

УДК 57.026+575.174

СОПОСТАВЛЕНИЕ ГЕНЕТИЧЕСКИХ И ЛИНГВИСТИЧЕСКИХ ДАНЫХ КАК ПОДХОД К ИЗУЧЕНИЮ МИГРАЦИЙ НАРОДОВ

Е.А. Булыгина, В.В. Бочкарев, В.Д. Соловьев

Казанский (Приволжский) федеральный университет, г. Казань, 420008, Россия

Аннотация

Вопросы происхождения человека и его расселения по земному шару рассматриваются с применением методов различных наук. Особый интерес для исследователя в этой области представляют доступные данные геномного профилирования, объем которых ежедневно возрастает. На основе совокупного анализа информации о генетическом разнообразии народов Евразии (по маркерам Y-хромосомы, митохондриальной ДНК и по панели однонуклеотидных полиморфизмов) и их языковой характеристике (грамматические свойства), полученной из современных баз данных, предложен способ изучения и реконструкции миграционных событий, а также описаны некоторые модели корреляционных взаимоотношений между генетическим пулом популяции и ее языковыми особенностями. Настоящее исследование впервые охватывает столь большую территорию (вся Евразия, за исключением Юго-Восточной) и основывается на представительных генетических и лингвистических базах данных.

Ключевые слова: коэволюция языков и генов, геногеография, миграции, Y-хромосома, митохондриальная ДНК, SNP

Введение

В настоящее время вопрос происхождения и генетической истории человечества рассматривается с позиций многих наук, включая археологию, антропологию, нейробиологию, молекулярную генетику, лингвистику, микроэкономику и т. п. Наиболее перспективными для исследователей оказываются комплексные междисциплинарные подходы, интегрирующие данные нескольких источников. В частности, с наступлением постгеномной эры и появлением большого количества доступной информации по генетическому профилированию разнообразных этносов стало возможным уточнить направление путей их миграции и распространения по земному шару. В этом контексте коэволюция языков и генов находится в фокусе внимания исследователей, начиная с работы Кавалли-Сфорца, в которой продемонстрировано наличие глобальной положительной корреляции между генетическим древом народов и лингвистической классификацией их языков, а также описаны возможные причины отсутствия корреляции в отдельных случаях [1]. В детальных исследованиях на локальном уровне было показано наличие значительных вариаций в характере корреляции генетической и языковой характеристик [2–4], что ставит вопрос об условиях, порождающих

различные корреляционные паттерны. В одной из работ [5] приведен пример реконструкции миграционных событий для популяции острова Сумба (Восточная Индонезия), язык которой относится к австронезийской группе. Авторами этой работы вычислены генетические, лингвистические и географические расстояния для нескольких регионов острова, и регион на севере острова, показавший наиболее сильную корреляцию между данными показателями, был обозначен как место первого контакта австронезийцев с местными аборигенами, что косвенно подтверждается существующими легендами. В [5] показано, как место первичной экспансии популяции сохраняет специфические количественные отношения между языком и генофондом, что может быть использовано для реконструкции путей миграции. В настоящей работе мы описываем возможные модели таких отношений на примере нескольких евразийских популяций.

1. Материалы и методы

1.1. Генетический профиль популяций. Для анализа генетического разнообразия были использованы три массива данных. Первый представляет собой базу данных частот гаплогрупп Y-хромосомы у народов мира, характеризующую 252 популяции по частотам 335 Y-гаплогрупп [6]. Эти данные взяты из опубликованных и архивных источников.

Описание популяционно-генетического маркера другого рода приводится в базе данных по 11 западно-евразийским митохондриальным гаплогруппам [7]. В качестве третьего источника была использована матрица генетических расстояний между 36 популяциями Евразии [8], созданная на основе 200 тыс. полиморфных локусов по методике, описанной ранее [9].

1.2. Лингвистические данные. Основой для языковой характеристики популяций послужила созданная в Институте языкознания РАН (г. Москва) грамматическая база данных «Языки мира». В ней содержится описание 315 языков, распространенных в большинстве на территории Евразии, по 3821 бинарному признаку [10].

Использованные источники данных отличались по набору этносов, поэтому итоговая выборка для корреляционного анализа состояла из 53, 36 и 32 популяции для баз данных по Y-хромосоме, митохондриальной ДНК (мтДНК) и однонуклеотидным полиморфизмам (SNP) соответственно (табл. 1).

