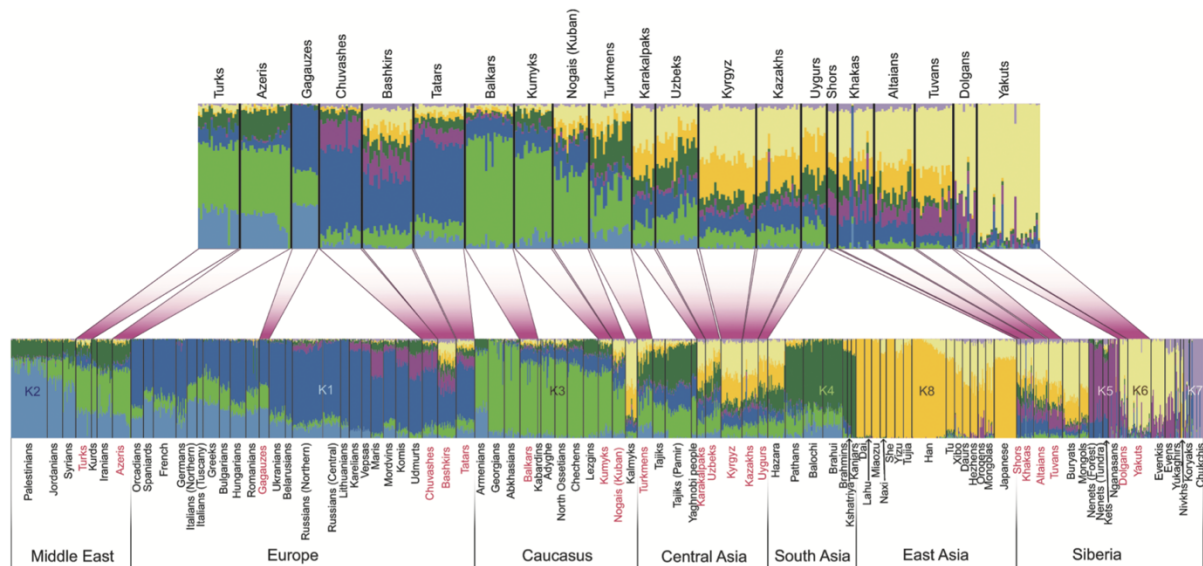
شکل ۱. نقشه جغرافیایی نمونه‌های مورد مطالعه و درخت زبانی زبان‌های ترکی. بخش A. جمعیت‌های غیرترک‌زبان بسته به منطقه، با دایره‌هایی به رنگ‌های آبی روشن، سبز روشن، سبز تیره، قهوه‌ای روشن و زرد نشان داده شده‌اند. جمعیت‌های ترک‌زبان، صرف‌نظر از منطقه نمونه‌برداری، با دایره‌های قرمز مشخص شده‌اند. نام کامل جمعیت‌ها در جدول S1 آمده است. بخش B. درخت زبانی زبان‌های ترکی برگرفته از پژوهش «دیبو» در سال ۲۰۰۴ است و تنها شامل زبان‌هایی می‌شود که توسط مردمان ترک‌زبان مورد بررسی در این مطالعه صحبت می‌شوند. محور افقی نشان‌دهنده مقیاس زمانی بر حسب هزار سال (kya) است. شاخه‌های داخلی با رنگ‌های متفاوت نمایش داده شده‌اند.

در اینجا ما می‌پرسیم که آیا می‌توان سیگنال (های) ژنتیکی صریحی را که توسط همه مردمان ترک‌زبان که احتمالاً از ترک‌های کوچ‌نشین پیشاتاریخی فرضی نسب گرفته‌اند، شناسایی کرد یا خیر. به طور خاص، ما آزمایش می‌کنیم که آیا مردمان مختلف ترک‌زبان میراث ژنتیکی مشترکی دارند که بتوان آن را به IAH فرضی ردیابی کرد یا خیر. به‌طور مشخص‌تر، این پرسش را مطرح می‌کنیم که آیا این نیای مشترک در بازه‌ای زمانی و تاریخی رخ داده است؛ امری که با وجود قطعات کروموزومی طویل و یکسان از نظر نسب که میان مردمان ترک‌زبان سراسر اوراسیای غربی و ساکنان منطقه IAH مشترک است، تأیید می‌شود. برای بررسی این پرسش‌ها، ما از یک آرایه ژنوتیپ‌گذاری با تراکم بالا و پوشش سراسری ژنوم استفاده کردیم تا داده‌هایی را در باره مردمان ترک‌زبان، که نماینده تمامی شاخه‌های اصلی این خانواده زبانی هستند، گردآوری کنیم (شکل ۱ B).

نتایج

به‌منظور تعیین ویژگی‌های ساختار جمعیتی گروه‌های ترک‌زبان در مقایسه با همسایگان جغرافیایی‌شان در سراسر اوراسیا، ما ۳۲۲ نمونه جدید از ۳۸ جمعیت اوراسیایی را تعیین ژنوتیپ کردیم و آن‌ها را با داده‌های منتشرشده پیشین (برای جزئیات به جدول S1 و بخش مواد و روش‌ها مراجعه کنید) ترکیب نمودیم تا مجموعه‌ای شامل ۱۴۴۴ نمونه با ۵۱۵,۸۴۱ نشانگر ژنتیکی حاصل شود. نمونه‌های جدید ارائه‌شده در این مطالعه، از نظر جغرافیایی مناطقی را پوشش می‌دهند که پیش‌تر کمتر مورد بررسی قرار گرفته بودند؛ مناطقی چون اروپای شرقی (ناحیه ولگا-اورال)، آسیای مرکزی، سیبری و خاورمیانه. ما از رویکردی مشابه ساختار-مانند (۲۷) که در نرم‌افزار آمیختگی (۲۸) پیاده‌سازی شده است، برای بررسی ساختار ژنتیکی جمعیت‌های اوراسیایی استفاده کردیم؛ بدین‌صورت که محتمل‌ترین تعداد خوشه‌های ژنتیکی و نسبت‌های اختلاط سازگار با داده‌های ژنوتیپی مشاهده‌شده (در بازه گروه‌های $K=3$ تا $K=14$) را استنباط نمودیم (شکل S1 دیده شود). طوری که در مطالعات پیشین (۱۵، ۲۰ و ۲۹) نشان داده شده است، جمعیت‌های آسیای شرقی معمولاً دارای آللی‌هایی بودند که در دو خوشه کلی قرار می‌گرفتند؛ خوشه‌هایی که در اینجا با نام‌های $k6$ و $k8$ (در مدلی با فرض وجود ۸ جمعیت «اجدادی») نمایش داده شده‌اند (شکل ۲ دیده شود). از نظر جغرافیایی، کانون گسترش این دو مؤلفه (خوشه) به ترتیب در سیبری و آسیای شرقی قرار داشت. فراوانی ترکیبی آن‌ها با حرکت از آسیای شرقی به سمت غرب کاهش می‌یافت (همبستگی با طول جغرافیایی: $p = 8.8 \times 10^{-16}$ ، $R = 0.77$ ، CI ۹۵٪: ۰.۶۶–۰.۸۵). در کل، آللی‌های موجود در جمعیت‌های ترک‌زبان نمونه‌برداری‌شده از سراسر اوراسیای غربی، همانند همسایگان جغرافیایی‌شان، در همان مجموعه

خوشه‌های ژنتیکی اوراسیای غربی یعنی k1 تا k4 قرار می‌گرفتند. علاوه بر این، اقوام ترک‌زبان منطقه ولگا-اورال (شامل چوواش‌ها، تاتارها و باشقیرها) نیز در خوشه k5 جای می‌گرفتند؛ خوشه‌ای که شامل جمعیت‌های اورال‌زبان سیبری (نگاناسان‌ها و نینیتس‌ها) بود و تا برخی از اورال‌زبانان اروپایی (ماری‌ها، اودمورت‌ها و کومی‌ها) نیز امتداد داشت. با این حال، در بیشتر موارد، مردمان ترک‌زبان در مقایسه با همسایگان جغرافیایی خود، مجموع حضور بیشتری از «مؤلفه‌های شرقی» k6 و k8 را نشان دادند.



شکل ۲. ساختار جمعیتی استنباط‌شده با استفاده از تحلیل آمیختگی. نتایج آمیختگی برای مقدار $K=8$ نمایش داده شده است. هر فرد با یک ستون عمودی (۱۰۰٪) انباشته نشان داده شده که نسبت‌های اجدادی را در K جمعیت اجدادی تشکیل‌دهنده نشان می‌دهد. جمعیت‌های ترک‌زبان با رنگ قرمز مشخص شده‌اند. نمودار میله‌ای بالایی تنها جمعیت‌های ترک‌زبان را نشان می‌دهد.

آزمون سه-جمعیتی

«مؤلفه‌های شرقی» k6 و k8 که در میان مردمان ترک‌زبان و غیرترک‌زبان سراسر اوراسیای غربی شناسایی شده‌اند، و هم «مؤلفه‌های غربی» k1، k2 و k3 که در جمعیت‌های سیبری وجود دارند، ممکن است ناشی از رویدادهای جریان ژنی در جهت‌های مخالف در گذشته باشند؛ این تاریخچه اختلاط جمعیتی را می‌توان با استفاده از آماره f_3 به لحاظ آماری آزمود (۳۰ و ۳۱). به‌منظور ارزیابی سناریوهای اختلاط پیشنهادشده توسط تحلیل آمیختگی، ما تمام ترکیبات سه‌جمعیتی ممکن در مجموعه داده خود را با استفاده از آزمون سه-جمعیتی (آماره f_3) بررسی کردیم (۳۰ و ۳۱). ما تنها آن دسته