Для каждого списка популяций были вычислены матрицы генетических и лингвистических дистанций (евклидовы расстояния), а для каждой пары «генофонд – язык» рассчитаны корреляционные коэффициенты по непараметрическому методу Спирмена. Для некоторых этносов построены диаграммы рассеяния, на которых ось абсцисс отражает генетическое, а ось ординат – лингвистическое расстояние от заданной популяции до остальных народов. Статистический анализ выполнен в среде R (R Development Core Team).

2. Результаты и их обсуждение

2.1. Гено-лингвистические корреляции. В табл. 2 приведены статистически значимые ($p < 0.01$) коэффициенты корреляции между генетическими и лингвистическими расстояниями для выбранных народов.

Табл. 1

Представленность популяций в генетических базах данных и классификация их языков. В скобках указано условное обозначение популяции на диаграммах рассеяния

Языковая принадлежность	Языковая ветвь	БД «Языки мира»	БД Y-гаплогрупп	БД мтДНК-гаплогрупп	БД SNP
Индоевропейская семья	Кельтская	Ирландский (ИРЛ)		+	
		Шотландский (ШОТЛ)	+	+	
	Германская	Английский (АНГ)	+	+	
		Голландский (ГОЛ)	+		
		Исландский (ИСЛ)		+	
		Немецкий (НЕМ)	+	+	
		Норвежский (НОРВ)	+	+	
		Шведский (ШВЕД)	+	+	
	Индоиранская	Таджикский (ТАДЖ)	+		
	Романская	Испанский (ИСП)	+	+	
		Итальянский (ИТАЛ)	+	+	+
		Каталанский (КАТ)	+		
		Молдавский (МЛД)			+
Португальский (ПОРТ)		+	+		
Румынский (РУМ)		+	+	+	
Сардинский (САРД)		+	+	+	
Французский (ФР)	+	+	+		
Армянская	Армянский (АРМ)	+	+		
Славянская	Болгарский (БОЛГ) Белорусский (БЕЛ) Македонский (МАК) Польский (ПОЛ) Русский (РУС) Сербский (СЕРБ) Словацкий (СЛК) Словенский (СЛВ) Украинский (УКР)		+	+	+
		+	+	+	
		+	+	+	
		+	+	+	
		+	+	+	
		+	+	+	
		+	+	+	
		+	+	+	
Алтайская семья	Монгольская	Бурятский (БУР)	+		+
		Даурский (ДАУР)			+
		Калмыцкий (КАЛМ)		+	
		Монгольский (МНГ)	+		+
	Монгорский (МГР)			+	
	Тунгусская	Эвенкийский (ЭВК)	+		
		Орочонский (ОРОЧ)			+
Нанайский (НАН)				+	

	Тюркская	Алтайский (АЛТ)	+		
		Азербайджанский (АЗБ)		+	
		Башкирский (БАШ)	+	+	+
		Гагаузский (ГАГ)	+		
		Каракалпакский (КРК)	+		
		Карачаево-балкарский (КРЧ)	+		+
		Казахский (КАЗ)	+	+	+
		Киргизский (КИРГ)	+		+
		Ногайский (НОГ)		+	
		Татарский (ТАТ)	+	+	+
		Турецкий (ТУР)	+	+	
		Туркменский (ТУРК)	+		
		Тувинский (ТУВ)	+		+
		Уйгурский (УЙГ)	+		+
		Узбекский (УЗБ)	+	+	+
		Хакасский (ХАК)	+		
		Чувашский (ЧУВ)	+	+	+
Шорский (ШОР)	+				
Якутский (ЯКУТ)	+		+		
Уральская семья	Финно-угорская	Венгерский (ВЕНГ)	+	+	+
		Карельский (КАР)		+	+
		Коми-зырянский (КМЗ)	+		+
		Коми-пермяцкий (КМП)	+	+	+
		Марийский (МАР)	+	+	
		Саамский (СААМ)	+	+	
		Финский (ФИН)	+	+	
		Хантыйский (ХАНТ)	+		+
		Эстонский (ЭСТ)	+	+	
		Самоедская	Ненецкий (НЕН)	+	
Селькупский (СЕЛ)	+				
Енисейская семья	Североенисейская	Кетский (КЕТ)			+

Можно заметить, что митохондриальный профиль финно-угорских популяций (коми-пермяки, марийцы, финны) показывает отрицательную корреляцию с их языковой характеристикой. Это может быть либо следствием отсутствия в базе мтДНК информации по восточноевропейским и азиатским гаплогруппам, либо признаком обратной зависимости между частотами митохондриальных гаплогрупп и грамматическими свойствами, что следует проверить на большей выборке финно-угорских народов.

Эффективный размер популяции для митохондриальной ДНК составляет 1/4 от соответствующего показателя для ядерной ДНК (митохондриальные гены наследуются в единственной копии и только по материнской линии) [11], а скорость возникновения мутаций в несколько раз выше по сравнению с ядерной ДНК [12], что делает генетический материал митохондрий более чувствительным к случайному дрейфу генов. Ранее было показано, что Y-хромосомные варианты отличаются более строгой географической локализацией [13], однако наличие информации по обеим генетическим системам в данных SNP придает

Табл. 2

Значения корреляции между генетическими и лингвистическими расстояниями ($p < 0.01$). Жирным шрифтом выделены значения корреляции для народов урало-алтайской языковой семьи

Популяция	Y	mt	SNP	Популяция	Y	mt	SNP
Башкиры		0.55		Ногайцы		0.46	
Белорусы		0.49	0.73	Орочоны			0.79
Буряты			0.72	Португальцы	0.36		
Венгры	-0.45			Русские		0.48	0.67
Дауры			0.73	Саами	0.38		
Испанцы	0.51			Сербы	0.40		
Калмыки		0.46		Словаки			0.53
Каталонцы	0.41			Татары			-0.50
Казахи		0.47	0.80	Тувинцы	0.39		0.70
Кеты			0.61	Узбеки		0.55	
Киргизы			0.76	Уйгуры			0.84
Коми-зыряне	0.40			Украинцы	0.47		0.50
Коми-пермяки	0.37	-0.46		Финны	0.47	-0.47	
Македонцы	0.38			Французы	0.45		
Марийцы	0.36	-0.43		Хакасы	0.39		
Монголы	0.39		0.86	Ханты	0.55		0.54
Монгоры			0.77	Шведы	0.36		
Нанайцы			0.79	Шотландцы		-0.46	
Немцы	0.41			Эвенки	0.38		
Ненцы	0.35			Якуты			0.79

им универсальные свойства, позволяющие, с одной стороны, обозначить пути миграций народов на основе однородительских маркеров, а с другой – отразить современное генетическое разнообразие человеческих популяций в геномном разрешении.

Интересным выглядит неравномерное распределение коэффициентов корреляций среди языковых групп. Для Y-хромосомного маркера группой с наиболее ярко выраженными гено-лингвистическими ассоциациями стали уральские народы, для данных SNP – алтайские (табл. 2, жирный шрифт). Кластер урало-алтайских языков распространен на обширной территории – от северо-восточной и центральной Азии до северо-восточной Европы, и более богатое межпопуляционное разнообразие народов-носителей данных языков (к примеру, по сравнению с индоевропейским кластером) и меньшее число контактов между ними ведет к более частому обнаружению значимых взаимосвязей между их языками и генетическим пулом.

2.2. Типичные корреляционные паттерны. Мы можем ожидать следующие варианты взаимоотношений близкородственных народов (народов, принадлежащих одной языковой группе).

1. Группа народов, компактно проживающих на своей исконной территории, генетически и лингвистически однородна благодаря общему происхождению и смешанным бракам. Наблюдается положительная корреляция между дистанциями пар «генофонд – язык» – чем дальше генетически этнос от своих соседей, тем меньше с ними общих языковых свойств.

Примером, иллюстрирующим данную модель, может служить монгольская семья. На рис. 1 приведены диаграммы рассеяния для расстояний между монголами / монгольским языком и другими народами / языками выборки. В левом нижнем углу диаграммы располагаются точки, представляющие народы/языки монгольской семьи (буряты, дауры, монголы), в правом верхнем – индоевропейские народы, далеко отстоящие от монгольских и генетически, и лингвистически, и географически. От линии регрессии на обеих диаграммах удалены координаты «гены – язык» для русской популяции в связи с избыточным описанием русского языка в базе данных «Языки мира». Значение корреляции в данном случае составляет 0.86 для данных SNP и 0.39 для Y-хромосомы ($p < 0.01$).

2. На одной территории проживают народности из различных этнических и языковых групп с тесными культурными контактами и частыми межэтническими браками. Было показано, что в некоторых случаях грамматические заимствования происходят медленнее, чем изменения в генофонде, и лингвистические расстояния между этносами превышают генетические [14]. Обнаруживается слабая положительная гено-лингвистическая корреляция.