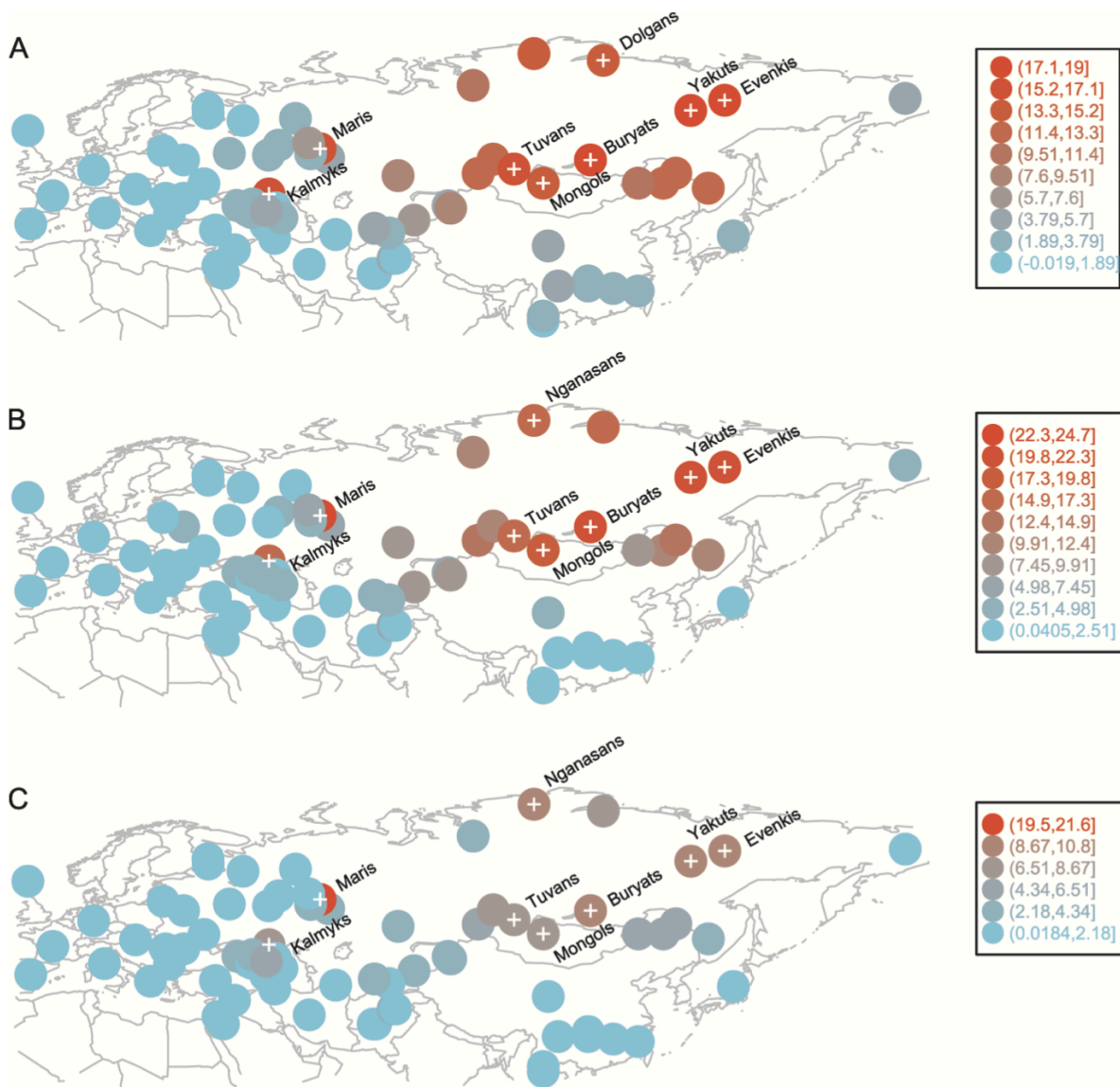
از گروه‌های سه‌تایی جمعیتی f3 (هدف، منبع ۱ و منبع ۲) را گزارش کردیم که دارای منفی‌ترین مقادیر آماره f3 بودند (جدول S2 دیده شود) و جمعیت‌ها را زمانی دارای اختلاط معنادار در نظر گرفتیم که مقدار Z آن‌ها کمتر از ۱.۶۴ بود (یعنی مقدار p برای آزمون یک‌دامنه کمتر از ۰.۰۵ بود). نتایج آزمون‌های سه-جمعیتی ما نشان داد که تقریباً تمام مردمان ترک‌زبان اوراسیای غربی (۱۵ مورد از ۱۶ مورد) و همسایگان غیرترک‌زبان آن‌ها (۴۹ مورد از ۶۱ مورد) با جمعیت‌های مرتبط با آسیای شرقی و سیبری اختلاط داشته‌اند (برای تقسیم‌بندی جغرافیایی به جدول S2 مراجعه کنید). به همین ترتیب، تمام جمعیت‌های ترک‌زبان سیبری و هم برخی (۱۱ مورد از ۲۷ مورد) از جمعیت‌های غیرترک‌زبان اوراسیای شرقی، نشانه‌هایی از اختلاط با جمعیت‌های مرتبط با اوراسیای غربی را بروز دادند. در تفسیر نتایج آماره f3، ذکر این نکته حائز اهمیت است که جمعیت‌های منبع گزارش‌شده، لزوماً نشان‌دهنده جمعیت‌های واقعی دخیل در اختلاط نیستند (۳۱). اگرچه جمعیت‌های منبع دقیق نامشخص بودند، اما مقادیر منفی و معنادار آماره f3 شواهد محکمی دال بر اختلاط در اکثر جمعیت‌های ترک‌زبان و غیرترک‌زبان موجود در مجموعه داده ما فراهم آورد. برای بررسی اینکه آیا این سیگنال‌های اختلاط ناشی از رویدادهای اخیر جریان ژن هستند یا خیر، در مرحله بعد توزیع قطعات کروموزومی طویل مشترک بین جمعیت‌ها در مجموعه داده‌های خود را بررسی کردیم.

توزیع جغرافیایی نیای مشترک اخیر

مطالعه‌ای اخیر نشان می‌دهد که حتی دو فرد غیرخویشاوند از دو سر مخالف اروپا، صدها قطعه کروموزومی IBD (ناشی از نیای مشترک) را از اجدادی که در ۳۰۰۰ سال گذشته می‌زیسته‌اند، به اشتراک می‌گذارند. میزان چنین نیای مشترکِ اخیر با افزایش فاصله جغرافیایی میان جفت‌جمعیت‌ها به صورت نمایی کاهش می‌یابد؛ البته این الگوی وابسته به فاصله ممکن است بر اثر گسترش جمعیت یا شارش ژنی دچار تغییر و انحراف شود (۳۲). ما در مجموعه جمعیت‌های اوراسیایی خود، همبستگی نسبتاً بالایی (ضریب همبستگی پیرسون = ۰.۷۷؛ ۹۵٪ CI: ۰.۷۶ تا ۰.۷۹؛ $p < 2.2 \times 10^{-16}$) میان نرخ کاهش اشتراک IBD و فاصله جغرافیایی مشاهده کردیم. این الگوی وابسته به فاصله احتمالاً هم تحت تأثیر جدایی ناشی از فاصله و هم تحت تأثیر شارش ژنی شکل گرفته است: بسیاری از این جمعیت‌ها آمیخته هستند (که با آماره f3 منفی در جدول S2 نشان داده شده است) و کاهشی وابسته به طول جغرافیایی در فراوانی «مؤلفه‌های شرقی k6 و k8» مشاهده می‌شود. برخی جمعیت‌ها ممکن است به دلیل انزوا، شارش ژنی بیشتر یا رانش ژنتیکی، در این الگوی وابسته به فاصله متمایز به نظر برسند. برای مثال، هنگامی که جمعیت‌های ترک‌زبان غرب اوراسیا (که نمونه‌برداری آن‌ها در

خاورمیانه، قفقاز، اروپای شرقی و آسیای مرکزی انجام شده بود) را از مجموعه داده خود حذف کردیم، همبستگی بهتری میان کاهش اشتراک IBD و فواصل جغرافیایی بین جمعیت‌ها مشاهده کردیم (ضریب همبستگی پیرسون = 0.83 ؛ 95% CI: 0.82 تا 0.85 ؛ $p < 2.2 \times 10^{-16}$). برای شناسایی جمعیت‌هایی که الگوی اشتراک IBD آن‌ها با جمعیت‌های ترک‌زبان از الگوی کاهش وابسته به فاصله تبعیت نمی‌کند، ابتدا میزان اشتراک IBD (میانگین طول ژنوم IBD بر حسب سانتی‌مورگان) را برای هر یک از ۱۲ جمعیت ترک‌زبان غربی در مقایسه با سایر جمعیت‌های موجود در مجموعه داده محاسبه کردیم (جدول S3 دیده شود) و سپس مقدار مشابهی را که برای همسایگان جغرافیایی آن‌ها محاسبه شده بود، از آن کسر نمودیم (برای جزئیات به بخش «مواد و روش‌ها» و برای نمایش شماتیک این تحلیل به شکل S2 مراجعه کنید). هنگامی که تفاوت‌های مربوط به هر ۱۲ جمعیت ترک‌زبان بر روی هم قرار داده شدند، ما سیگنال به‌طور غیرمعمولی قوی از اشتراک تجمعی IBD (که در شکل‌های 3A تا 3C با علامت «مثبت» مشخص شده‌اند) را در جمعیت‌های خارج از منطقه اوراسیای غربی شناسایی کردیم.

سیگنال همبسته‌ی اشتراک IBD برای این جمعیت‌های دوردست، فراتر از حد انتظار بر اساس الگوی کاهش وابسته به فاصله بود. بیشتر این جمعیت‌های دوردست در جنوب سیبری و مغولستان (SSM) و شمال شرقی سیبری واقع شده‌اند، به استثنای دو نمونه در اروپای شرقی (ماری‌ها) و قفقاز شمالی (کالمیک‌ها). اصولاً، هنگامی که الگوی اشتراک IBD را به این شیوه میان جمعیت‌های ترک‌زبان و غیرترک‌زبان همسایه مقایسه می‌کنیم، ممکن است به دلیل رانش ژنتیکی در یکی از جمعیت‌های مورد مقایسه، سیگنال بالایی از اشتراک IBD با برخی جمعیت‌های سیبری مشاهده کنیم؛ اما احتمال اینکه چنین سیگنال‌های تصادفی‌ای میان چندین جمعیت ترک‌زبان همبستگی داشته و در یک منطقه خاص انباشته شوند، ناچیز است. در واقع، فرضیه صفر در این تحلیل، عدم وجود تفاوت سیستماتیک میان هر یک از جمعیت‌های ترک‌زبان و همسایگان جغرافیایی مربوط به آن‌ها را فرض می‌کرد. بنابراین، فرضیه صفر پیش‌بینی می‌کرد که تفاوت‌های تصادفی در سراسر محدوده جغرافیایی جمعیت‌های ترک‌زبان غربی انباشته شوند. برای نشان دادن این انتظار مبتنی بر فرضیه صفر، ما هر یک از جمعیت‌های ترک‌زبان غربی را با جمعیت‌هایی که به‌طور تصادفی از میان مجموعه‌های همسایگان غیرترک‌زبان مربوطه انتخاب شده بودند جایگزین کردیم و تحلیل تفریق/انباشت را تکرار نمودیم (طوری که در شکل S2 نشان داده شده است). هنگامی که مجموعه‌های تصادفی نمونه‌های غیرترک‌زبان مورد آزمایش قرار گرفتند، سیگنال انباشته‌شده، مطابق با پیش‌بینی فرضیه صفر، محدود به جمعیت‌هایی (که در شکل S3 با علامت «مثبت» مشخص شده‌اند) در داخل اوراسیای غربی بود.



شکل ۳. جمعیت‌هایی که دارای سیگنال‌های قوی و همبسته از اشتراک IBD (بخش‌های یکسان از نظر اجداد) با اقوام ترک‌زبان غربی هستند. موقعیت دایره‌ها با محل جغرافیایی جمعیت‌ها مطابقت دارد. رنگ دایره نشان‌دهنده میزان مازاد اشتراک IBD (مطابق راهنمای نمودار) است که یک جمعیت با تمام ۱۲ جمعیت ترک‌زبان غربی دارد. جمعیت‌هایی که میزان اشتراک IBD آن‌ها فراتر از صدک (کوانتایل) ۰,۹۰ است، با علامت «مثبت» (+) مشخص شده‌اند. بخش A سیگنال اشتراک IBD بر اساس قطعات IBD با طول ۱ تا ۲ سانتی‌مورگان. بخش B سیگنال اشتراک IBD بر اساس قطعات IBD با طول ۲ تا ۳ سانتی‌مورگان. بخش C سیگنال اشتراک IBD بر اساس قطعات IBD با طول ۳ تا ۴ سانتی‌مورگان.

با این حال، دو استثنا (انگاسان‌ها و نینیتس‌ها) وجود دارند که بررسی دقیق آن‌ها به یافته‌ای جالب و سازگار با نتایج آمیختگی ما اشاره دارد. این دو جمعیت سیبریایی، یعنی انگاسان‌ها و ننتس‌ها (شکل‌های S3A، S3B، S3E، S3I و S3J)، به زبان‌های اورالی صحبت می‌کنند و تنها زمانی سیگنال انباشته‌ی بالایی را نشان دادند که مجموعه‌های مورد آزمایش ما شامل گویشوران اورالی غربی (ماری‌ها، کومی‌ها، ویپس‌ها و اودمورت‌ها) بودند. این یافته با نتایج تحلیل آمیختگی ما (شکل ۲ دیده شود) همخوانی داشت، زیر مؤلفه نیایی k5 به‌طور اختصاصی میان این گویشوران اورالی غربی و دو گروه اورالی‌زبان سیبریایی، یعنی نگاسان‌ها و نینیتس‌ها، مشترک بود. اکنون به تفاوت کلی میان سیگنال انباشته‌ی اشتراک IBD تحت فرضیه صفر (به شکل S3 مراجعه کنید) و آنچه برای مجموعه جمعیت‌های ترک‌زبان غربی مشاهده شد (شکل ۳ دیده شود)، بازمی‌گردیم. برخی از جمعیت‌های ساکن در مناطق SSM و شمال‌شرقی سیبری، سیگنال قوی اشتراک IBD با جمعیت‌های ترک‌زبان غربی نشان دادند و این الگو به احتمال زیاد حاکی از جریان ژنی اخیر از سوی سیبری است. برای تعیین دقیق‌تر منشای این جریان ژنی، دانستن اینکه کدامیک از جمعیت‌های سیبری بومی زیستگاه‌های کنونی خود هستند، حائز اهمیت است.

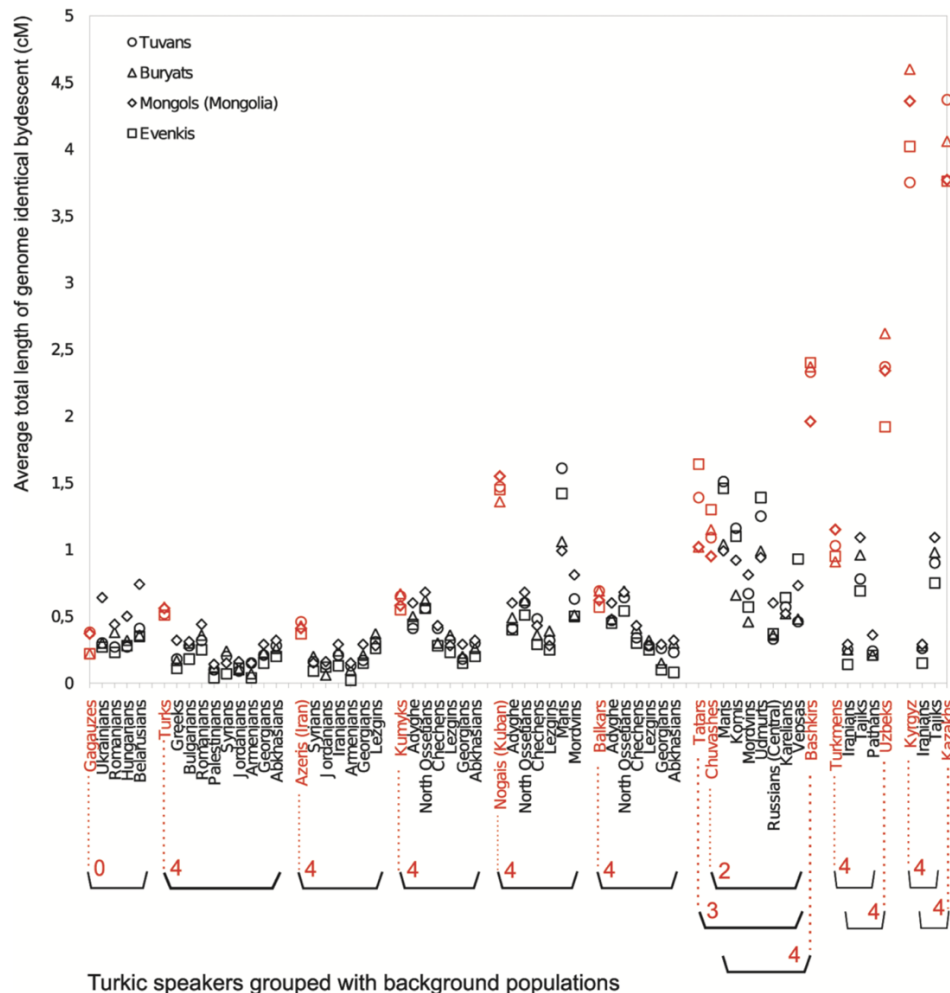
ما در بخش بحث و بررسی نشان می‌دهیم که تنها تووایی‌ها، بوریات‌ها و مغول‌ها ساکن در منطقه SSM بومی زیستگاه‌های کنونی خود هستند (دست‌کم در محدوده زمانی تاریخی شناخته‌شده) و از این رو، این منطقه بهترین گزینه برای منشای جریان ژنی اخیر به سوی جمعیت‌های ترک‌زبان غربی است. شایان ذکر است که این جهت‌گیری شرق به غرب از این واقعیت استنباط می‌شود که بعید است ۱۲ جمعیت نمونه‌برداری‌شده از نقاط مختلف اوراسیای غربی، همه سیگنال همبسته‌ای از اشتراک بالای IBD (بخش‌های یکسان بر اساس نسب) با یک منطقه واحد نشان دهند، مگر آنکه جریان ژنی از آن منطقه دریافت کرده باشند. در واقع، هنگامی که تحلیل خود را با انتخاب تصادفی جمعیت‌های غیرترک‌زبان تکرار کردیم (شکل S2 دیده شود)، نتوانستیم سیگنال همبسته مشابهی را بازتولید کنیم.

تحلیل پیشین ما نشان می‌دهد که جمعیت‌های ترک‌زبان غربی (شکل ۳ و جدول S3) در مقایسه با همسایگان غیرترک‌زبان خود، جریان ژنی قوی‌تری را از منطقه SSM تجربه کرده‌اند، اما مشخص نیست که آیا این سیگنال از نظر آماری معنادار است یا خیر. برای آزمون این موضوع، ما میزان اشتراک IBD را میان گروه جمعیت‌های SSM (شامل تووایی‌ها، مغول‌ها، بوریات‌ها و هم یک جمعیت مهاجر شناخته‌شده به نام ایوینکی‌ها) و تکتک جمعیت‌های ترک‌زبان غربی محاسبه کردیم. سپس، برای هر یک از جمعیت‌های ترک‌زبان غربی، داده‌های همسایگان غیرترک‌زبان آن‌ها را با هم ادغام کردیم و ۱۰,۰۰۰ نمونه جایگشتی ایجاد نمودیم تا بررسی کنیم که آیا میزان اشتراک IBD با آن چهار

جمعیت سیبریایی – که در جمعیت‌های ترک‌زبان مورد آزمایش مشاهده شده بود – ممکن است به صورت تصادفی نیز حاصل شود یا خیر. اشتراک IBD به‌طور جداگانه برای رده‌های مختلف قطعات کروموزومی (۱-۲ سانتی‌مورگان، ۲-۳ سانتی‌مورگان، ۳-۴ سانتی‌مورگان و غیره) برآورد شد و آزمون‌های جایگشت انجام گرفت. در بیشتر موارد، میزان بالاتر اشتراک IBD میان جمعیت‌های ترک‌زبان غربی (در مقایسه با همسایگان غیرترک‌زبان) و جمعیت‌های سیبریایی، از نظر آماری معنادار بود (شکل ۴ و شکل S4 دیده شوند؛ اعداد سرخ نشان‌دهنده تعداد جمعیت‌های سیبریایی با مقدار p کمتر از ۰.۰۱ هستند). برخی از همسایگان غیرترک‌زبان، مانند آدیغه‌ها، ماری‌ها، اودمورت‌ها و اوسیتیایی‌های شمالی نیز تعداد نسبتاً بالایی از قطعات IBD (مشترک از طریق نسب) را با جمعیت‌های SSM به اشتراک می‌گذاشتند (شکل ۴ و شکل S4 دیده شوند). ما نتیجه می‌گیریم که جریان ژنی اخیر از منطقه SSM که در تحلیل پیشین ما استنباط شده بود، محدود به اقوام ترک‌زبان غربی نبوده است؛ همچنین، میزان بالاتر اشتراک IBD گواهی بر متمایز بودن جمعیت‌های ترک‌زبان از همسایگان غیرترک‌زبان‌شان است.