В качестве примера рассматривается португальский язык, носители которого и в Европе, и в других частях света активно контактируют с различными этносами (рис. 2). В этом варианте отличным от предыдущего случая является и распределение точек на диаграмме — они образуют форму, напоминающую облако.

Коэффициент корреляции для Y-хромосомы составляет 0.36 ($p < 0.01$).

3. Ситуация, прямо противоположная первой: народ мигрирует на территорию другой языковой группы и тесно контактирует с ней. Язык данной популяции остается практически неизменным, в отличие от генофонда. В качестве примера рассматриваются венгры (рис. 3), чей язык принадлежит уральской семье, а генетический пул сближается с пулом соседних неродственных народов благодаря смешанным бракам, что приводит к возникновению отрицательной корреляции между генетическими и лингвистическими дистанциями по Y-хромосоме ($r = -0,45$, $p < 0.01$).

Обращает на себя внимание конфигурация облака точек на рис. 3: вдоль оси абсцисс располагаются точки, соответствующие уральским и алтайским народам, с которыми венгерский народ связывает общая языковая принадлежность, а вдоль оси ординат — точки, соответствующие индоевропейским популяциям со схожим генофондом.

4. Еще одна типичная, но малораспространенная ситуация — народ мигрирует на территорию другой языковой группы, но продолжает жить крайне изолированно, не ассимилируя с местным населением. Наблюдается положительная гено-языковая корреляция.

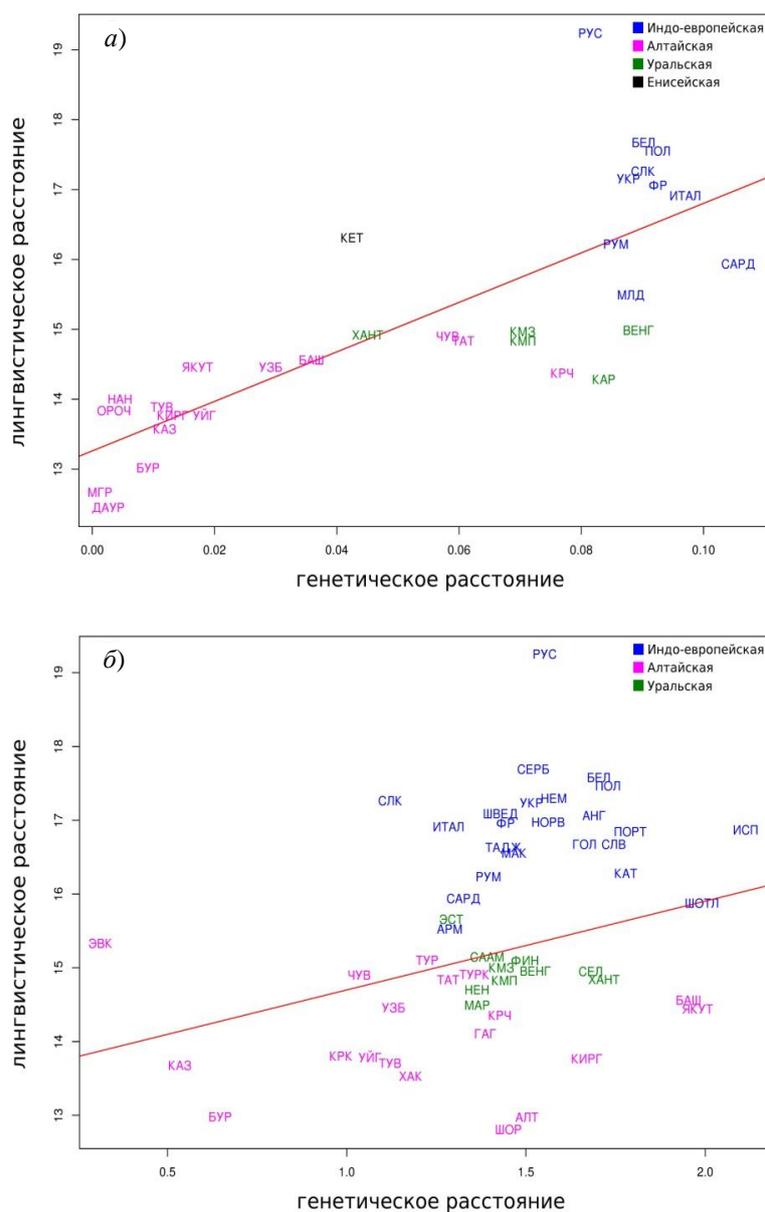


Рис. 1. Диаграмма рассеяния генетических и лингвистических расстояний от монгольской популяции/языка до остальных популяций/языков выборки; а) данные SNP, б) данные Y-гаплогрупп. Цветом обозначена принадлежность языков к языковым семьям

Наиболее подходящим примером служат калмыки (рис. 4). Для них фиксируется положительная корреляция между генетическими и лингвистическими расстояниями по мтДНК ($r = 0.46$, $p < 0.01$). К сожалению, данные по Y-хромосоме и однонуклеотидным полиморфизмам противоречивы и недостаточны, что вызывает необходимость дополнительных исследований.