هنگامی که قطعات IBD در رده‌های طولی مختلف به‌صورت جداگانه بررسی شدند، الگویی مکانی در اشتراک IBD مشاهده گردید. در رده‌های طولی ۱ تا ۲ سانتی‌مورگان و ۲ تا ۳ سانتی‌مورگان، میزان بالاتر اشتراک IBD برای اکثر ترک‌زبانان – به‌جز گائوزها و چوواش‌ها (و هم‌تاتارها در مورد رده ۲ تا ۳ سانتی‌مورگان) – از نظر آماری معنادار است. برای قطعات IBD طولانی‌تر (۳ تا ۴ سانتی‌مورگان)، شواهد آماری دال بر اشتراک بالاتر IBD در برخی نمونه‌های خاورمیانه و قفقاز (آذری‌ها، کومیک‌ها و بالکارها) ضعیف‌تر می‌شود. منظور از شواهد ضعیف‌تر این است که مازاد معنادار آماری در اشتراک IBD، تنها به زیرمجموعه‌ای از چهار گروه اجدادی کاندیدا محدود می‌شد. در منطقه ولگا-اورال، برای همین رده از قطعات (۳ تا ۴ سانتی‌مورگان)، تنها باشقیرها همچنان شواهد قوی از جریان ژنی نشان می‌دادند، در حالی که تاتارها و چوواش‌ها چنین نبودند. برای این دو جمعیت ترک‌زبان، نتایج تمام آزمون‌ها از نظر آماری معنادار نبود؛ زیرا گروه مرجع (که نمونه‌های جایگشت‌یافته از آن استخراج می‌شدند) شامل جمعیت ماری (با زبان فینی) بود که سطوح قابل‌مقایسه‌ای از آمیختگی آسیایی (شکل ۲) و اشتراک IBD (شکل S4 دیده شود) را نشان می‌داد. هنگامی که قطعات حتی طولانی‌تری (۴ تا ۵ و ۵ تا ۶ سانتی‌مورگان) را در نظر گرفتیم، دیگر مازاد نظام‌مندی در اشتراک IBD برای اقوام ترک‌زبان در خاورمیانه، قفقاز یا منطقه ولگا-اورال مشاهده نکردیم. در مقابل، جمعیت‌های نزدیک‌تر به منطقه SSM (اوزبک‌ها، قزاق‌ها، قرقیزها و اویغورها، و هم باشقیرهای منطقه ولگا-اورال) همچنان میزان به‌طور آماری معناداری از اشتراک قطعات IBD (هویت

بر اساس نسب مشترک) را نشان دادند. این الگوی مکانی را می‌توان تا حدی با کمیاب بودن نسبی قطعات IBD طولانی‌تر در مقایسه با قطعات کوتاه‌تر و هم رویدادهای مکرر جریان ژنی به سمت جمعیت‌های نزدیک‌تر به منطقه SSM توضیح داد.

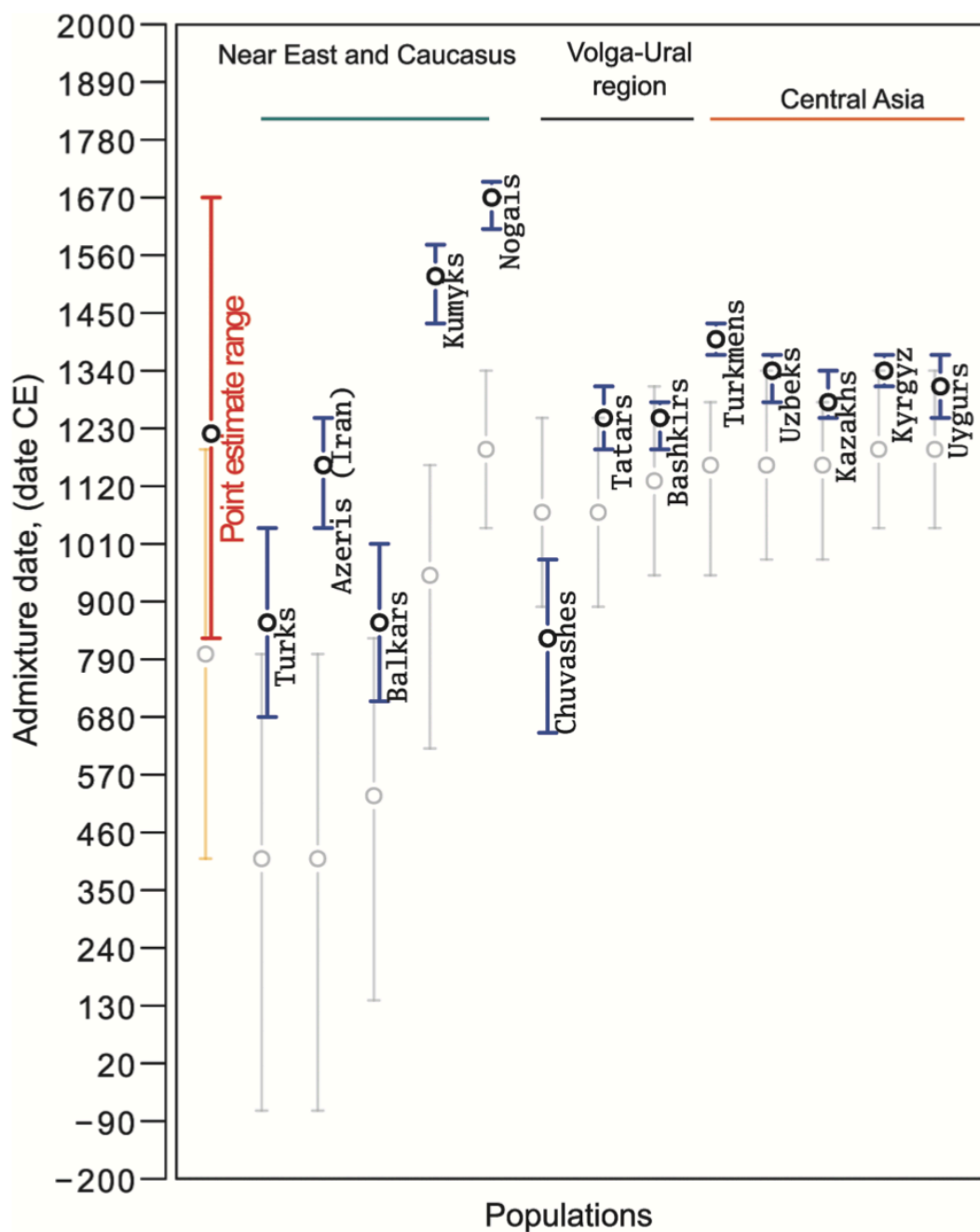


شکل ۴. اشتراک IBD جفت‌به‌جفت بر اساس قطعاتی با طول ۱ تا ۲ سانتی‌مورگان. برای هر جمعیت که در امتداد محور افقی مرتب شده است، میزان اشتراک IBD با سه جمعیت SSM (تووانی‌ها، بوریات‌ها، مغول‌ها) و هم ایوینکی محاسبه شده است. هر جمعیت ترک‌زبان (که با رنگ قرمز نشان داده شده) با استفاده از پیرانتز در کنار همسایگان جغرافیایی خود گروه‌بندی شده است. همسایگان جغرافیایی گروه‌بندی‌شده، جمع‌آوری و برای انجام آزمون جایگشت، همانطور که در بخش M&M توضیح داده شده، مورد استفاده قرار گرفتند. اعداد قرمز زیر نام جمعیت ترک نشان می‌دهد که چه تعداد از جمعیت‌های SSM، اشتراک آماری معناداری از بیماری التهابی روده (IBD) با یک جمعیت ترک نشان می‌دهند. توجه داشته باشید که، برای مثال، باشقیرها، تاتارها و چوواش‌ها، همسایگان جغرافیایی خود را به اشتراک می‌گذارند.

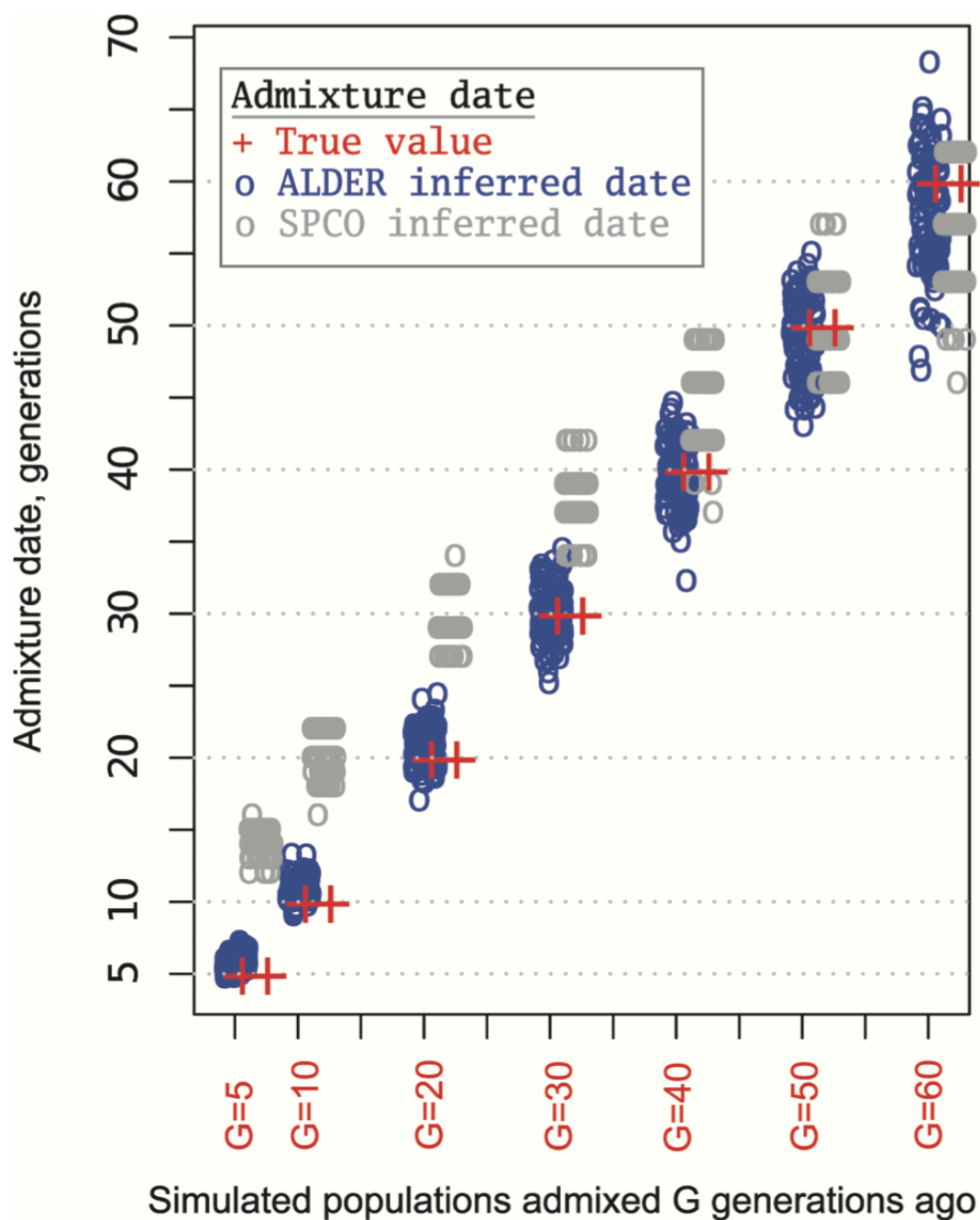
تعیین قدمت آمیختگی آسیایی با استفاده از روش‌های الدر و سپکو