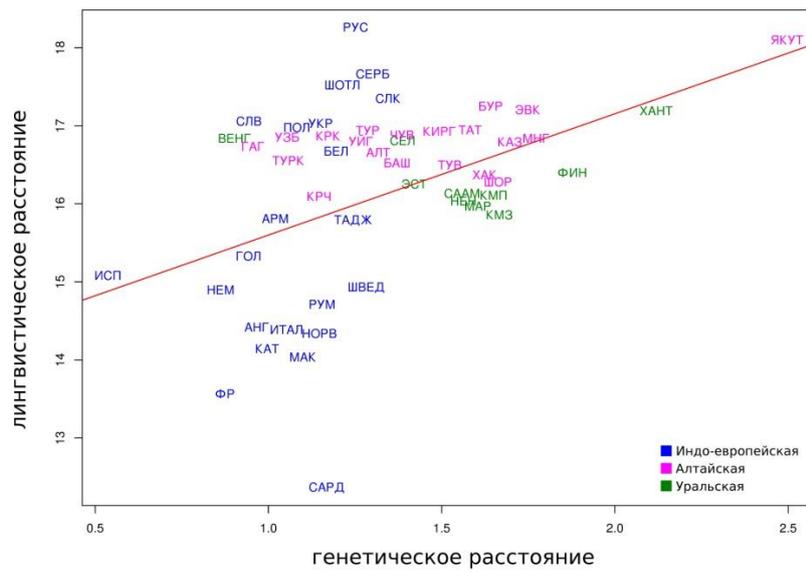


Рис. 2. Диаграмма рассеяния генетических и лингвистических расстояний от португальской популяции/языка до остальных популяций/языков выборки по данным Y-гаплогрупп. Цветом обозначена принадлежность языков к языковым семьям

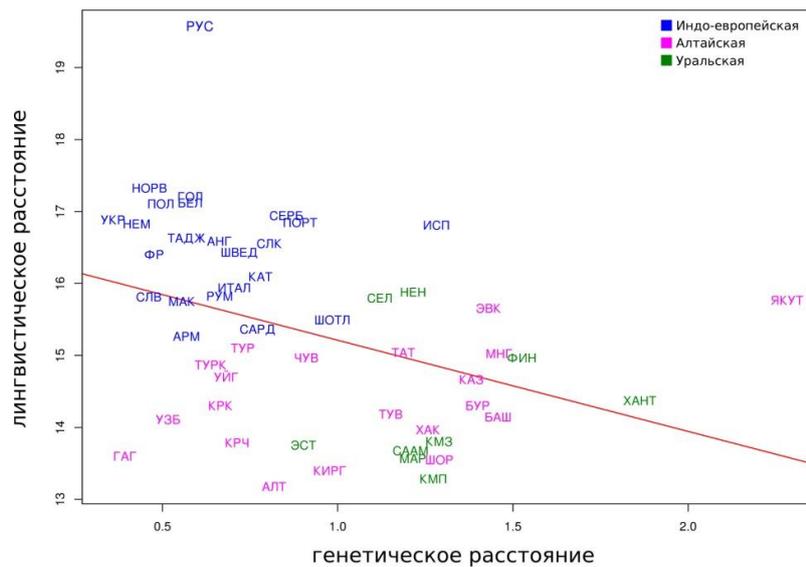


Рис. 3. Диаграмма рассеяния генетических и лингвистических расстояний от венгерской популяции/языка до остальных популяций/языков выборки по данным Y-гаплогрупп. Цветом обозначена принадлежность языков к языковым семьям

Заключение

В работе предложен междисциплинарный подход к реконструкции путей миграции народов на основе анализа корреляций между генетическими и лингвистическими данными, а также описаны их характерные паттерны. Показано, что существует зависимость между характером миграции народа и его контактов