بر اساس شواهد تاریخی، مهاجرت‌های اقوام ترک‌زبان عمدتاً در بازه زمانی سده‌های ۵ تا ۱۶ م (اطلاعات اندکی در باره دوره‌های پیش از آن وجود دارد) رخ داده و تا حدی با گسترش قلمرو مغول‌ها همپوشانی داشته است. با فرض ۳۰ سال برای هر نسل، اجداد مشترک سبیریایی مردمان مختلف ترک‌زبان، پیش از این دوره مهاجرت و در طول آن – یعنی بین ۲۰ تا ۵۳ نسل پیش – می‌زیسته‌اند. طول مورد انتظار برای یک قطعه IBD تک‌مسیری که از جد مشترکی با قدمت ۲۰ تا ۵۳ نسل به ارث رسیده باشد، بین ۲.۵ تا ۰.۹۴ سانتی‌مورگان متغیر است (برای جزئیات بیشتر به بخش روش‌ها مراجعه کنید). با در نظر گرفتن اینکه قطعات IBD چندمسیری به‌طور میانگین طولانی‌تر هستند (۳۳)، سیگنال اشتراک IBD در محدوده ۱ تا ۵ سانتی‌مورگان که میان مردمان ترک‌زبان غربی و جمعیت‌های منطقه SSM مشاهده شده است، ممکن است ناشی از گسترش تاریخی مردمان ترک‌زبان و مغول‌زبان از مبدأ منطقه SSM باشد. اگرچه می‌توان قدمت اجداد مشترک را به‌طور تقریبی و مستقیماً از روی توزیع قطعات IBD مشترک تعیین کرد (۳۲)، اما چنین برآوردی برای اهداف ما بیش از حد کلی و غیردقیق خواهد بود.

در اینجا ما از دو روش مختلف پیاده‌سازی شده در الدر (۳۴) و سپکو (۳۵) برای تعیین زمان آمیختگی سبیریایی/آسیایی در میان اقوام ترک‌زبان استفاده می‌کنیم. تاریخ‌های آمیختگی برای تمام اقوام ترک‌زبان مورد تحلیل (شکل ۵ و جدول S4) در بازه زمانی تاریخی (سده‌های ۵ تا ۱۷) قرار گرفتند؛ بازه‌ای که با دوره مهاجرت‌های کوچ‌نشینان ناشی از گسترش قلمرو مردمان ترک‌زبان (سده‌های ۶ تا ۱۶) و مغول‌ها (سده ۱۳) همپوشانی دارد (۲ و ۳). با این حال، تاریخ‌های آمیختگی برآوردشده توسط این دو روش تنها تا حدی همپوشانی داشتند و در مورد بیشتر جمعیت‌ها با یکدیگر متفاوت بودند (شکل ۵ دیده شود). از این رو، ما مجموعه‌ای از رویدادهای آمیختگی را در بازه زمانی تاریخی مورد نظر شبیه‌سازی کردیم و عملکرد این دو روش را مقایسه نمودیم (برای جزئیات، به بخش مواد و روش‌ها مراجعه کنید). تاریخ‌های تعیین‌شده توسط الدر به مقادیر واقعی شبیه‌سازی شده نزدیک‌تر بودند، در حالی که سپکو به‌طور مداوم تاریخ‌هایی کمی قدیمی‌تر را برآورد می‌کرد (شکل ۶ دیده شود). نکته مهم اینکه تاریخ‌های به‌دست‌آمده از سپکو برای مجموعه داده‌های واقعی ما (شکل ۵) نیز تمایل به قدیمی‌تر بودن داشتند و از این رو، ما احتمال وجود سوگیری در برآوردهای سپکو را مطرح می‌کنیم. از این پس، تنها به بحث در باره تاریخ‌های تعیین‌شده توسط الدر می‌پردازیم.



شکل ۵. تاریخ‌های آمیختگی برای جمعیت‌های ترک‌زبان در مقیاس تاریخ مطلق. دایره‌های آبی، تخمین‌های نقطه‌ای استنباط‌شده توسط الدر را نشان می‌دهند و میله‌های خطا، فواصل اطمینان ۹۵٪ را نشان می‌دهند. دایره‌های خاکستری، تخمین‌های نقطه‌ای استنباط‌شده توسط سپکو را نشان می‌دهند و میله‌های خطا به رنگ خاکستری، فواصل اطمینان ۹۵٪ را نشان می‌دهند. میله قرمز، محدوده تخمین نقطه‌ای (استنتاج‌شده با استفاده از الدر) را در تمام نمونه‌های تحلیل‌شده نشان می‌دهد و میله نارنجی، همین را برای تاریخ‌های استنباط‌شده توسط سپکو نشان می‌دهد. تاریخ‌های اختلاط قبل از عصر مشترک یا جدید با علامت منفی نشان داده شده‌اند.



شکل ۶. تاریخ‌های اختلاط برای جمعیت‌های شبیه‌سازی‌شده. جمعیت‌های شبیه‌سازی‌شده با ترکیب دو جمعیت اجدادی در نسل‌های قبل، همانطور که در بخش M&M توضیح داده شده است، تولید شدند. ما هر سناریوی اختلاط را ۱۲۰ بار تکرار کردیم و با دو روش تاریخ‌گذاری آمیختگی: الدر و سپکو تجزیه و تحلیل کردیم. دایره‌ها نشان‌دهنده تاریخ آمیختگی برای یک جمعیت شبیه‌سازی‌شده هستند و رنگ دایره نشان‌دهنده روش استنباط اختلاط است، همانطور که در راهنما نشان داده شده است. نمادهای قرمز «به‌علاوه» تاریخ آمیختگی واقعی را نشان می‌دهند.

سیگنال عدم تعادل پیوستگی (LD) ناشی از آمیختگی در یک جمعیت آمیخته، با استفاده از دو جمعیت جایگزین (جمعیت‌های مرجع) بازسازی می‌شود؛ هنگامی که چندین جفت از این نوع آزمایش می‌شوند، در امکان مقایسه نزدیکی ژنتیکی آن‌ها با جمعیت‌های آمیزش‌کننده واقعی را بر اساس دامنه منحنی LD وزن‌دار فراهم می‌سازد (۳۶). با این حال، باید توجه داشت که استنباط‌های انجام‌شده به این روش ممکن است گمراه‌کننده باشند، به‌ویژه زمانی که جمعیت مرجع (که از نظر ژنتیکی با جمعیت آمیزش‌کننده واقعی مرتبط است) خود نیز دچار آمیختگی شده باشد. در این حالت، جمعیتی دیگر که ارتباط کمتری دارد، به‌عنوان جایگزین بهتری برای جمعیت آمیزش‌کننده واقعی شناسایی می‌شود. ما تمام ترکیبات ممکن جمعیت‌ها در مجموعه داده خود را به‌عنوان مرجع برای جمعیت‌های ترک‌زبان در نظر گرفتیم و مجموعه‌ای از «بهترین» جفت‌ها (جدول S4) را که بالاترین دامنه منحنی LD وزن‌دار را نشان می‌دادند، گزارش کردیم.

با این حال، تفسیر این نتایج پیچیده است، زیرا بر اساس آزمون‌های سه‌جمعیتی ما (جدول S2)، برخی از جمعیت‌های مرجع خود نیز حاصل آمیختگی هستند. نمی‌توان این احتمال را رد کرد که برخی از گروه‌های مرجع مرتبط با نیاکان تاریخی، به دلیل آمیختگی ژنتیکی، دامنه کمتری را نشان داده‌اند. برای مثال، جمعیت‌های SSM که نشانه‌هایی از جریان ژنی اخیر (فراوانی بیش‌از‌حد قطعات IBD) به سمت جمعیت‌های ترک‌زبان غربی را بروز می‌دادند، در مقایسه با گروه‌های مرجع بدون آمیختگی (که در جدول S4 گزارش کردیم)، دامنه کمتری در شاخص LD (عدم تعادل پیوستگی) ناشی از آمیختگی وزن‌دهی‌شده نشان دادند. جمعیت‌های SSM به‌طور قابل‌توجهی دارای آمیختگی ژنتیکی هستند (جدول S2 دیده شود) و این رویداد احتمالاً در دوران مهاجرت اقوام ترک‌زبان رخ داده است (جدول S4 دیده شود).

اگرچه ما برای هر جمعیت تنها یک تاریخ واحد برای آمیختگی گزارش می‌کنیم، اما باید خاطر‌نشان کرد که احتمالاً شکل‌گیری اقوام ترک‌زبان امروزی حاصل چندین موج مهاجرتی بوده است (۲-۴ و ۳۷). در واقع، مردمان ترک‌زبانی که به منطقه SSM نزدیک‌تر بودند (مانند ساکنان منطقه ولگا-اورال و آسیای مرکزی)، تاریخ‌های متأخرتری (جدیدتری) را نسبت به جمعیت‌های دورتر – مانند ترک‌های آناتولی، آذری‌های ایران و بالکارهای قفقاز شمالی – نشان دادند. تنها نوگای‌ها (که در گذشته کوچ‌نشینان کمربند استپی بودند) و کومیک‌ها (ساکن دامنه‌های شمالی قفقاز) از این الگوی مکانی متمایز هستند.