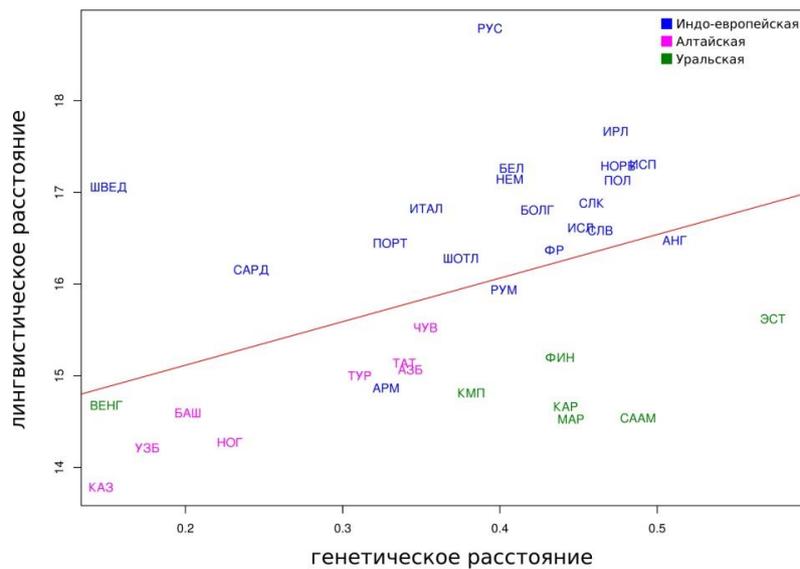


Рис. 4. Диаграмма рассеяния генетических и лингвистических расстояний от калмыцкого народа/языка до остальных народов/языков выборки по данным мтДНК. Цветом обозначена принадлежность языков к языковым семьям

с соседними народами и коэффициентом корреляции генетических и лингвистических расстояний и формой облака точек на диаграмме рассеяния. Наличие значимой отрицательной взаимосвязи является надежным сигналом произошедшего миграционного события. Данное исследование впервые охватывает столь большую территорию (вся Евразия, за исключением Юго-Восточной) и основывается на представительных генетических и лингвистических базах данных.

Благодарности. Работа выполнена в рамках госзадания Минобрнауки РФ (соглашение №34.5517.2017/ВУ) на базе Междисциплинарного центра коллективного пользования Казанского (Приволжского) федерального университета.

Литература

1. Cavalli-Sforza L.L. Genes, peoples, and languages // PNAS. – 1997. – V. 94, No 15. – P. 7719–7724. – doi: 10.1073/pnas.94.15.7719.
2. Sajantila A., Lahermo P., Anttinen T., Lukka M., Sistonen P., Savontaus M.L., Aula P., Beckman L., Tranebjaerg L., Gedde-Dahl T., Issel-Tarver L., DiRienzo A., Pääbo S. Genes and languages in Europe: An analysis of mitochondrial lineages // Genome Res. – 1995. – V. 5, No 1. – P. 42–52.
3. Hunley K., Dunn M., Lindström E., Reesink G., Terrill A., Healy M.E., Koki G., Friedlaender F.R., Friedlaender J.S. Genetic and linguistic coevolution in Northern Island Melanesia // PLoS Genet. – 2008. – V. 4, No 10. – Art. e1000239, P. 1–14. – doi: 10.1371/journal.pgen.1000239.
4. Balanovsky O., Dibirova K., Dybo A., Mudrak O., Frolova S., Pocheshkhova E., Haber M., Platt D., Schurr Th., Haak W., Kuznetsova M., Radzhabov M., Balaganskaya O., Romanov A., Zakharova T., Soria-Hernanz D.F., Zalloua P., Koshel S., Ruhlen M., Renfrew C., Wells R. S., Tyler-Smith Ch., Balanovska E., The Genographic Consortium 2011 Parallel