بحث و بررسی

تحلیل آمیختگی ما (شکل ۲) نشان داد که جمعیت‌های ترک‌زبان پراکنده در سراسر اوراسیا، بخش عمده‌ای از نسب ژنتیکی خود را با همسایگان جغرافیایی غیرترک‌زبان کنونی‌شان مشترک هستند. این موضوع به‌ویژه در مورد مردمان ترک‌زبان ساکن آناتولی، ایران، قفقاز و اروپای شرقی مشهود است، اما تعیین آن برای ترک‌زبانان شمال شرقی سیبری – یعنی یاکوت‌ها و دولگان‌ها – دشوارتر است، زیرا جمعیت‌های مرجع غیرترک‌زبان برای مقایسه با آن‌ها وجود ندارند. ما همچنین دریافتیم که سهم بالاتر مؤلفه‌های ژنتیکی آسیایی، ترک‌زبانان سراسر اوراسیای غربی را از همسایگان غیرترک‌زبان بلافصل‌شان متمایز می‌کند. این نتایج مؤید مدلی است که بر اساس آن، گسترش خانواده زبانی ترک به خارج از هسته اصلی احتمالی‌اش در اوراسیای شرقی، عمدتاً از طریق «جایگزینی زبانی» – و احتمالاً بر اساس سناریوی «سلطه نخبگان» – رخ داده است؛ بدین معنا که کوچ‌نشینان ترک‌زبان مهاجر، به دلیل برتری در سازمان‌دهی نظامی و/یا اجتماعی، زبان خود را بر مردمان بومی تحمیل کرده‌اند.

هنگامی که مردمان ترک‌زبان ساکن در سراسر اوراسیای غربی با همسایگان غیرترک‌زبان خود مقایسه می‌شوند، میزان بالاتری از اشتراک IBD (بخش‌های یکسان بر اساس نسب) را با جمعیت‌های مناطق SSM و شمال شرقی سیبری نشان می‌دهند (شکل ۳ دیده شود). با این حال، دو جمعیت غیرسیبریایی نیز وجود دارند که اشتراک IBD بالایی با گروه‌های ترک‌زبان مورد مطالعه نشان می‌دهند: کالمیک‌ها و ماری‌ها. این موارد استثنایی نیازمند بررسی دقیق با در نظر گرفتن داده‌های تاریخی و مطالعات پیشین هستند. برای نمونه، کالمیک‌های مغول‌زبان تنها در سده ۱۷ از جونگاریا (استانی در شمال غربی چین در مرز مغولستان) به قفقاز شمالی مهاجرت کردند (۳۷)؛ در حالی که ماری‌ها به دلیل آمیختگی ژنتیکی اخیر و به‌طور غیرمعمول بالا با باشقیرها، از سایر همسایگان جغرافیایی خود متمایز می‌شوند: آن‌ها در مقایسه با سایر جمعیت‌های منطقه، در تمامی رده‌های طولی قطعات IBD (از ۱-۲ سانتی‌مورگان تا ۱۱-۱۲ سانتی‌مورگان) میزان بالاتری از اشتراک IBD با باشقیرها نشان می‌دهند ($p > 0.05$). این امر ممکن است ناشی از آن باشد که ما نمونه‌های مردم ماری را در جمهوری باشقورتوستان گردآوری کردیم؛ جایی که به نظر می‌رسد آن‌ها تا حدی با باشقیرها وصلت داشته‌اند. در نهایت، برخی از جمعیت‌های سیبری در واقع مهاجرانی هستند که در مکان‌های فعلی خود ساکن شده‌اند. برای مثال، یاکوت‌ها، ایوینکی‌ها و دولگان‌ها عمدتاً ریشه در منطقه دریاچه بایکال دارند که در اصل همان منطقه SSM است (۲۳).

مشخص شده است که بیشتر جمعیت‌هایی که سیگنال بالایی از اشتراک IBD با جمعیت‌های ترک‌زبان غربی نشان می‌دهند، یا از منطقه SSM منشأ گرفته‌اند و یا با یکی از جمعیت‌های ترک‌زبان مورد

آزمایش آمیختگی داشته‌اند. تنها استثنا، مردم نگاناسان است؛ آن‌ها اشتراک IBD به‌طور غیرمعمول بالایی را هم با مردمان ترک‌زبان غربی (شکل ۳) و هم با جمعیت‌های غیرترک‌زبان انتخاب‌شده به‌صورت تصادفی (شکل S3) نشان می‌دهند. با در نظر گرفتن اینکه جمعیت‌های منطقه SSM (توانی‌ها، مغول‌ها و بوریات‌ها) را می‌توان با اطمینان بومی مناطق خود دانست، و اینکه سایر جمعیت‌های سیبری و غیرسیبری (که در شکل ۳ اشتراک IBD بالایی با مردمان ترک‌زبان غربی نشان می‌دهند) همه ریشه در SSM دارند، ما پیشنهاد می‌کنیم که جمعیت‌های اجدادی این منطقه در جریان اخیر تبادل ژنی به سوی مردمان ترک‌زبان غربی نقش داشته‌اند. همچنین خاطرنشان می‌کنیم که قطعات IBD منطبق با SSM با فراوانی قابل‌توجهی در میان برخی مردمان غیرترک‌زبان، مانند آدیغه‌ها، ماری‌ها، اوسیتیایی‌های شمالی و اودمورت‌ها نیز مشاهده شد (شکل ۴ و شکل S4) که نشان می‌دهد جریان ژنی از منطقه SSM در جمعیت‌های غیرترک‌زبان نیز نقش داشته است. در این خصوص، نمی‌توان توضیحات جایگزینی را که به مهاجرت‌های مردمان ترک‌زبان و مغول‌زبان مربوط نمی‌شوند رد کرد، اما این رویدادهای تاریخی همچنان محتمل‌ترین سناریو محسوب می‌شوند؛ زیرا نسبت بالای قطعات منطبق با SSM، ویژگی مشترک و بارز بسیاری از مردمان ترک‌زبان غربی است و چنین سیگنال همبسته‌ای از اشتراک با جمعیت‌های سیبری در هیچ گروه جمعیتی دیگری مشاهده نمی‌شود (شکل S3). بنابراین، احتمال دارد که مهاجران دارای منشاء SSM با بسیاری از نیاکان جمعیت‌های امروزی اوراسیای غربی تعامل داشته‌اند، اما تعامل قوی‌تر (که در میزان بالاتر اشتراک IBD بازتاب یافته است) با نیاکان مهاجر SSM بود که موجب «ترک‌زبانی» (Turkicization) گردید. ما برای هر یک از جمعیت‌های ترک‌زبان غربی یک آزمون جایگزینی انجام دادیم و مشاهده کردیم که میزان مازاد اشتراک IBD با جمعیت‌های منطقه SSM (در مقایسه با همسایگان غیرترک‌زبان) از نظر آماری معنادار است (شکل ۴ و شکل S4 دیده شوند).

یکی دیگر از نتایج مهم تحلیل اشتراک IBD ما، این یافته است که دو مورد از سه جمعیت SSM که ما آن‌ها را جمعیت‌های منشاء یا نمونه‌های جایگزین امروزی برای جمعیت‌های منشاء می‌دانیم، هر دو گویشور زبان‌های مغولی هستند. این مشاهده را می‌توان به روش‌های گوناگونی تبیین کرد. برای مثال، می‌توان فرض کرد که فتوحات مغول که از سده ۱۳ آغاز شد، با گسترش جمعیتی آن‌ها در سرزمین‌هایی همراه بود که پیش‌تر تا حدی توسط ترک‌زبانان اشغال شده بودند، و این امر منجر به اختلاط/آمیختگی میان ترک‌زبانان و مغول‌زبانان گردید. از سوی دیگر، این احتمال نیز وجود دارد که نیاکان قبایل ترک‌زبان و مغول‌زبان از یک منطقه (یا منطقه‌ای تقریباً یکسان) برخاسته باشند و پیش از گسترش هر یک از این گروه‌ها، دوره‌های متعددی از اختلاط را تجربه کرده باشند. گواه غیرمستقیم بر تبیین دوم،

لایه‌بندی پیچیده و دیرپای وام‌واژه‌های مغولی در زبان‌های ترکی و برعکس است (۳۸). تبیین نخست از منظر تاریخی بعید به نظر می‌رسد؛ زیرا اگرچه فتوحات مغول توسط سپاهیان چنگیزخان در اوایل سده ۱۳ آغاز شد، اما به خوبی می‌دانیم که این فتوحات شامل اسکان مجدد و گسترده‌ی مغول‌ها در سرزمین‌های اشغالی نبود. در عوض، ماشین جنگی مغول‌ها هم‌زمان با پیشروی و گسترش قلمرو، به‌تدریج با پیوستن قبایل گوناگون ترک‌زبان [و ایرانی‌زبان، لعل زاد] تقویت می‌شد و بدین ترتیب، مردمان ترک‌زبان سرانجام به گسترش حضور خود در استپ‌های اوراسیا و فراتر از آن یاری رساندند (۳۹). بنابراین، ما تبیین دوم را ترجیح می‌دهیم، هرچند نمی‌توانیم نقش مغول‌ها را در کل نادیده بگیریم، به‌ویژه با توجه به تاریخ‌های تخمینی اختلاط که با دوره گسترش قلمرو مغول‌ها هم‌پوشانی دارند.

در نهایت، تحلیل اشتراک IBD ما نشان داد که منطقه SSM منشاء جریان ژنی اخیر بوده است. این منطقه یکی از خاستگاه‌های احتمالی برای مردمان ترک‌زبان و هم مغول‌هایی است که از نظر زبانی با آن‌ها خویشاوند هستند. اگرچه حضور امپراتوری مغول در این قلمرو به خوبی ثبت شده است، اما منابع تاریخی به تنهایی برای نسبت دادن بی‌چون و چرای این منطقه به خاستگاه اقوام ترک‌زبان کافی نیستند؛ زیرا حضور برخی از گروه‌های ترک‌زبان متعلق به شاخه اوغوری در مناطق غربی‌تر، یعنی استپ‌های پونتیک-خزر، در اواسط تا اواخر سده ۵ م تأیید شده است. این موقعیت جغرافیایی با منطقه SSM فاصله زیادی دارد و از نظر زمانی نیز بسیار پیش‌تر از دوران تأسیس امپراتوری گوک تورک در منطقه SSM است. بنابراین، مطالعه ما نخستین شواهد ژنتیکی را ارائه می‌دهد که مؤید یکی از فرضیه‌های پیشین مبنی بر وجود یک «خاستگاه یا سرزمین مادری آسیای مرکزی» (IAH) در نزدیکی مغولستان و جنوب سیبری است.