- evolution of genes and languages in the Caucasus region // *Mol. Biol. Evol.* – 2011. – V. 28, No 10. – P. 2905–2920. – doi: 10.1093/molbev/msr126.
5. *Lansing S., Cox M.P., Downey S.S., Gabler B.M., Hallmark B., Karafet T.M., Norquest P., Schoenfelder J.W., Sudoyo H., Watkins J.C., Hammer M.F.* Coevolution of languages and genes on the island of Sumba, eastern Indonesia // *Proc. Natl. Acad. Sci. U.S.A.* – 2007. – V. 104, No 41. – P. 16022–16026. – doi: 10.1073/pnas.0704451104.
 6. *Балановский О.П., Пшеничнов А.С., Сычев Р.С., Евсеева И.В., Балановская Е.В.* Y-base: частоты гаплогрупп Y хромосомы у народов мира. – 2010. – URL: <http://www.genofond.ru/genofond.ru/default22a2e.html?s=0&p=711>, свободный.
 7. *Запорожченко В.В., Балановский О.П., Пшеничнов А.С., Балановская Е.В.* База данных «Частоты гаплогрупп мтДНК в Западной Евразии». Версия 1.0. – 2007. – URL: <http://www.genofond.ru/genofond.ru/default26226.html?s=0&p=333>, свободный.
 8. *Purdue M.P., Johansson M., Zelenika D., Toro J.R., Scelo G., Moore L.E., Prokhortchouk E., Wu X., Kiemeny L.A., Gaborieau V., Jacobs K.B., Chow W.H., Zaridze D., Matveev V., Lubinski J., Trubicka J., Szeszenia-Dabrowska N., Lissowska J., Rudnai P., Fabianova E., Bucur A., Bencko V., Foretova L., Janout V., Boffetta P., Colt J.S., Davis F.G., Schwartz K.L., Banks R.E., Selby P.J., Harnden P., Berg C.D., Hsing A.W., Grubb R.L. 3rd, Boeing H., Vineis P., Clavel-Chapelon F., Palli D., Tumino R., Krogh V., Panico S., Duell E.J., Quirós J.R., Sanchez M.J., Navarro C., Ardanaz E., Dorronsoro M., Khaw K.T., Allen N.E., Bueno-de-Mesquita H.B., Peeters P.H., Trichopoulos D., Linseisen J., Ljungberg B., Overvad K., Tjønneland A., Romieu I., Riboli E., Mukeria A., Shangina O., Stevens V.L., Thun M.J., Diver W.R., Gapstur S.M., Pharoah P.D., Easton D.F., Albanes D., Weinstein S.J., Virtamo J., Vatten L., Hveem K., Njølstad I., Tell G.S., Stoltenberg C., Kumar R., Koppova K., Cussenot O., Benhamou S., Oosterwijk E., Vermeulen S.H., Aben K.K., van der Marel S.L., Ye Y., Wood C.G., Pu X., Mazur A.M., Boulygina E.S., Chekanov N.N., Foglio M., Lechner D., Gut I., Heath S., Blanche H., Hutchinson A., Thomas G., Wang Z., Yeager M., Fraumeni J.F. Jr., Skryabin K.G., McKay J.D., Rothman N., Chanock S.J., Lathrop M., Brennan P.* Genome-wide association study of renal cell carcinoma identifies two susceptibility loci on 2p21 and 11q13.3 // *Nat. Genet.* – 2011. – V. 43, No 1. – P. 60–65. – doi: 10.1038/ng.723.
 9. *Excoffier L., Smouse P.E., Quattro J.M.* Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data // *Genetics.* – 1992. – V. 131, No 2. – P. 479–491.
 10. *Поляков В.Н., Соловьев В.Д.* Компьютерные модели и методы в типологии и компаративистике. – Казань: Казан. гос. ун-т, 2006. – 210 с.
 11. *Nei M., Tajima F.* Genetic drift and estimation of effective population size // *Genetics.* – 1981. – V. 98, No 3. – P. 625–640.
 12. *Moritz C., Dowling T.E., Brown W.M.* Evolution of animal mitochondrial DNA: Relevance for population biology and systematics // *Annu. Rev. Ecol. Syst.* – 1987. – V. 18, No 1. – P. 269–292.
 13. *Seielstad M., Minch E., Cavalli-Sforza L.* Genetic evidence for a higher female migration rate in humans // *Nat. Genet.* – 1998. – V. 20, No 3. – P. 278–280.
 14. *Tallerman M., Gibson K.* The Oxford Handbook of Language Evolution. – Oxford: Oxford Univ. Press, 2012. – 763 p.

Поступила в редакцию
24.04.17

Булыгина Евгения Александровна, младший научный сотрудник Openlab «Омиксные технологии» Института фундаментальной медицины и биологии

Казанский (Приволжский) федеральный университет
ул. Кремлевская, д. 18, г. Казань, 420008, Россия
E-mail: *boulygina@gmail.com*

Бочкарев Владимир Владимирович, электроник I категории Института физики

Казанский (Приволжский) федеральный университет
ул. Кремлевская, д. 18, г. Казань, 420008, Россия
E-mail: *vbochkarev@mail.ru*

Соловьев Валерий Дмитриевич, доктор физико-математических наук, профессор, ведущий научный сотрудник Института филологии и межкультурной коммуникации