جریان ژنی از ناحیه SSM که بر اساس تحلیل اشتراک IBD استنباط کردیم، باید با استفاده از روشی جایگزین مانند الدر نیز قابل شناسایی باشد؛ روشی که بر تحلیل الگوهای عدم تعادل پیوستگی (LD) ناشی از اختلاط ژنتیکی استوار است. با استفاده از روش الدر، تمام ترکیبات ممکن از جمعیت‌های مرجع موجود در مجموعه داده خود را بررسی کردیم. الگوهای افت LD مشاهده‌شده در میان جمعیت‌های ترک‌زبان غربی، با اختلاط میان جمعیت‌های اوراسیای غربی و جمعیت‌های آسیای شرقی/سیبری همخوانی داشت (جمعیت‌های مرجع شناسایی‌شده در جدول S4 دیده شوند).

تاریخ‌گذاری اختلاط/آمیختگی با استفاده از مجموعه جمعیت‌های آسیای شرقی/سیبری (جدول S4)، رویدادهای اختلاطی را در بازه زمانی بین سال‌های ۸۱۶ (برای چوواش‌ها) تا ۱۶۵۷ (برای نوگای‌ها) نشان داد. ما این جمعیت‌های مرجع را بر اساس بیشترین دامنه منحنی LD انتخاب کردیم، طوری که توسط طراحان این روش پیشنهاد شده است. شایان ذکر است که تمام جمعیت‌های SSM که به عنوان

منبع جریان ژنی SSM شناسایی شده بودند، یا در مرحله پیش‌آزمون الدر به دلیل داشتن سیگنال اختلاط مشترک با جمعیت‌های ترک‌زبان مورد مطالعه حذف شدند، و یا دامنه منحنی LD وزن‌دار آن‌ها در مقایسه با مراجع بدون اختلاط، پایین‌تر بود. در واقع، همان‌طور که نشان می‌دهیم، جمعیت‌های SSM و دو جمعیت شمال شرقی سیبری، همه سیگنال اختلاطی را نشان دادند که از نظر آماری معنادار بود و میان همان مجموعه جمعیت‌های اوراسیای غربی و آسیای شرقی (مشابه آنچه در مردمان ترک‌زبان غربی دیده می‌شود) رخ داده بود (جدول‌های S2 و S4 دیده شوند). بنابراین، مجموعه جمعیت‌های مرجع گزارش‌شده در جدول S4 که بیشترین دامنه منحنی LD را نشان می‌دهند، در حقیقت همان مجموعه جمعیت‌های مرجع بدون اختلاط (جدول S2) هستند که از مرحله فیلترینگ ادلیر عبور کرده‌اند. این فیلتر، هر جمعیت مرجعی را که سیگنال اختلاط مشترکی با جمعیت مورد مطالعه نشان دهد، حذف می‌کند. برای مطالعه ما حائز اهمیت بود که بازه زمانی تاریخ‌های اختلاط استنباط‌شده توسط ادلیر با مهاجرت‌های عمده مردمان ترک‌زبان و گسترش بعدی مغول‌ها (شکل ۵) همپوشانی داشته باشد؛ رویدادهایی که می‌دانیم موجب مهاجرت‌های کوچ‌نشین به آسیای مرکزی سده‌های میانه، خاورمیانه و اروپا شده‌اند. علاوه بر این، با در نظر گرفتن طبقه‌بندی زبانی و بافت منطقه‌ای، شباهت‌هایی با رویدادهای تاریخی در مقیاس بزرگ یافتیم. برای مثال، مردمان امروزی تاتار، باشقیر، قزاق، اوزبک و قرغیز در گستره‌ای از حوضه رود ولگا تا کوه‌های تیان‌شان در آسیای مرکزی ساکن هستند؛ با این حال، شواهد (شکل ۵) نشان‌دهنده آمیختگی ژنتیکی در دوران‌های متأخر، یعنی بین سده‌های ۱۳ و ۱۴، در میان آن‌هاست. این مردمان به زبان‌های ترکی از شاخه قپچاق-قارلوق سخن می‌گویند و زمان آمیختگی آن‌ها پس از مهاجرت‌های احتمالی نیاکان ترک‌زبان قپچاق از مناطق رودخانه‌های ایرتیش و آب در سده ۱۱ بوده است (۳۷). البته استثنائاتی نیز وجود دارد؛ از جمله مردمان بالکار، کومیک و نوگای در قفقاز شمالی که یا تاریخ آمیختگی‌شان به دوره‌ای بسیار پیش‌تر (سده ۸) بازمی‌گردد و یا آمیختگی آن‌ها در دوره‌ای بسیار متأخرتر، یعنی بین سده ۱۵ (برای کومیک‌ها) و سده ۱۷ (برای نوگای‌ها)، رخ داده است.

چوواش‌ها که تنها گویشوران باقی‌مانده از شاخه اوغور هستند، تاریخ آمیختگی قدیمی‌تری (سده ۹) نسبت به همسایگان ترک‌زبان شاخه قپچاق خود در منطقه ولگا نشان دادند. بر اساس منابع تاریخی، هنگامی که امپراتوری اونوگور-بولگار (مستقر در استپ‌های شمال دریای سیاه) در سده ۷ م فروپاشید، برخی از بازماندگان آن به سمت شمال و در امتداد کرانه راست رود ولگا مهاجرت کردند و گروهی را تشکیل دادند که بعدها به نام بولگارهای ولگا شناخته شدند؛ گروهی که نخستین اطلاعات مکتوب

در باره آن‌ها تنها در اواخر سده ۹ در منابع اسلامی پدیدار می‌شود (۴۰). بدین ترتیب، نشانه آمیختگی ژنتیکی در چووش‌ها با زمان احتمالی ورود گویشوران زبان‌های اوغور به منطقه ولگا همخوانی دارد. تفاوت در تاریخ‌های آمیختگی برای سه جمعیت ترک‌زبان شاخه اوغوززبان (آذری‌ها، ترک‌های آناتولی و ترکمن‌ها) قابل توجه بود و موقعیت جغرافیایی آن‌ها توضیحی احتمالی برای این تفاوت ارائه می‌دهد. ترک‌های آناتولی و آذری‌ها که نیاکان آسیای مرکزی‌شان از فلات ایران عبور کرده و تا حد زیادی از جریان ژنی (تبادل ژنتیکی) بعدی با سایر مردمان ترک‌زبان دور مانده بودند، هر دو شواهدی از رویدادهای آمیختگی در دوره‌هایی پیش‌تر (به ترتیب سده‌های ۱۳ و ۹) نسبت به ترکمن‌ها نشان می‌دهند. ترکمن‌های ساکن آسیای مرکزی، آمیختگی ژنتیکی به مراتب متأخرتری را نشان دادند که قدمت آن به سده ۱۴ باز می‌گردد؛ امری که با وضعیت سایر جمعیت‌های ترک‌زبان آسیای مرکزی همخوانی دارد و به احتمال زیاد ناشی از آمیختگی با امواج مهاجرتی متأخرتر – و شاید مکرر – از منطقه SSM به این ناحیه است.

خلاصه، مجموعه نمونه‌های ما که گستره کامل پراکندگی کنونی مردمان ترک‌زبان را در بر می‌گرفت، نشان می‌دهد که بخش قابل‌توجهی از ژنوم اکثر این مردمان با همسایگان جغرافیایی‌شان مشترک است؛ یافته‌ای که از مدل «سلطه نخبگان» برای گسترش زبان‌های ترکی پشتیبانی می‌کند. ما همچنین نشان دادیم که تقریباً تمام مردمان ترک‌زبان غربی، ردپای نیای مشترکی را در ژنوم خود حفظ کرده‌اند که می‌توان آن را تا منطقه SSM (جنوب سیبری و مغولستان) ردیابی کرد. بدین ترتیب، ما شواهد ژنتیکی برای فرضیه «سرزمین مادری یا خاستگاه آسیای درونی» (IAH) در مورد حاملان اولیه زبان‌های ترکی ارائه می‌دهیم؛ فرضیه‌ای که پیش‌تر توسط دیگران و بر اساس داده‌های تاریخی مطرح شده بود. افزون بر این، از آنجا که مردمان ترک‌زبان قطعات نیای مرتبط با SSM را در ژنوم خود حفظ کرده‌اند، امکان تعیین زمان آمیختگی برای ما فراهم شد و تاریخ‌های برآورده شده با دوره تاریخی مهاجرت‌های مردمان ترک‌زبان و گسترش هم‌زمان قلمرو مغول‌ها همخوانی مناسبی دارند. در نهایت، هنوز نکات بسیاری در باره پیامدهای جمعیتی این رویداد تاریخی پیچیده ناگفته مانده است و مطالعات آتی امکان تفکیک سیگنال‌های متعدد آمیختگی در ژنوم انسان و نیز تعیین دقیق خاستگاه جغرافیایی قطعات کروموزومی منفرد را فراهم خواهد ساخت.

مواد و روش‌ها

در زیر عنوان فوق زیرعنوان‌های «بیانیه اخلاقی»، «نمونه‌ها، تعیین ژنوتیپ و کنترل کیفیت»، «تحلیل آمیختگی»، «همبستگی بین طول جغرافیای و نسب ژنتیکی «شرقی»»، «آزمون سه-جمعیتی

(آماره‌های f3)، «تحلیل اشتراک IBD»، «همبستگی میان فاصله جغرافیایی و اشتراک IBD»، «شناسایی انحرافات از الگوی کاهش وابسته به فاصله در اشتراک IBD»، «آزمایش جایگشت»، «سنیابی ترکیبی با استفاده از الدر»، «تعیین قدمت آمیختگی با استفاده از روش تبدیل موجک»، «شبیه‌سازی انعقادی» و «شماره‌های ثبت»؛ و عنوان «اطلاعات تکمیلی (شامل شکل‌ها و جدول‌ها)» و «سیاسگزارى‌ها»ست که می‌توان آن را در متن انگلیسی¹ این پژوهش مطالعه کرد.

سرچشمه‌ها

1. Renfrew C. Archaeogenetics—towards a 'new synthesis'? *Curr Biol.* 2010; 20(4):R162–5. doi: 10.1016/j.cub.2009.11.056 PMID: 20178762
2. Sinor D. The establishment and dissolution of the Turk empire. In: Sinor D, editor. *The Cambridge History of Early Inner Asia*. Cambridge: Cambridge University Press 1990. p. 285–316.
3. Findley CV. *The Turks in world history*. New York: Oxford University Press; 2004. 320 p.
4. Golden PB. *Central Asia in World History*: Oxford University Press; 2011. 176 p.
5. Johanson L. The Structure of Turkic. In: Johanson L, Csátó EA, editors. *The Turkic languages*. London & New York: Routledge & Chapman & Hall; 1998. p. 67–80.
6. Dybo AV. *Chronology of Türkic languages and linguistic contacts of early Türks*. Moscow: Academia Publishers; 2004.
7. Rona-Tas A. The Reconstruction of Proto-Turkic and the Genetic Question. In: Johanson L, Csátó EA, editors. *The Turkic languages*. London & New York: Routledge & Chapman & Hall; 1998. p. 67–80.
8. Dybo AV. *Linguistic contacts of early Turks. Lexical fund. Pre-Turkic period*. Moscow: «Vostochnaya Literatura» Publishers; 2007.
9. Golden PB. Some Thoughts on the Origins of the Turks and the Shaping of the Turkic Peoples. In: Mair V, editor. *Contact and Exchange in the Ancient World Honolulu*: University of Hawaii Press; 2006. p. 136–57.
10. Bermisheva M, Tambets K, VILLEMS R, Khusnutdinova E. Diversity of mitochondrial DNA haplotypes in ethnic populations of the Volga-Ural region of Russia. *Mol Biol (Mosk)*. 2002; 36(6):990–1001. PMID: 12500536
11. 12. 13. 14. Nasidze I, Sarkisian T, Kerimov A, Stoneking M. Testing hypotheses of language replacement in the Caucasus: evidence from the Y-chromosome. *Hum Genet*. 2003; 112(3):255–61. PMID: 12596050
12. Cinnioglu C, King R, Kivisild T, Kalfoglu E, Atasoy S, Cavalleri GL, et al. Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet*. 2004; 114(2):127–48. PMID: 14586639
13. Varzari A, Kharkov V, Stephan W, Dergachev V, Puzyrev V, Weiss EH, et al. Searching for the origin of Gagauzes: inferences from Y-chromosome analysis. *Am J Hum Biol*. 2009; 21(3):326–36. doi: 10.1002/ajhb.20863 PMID: 19107901

- 14 Malyarchuk B, Derenko M, Denisova G, Kravtsova O. Mitogenomic diversity in Tatars from the Volga-Ural region of Russia. *Mol Biol Evol.* 2010; 27(10):2220–6. doi: 10.1093/molbev/msq065 PMID: 20457583
15. Yunusbayev B, Metspalu M, Jarve M, Kutuev I, Rootsi S, Metspalu E, et al. The Caucasus as an Asymmetric Semipermeable Barrier to Ancient Human Migrations. *Mol Biol Evol.* 2012; 29(1):359–65. doi: 10.1093/molbev/msr221 PMID: 21917723
16. Comas D, Plaza S, Wells RS, Yuldaseva N, Lao O, Calafell F, et al. Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages. *Eur J Hum Genet.* 2004; 12(6):495–504. PMID: 14872198
17. 18. 19. Quintana-Murci L, Chaix R, Wells S, Behar D, Sayar H, Scozzari R, et al. Where West meets East: The complex mtDNA landscape of the Southwest and Central Asian corridor. *Am J Hum Genet.* 2004; 74:827–45. PMID: 15077202
18. Derenko M, Malyarchuk B, Denisova GA, Wozniak M, Dambueva I, Dorzhu C, et al. Contrasting patterns of Y-chromosome variation in south Siberian populations from Baikal and Altai-Sayan regions. *Hum Genet.* 2006; 118(5):591–604. PMID: 16261343
19. Derenko M, Malyarchuk B, Grzybowski T, Denisova G, Dambueva I, Perkova M, et al. Phylogeographic analysis of mitochondrial DNA in northern Asian Populations. *Am J Hum Genet.* 2007; 81(5):1025–41. PMID: 17924343
20. Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, et al. Worldwide human relationships inferred from genome-wide patterns of variation. *Science.* 2008; 319(5866):1100–4. doi: 10.1126/science.1153717 PMID: 18292342
21. Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, et al. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature.* 2010; 463(7282):757–62. doi: 10.1038/nature08835 PMID: 20148029
22. Behar DM, Yunusbayev B, Metspalu M, Metspalu E, Rosset S, Parik J, et al. The genome-wide structure of the Jewish people. *Nature.* 2010; 466(7303):238–42. doi: 10.1038/nature09103 PMID: 20531471
23. 24. 25. Fedorova SA, Reidla M, Metspalu E, Metspalu M, Rootsi S, Tambets K, et al. Autosomal and uniparental portraits of the native populations of Sakha (Yakutia): implications for the peopling of Northeast Eurasia. *BMC evolutionary biology.* 2013; 13(1):127.
24. Di Cristofaro J, Pennarun E, Mazieres S, Myres NM, Lin AA, Temori SA, et al. Afghan Hindu Kush: where Eurasian sub-continent gene flows converge. *PLoS One.* 2013; 8(10):e76748. doi: 10.1371/journal.pone.0076748 PMID: 24204668
25. Hodoglugil U, Mahley RW. Turkish population structure and genetic ancestry reveal relatedness among Eurasian populations. *Ann Hum Genet.* 2012; 76(2):128–41. doi: 10.1111/j.1469-1809.2011.00701.x PMID: 22332727
26. 27. 28. 29. Hellenthal G, Busby GB, Band G, Wilson JF, Capelli C, Falush D, et al. A genetic atlas of human admixture history. *Science.* 2014; 343(6172):747–51. doi: 10.1126/science.1243518 PMID: 24531965
27. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. Association mapping in structured populations. *Am J Hum Genet.* 2000; 67(1):170–81. PMID: 10827107
28. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research.* 2009; 19(9):1655–64. doi: 10.1101/gr.094052.109 PMID: 19648217

29. Metspalu M, Romero IG, Yunusbayev B, Chaubey G, Mallick CB, Hudjashov G, et al. Shared and Unique Components of Human Population Structure and Genome-Wide Signals of Positive Selection in South Asia. *Am J Hum Genet.* 2011; 89(6):731–44. doi: 10.1016/j.ajhg.2011.11.010 PMID: 22152676
30. Reich D, Thangaraj K, Patterson N, Price AL, Singh L. Reconstructing Indian population history. *Nature.* 2009; 461(7263):489–U50. doi: 10.1038/nature08365 PMID: 19779445
31. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, et al. Ancient admixture in human history. *Genetics.* 2012; 192(3):1065–93. doi: 10.1534/genetics.112.145037 PMID: 22960212
32. Ralph P, Coop G. The Geography of Recent Genetic Ancestry across Europe. *PLoS Biol.* 2013; 11(5).
33. Browning SR. Estimation of pairwise identity by descent from dense genetic marker data in a population sample of haplotypes. *Genetics.* 2008; 178(4):2123–32. doi: 10.1534/genetics.107.084624 PMID: 18430938
34. Loh PR, Lipson M, Patterson N, Moorjani P, Pickrell JK, Reich D, et al. Inferring admixture histories of human populations using linkage disequilibrium. *Genetics.* 2013; 193(4):1233–54. doi: 10.1534/genetics.112.147330 PMID: 23410830
35. Pugach I, Matveyev R, Wollstein A, Kayser M, Stoneking M. Dating the age of admixture via wavelet transform analysis of genome-wide data. *Genome Biology.* 2011; 12(2):R19. doi: 10.1186/gb-2011-12-2-r19 PMID: 21352535
36. Williams AL, Patterson N, Glessner J, Hakonarson H, Reich D. Phasing of many thousands of geno-typed samples. *Am J Hum Genet.* 2012; 91(2):238–51. doi: 10.1016/j.ajhg.2012.06.013 PMID: 22883141
37. Grousset R. *The Empire of the Steppes: A History of Central Asia.* New Brunswick: Rutgers University Press; 1970.
38. Schöning C. *Turko-Mongolic Relations.* Janhunen J, editor. London: Routledge; 2003.
39. Saunders JJ. *The history of the Mongol conquests:* Routledge; 1971. 275 p.
40. Mako G. The Islamization of the Volga Bulgars: A Question Reconsidered. *Archivum Eurasiae Medii Aevi.* 2011; 18:199–223.
41. Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, et al. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature.* 2014;505:87–91. doi: 10.1038/nature12736 PMID: 24256729
42. Behar DM, Metspalu M, Baran Y, Kopelman NM, Yunusbayev B, Gladstein A, et al. No Evidence from Genome-Wide Data of a Khazar Origin for the Ashkenazi Jews. *Human Biology.* 2013; 85(6):859–900. PMID: 25079123
43. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics.* 2010; 26(22):2867–73. doi: 10.1093/bioinformatics/btq559 PMID: 20926424
44. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature.* 2007; 449(7164):851–61. PMID: 17943122
45. Browning BL, Browning SR. A fast, powerful method for detecting identity by descent. *Am J Hum Genet.* 2011; 88(2):173–82. doi: 10.1016/j.ajhg.2011.01.010 PMID: 21310274

46. Banerjee S. On geodetic distance computations in spatial modelling. *Biometrics*. 2005; 61:617–25. PMID: 16011712
47. Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, Piouffre L, et al. A human genome diversity cell line panel. *Science*. 2002; 296(5566):261–2. PMID: 11954565
48. Ma J, Amos CI. Principal components analysis of population admixture. *Plos One*. 2012; 7(7):e40115. doi: 10.1371/journal.pone.0040115 PMID: 22808102
49. Chen GK, Marjoram P, Wall JD. Fast and flexible simulation of DNA sequence data. *Genome Research*. 2009; 19(1):136–42. doi: 10.1101/gr.083634.108 PMID: 19029539
50. Schaffner SF, Foo C, Gabriel S, Reich D, Daly MJ, Altshuler D. Calibrating a coalescent simulation of human genome sequence variation. *Genome Research*. 2005; 15(11):1576–83. PMID: 16251467
-

1. The Genetic Legacy of the Expansion of Turkic-Speaking Nomads across Eurasia. Yunusbayev et al. 2015.

<https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1005068>

<https://www.researchgate.net/publication/275277706> The Genetic Legacy of the Expansion of Turkic-Speaking Nomads across Eurasia