Казанский (Приволжский) федеральный университет
ул. Кремлевская, д. 18, г. Казань, 420008, Россия
E-mail: *maki.solovyev@mail.ru*

ISSN 2542-064X (Print)
ISSN 2500-218X (Online)

UCHENYE ZAPISKI KAZANSKOGO UNIVERSITETA. SERIYA ESTESTVENNYE NAUKI
(Proceedings of Kazan University. Natural Sciences Series)

2018, vol. 160, no. 2, pp. 227–239

Comparison of Genetic and Linguistic Data as an Approach to Studying Human Migrations

E.A. Boulygina^{*}, *V.V. Bochkarev*^{**}, *V.D. Solovyev*^{***}

Kazan Federal University, Kazan, 420008 Russia

E-mail: ^{*}*boulygina@gmail.com*, ^{**}*vbochkarev@mail.ru*, ^{***}*maki.solovyev@mail.ru*

Received April 24, 2017

Abstract

The origin of humans and their settling around the globe are explored using various scientific methods. Genomic profiling data increasing in volume every day are of particular interest to researchers of these problems. In the current post-genomic era, a large amount of data on the genetic profiling of human ethnic groups has been accumulated, thereby making it possible to clarify the direction of human migration and distribution across the globe.

Here, we have presented a method for studying and reconstructing the migration events based on the joint analysis of data on the genetic (Y chromosomal and mitochondrial markers and single nucleotide variation panel) and linguistic (grammar features) diversity of Eurasian populations.

We have shown that, on the one hand, the direction of human migrations is associated with their contacts with neighboring ethnicities, and, on the other hand, the correlation coefficient of genetic and linguistic distances is closely related to and the shape of the points cloud on the scatter plots. The presence of a significant negative correlation is a reliable migration event signal.

Moreover, we have described a few patterns of correlation expected between population's genetic pool and linguistic features. The study covers such a large territory (the whole Eurasia, except the southeastern area) for the first time and is based on the representative genetic and linguistic databases.

Keywords: coevolution of genes and languages, gene geography, migrations, Y chromosome, mitochondrial DNA, SNP

Acknowledgments. The study was performed within the framework of the state task of the Ministry of Science and Education of the Russian Federation (agreement no. 34.5517.2017/VU) and using the equipment of the Interdisciplinary Center of Shared Facilities, Kazan Federal University.

Figure Captions

- Fig. 1. The scatter diagram of genetic and linguistic distances from the Mongolian population/language to other populations/languages; *a*) SNP data, *b*) data on Y haplogroups. The language families of the languages under study are shown with color.
- Fig. 2. The scatter diagram of genetic and linguistic distances from the Portuguese population/language to other populations/languages under study based on the data on Y haplogroups. The language families of the languages under study are shown with color.
- Fig. 3. The scatter diagram of genetic and linguistic distances from the Hungarian population/language to other populations/languages under study based on the data on Y haplogroups. The language families of the languages under study are shown with color.
- Fig. 4. The scatter diagram of genetic and linguistic distances from the Kalmyk population/language to other populations/languages under study based on the data on mtDNA. The language families of the languages under study are shown with color.

References

1. Cavalli-Sforza L.L. Genes, peoples, and languages. *PNAS*, 1997, vol. 94, no. 15, pp. 7719–7724. doi: 10.1073/pnas.94.15.7719.
2. Sajantila A., Lahermo P., Anttinen T., Lukka M., Sistonen P., Savontaus M.L., Aula P., Beckman L., Tranebjaerg L., Gedde-Dahl T., Issel-Tarver L., Di Rienzo A., Pääbo S. Genes and languages in Europe: An analysis of mitochondrial lineages. *Genome Res.*, 1995, vol. 5, no. 1, pp. 42–52.
3. Hunley K., Dunn M., Lindström E., Reesink G., Terrill A., Healy M.E., Koki G., Friedlaender F.R., Friedlaender J.S. Genetic and linguistic coevolution in Northern Island Melanesia. *PLoS Genet.*, 2008, vol. 4, no. 10, art. e1000239, pp. 1–14. doi: 10.1371/journal.pgen.1000239.
4. Balanovsky O., Dibirova K., Dybo A., Mudrak O., Frolova S., Pocheshkhova E., Haber M., Platt D., Schurr Th., Haak W., Kuznetsova M., Radzhabov M., Balaganskaya O., Romanov A., Zakharova T., Soria-Hernanz D.F., Zalloua P., Koshel S., Ruhlen M., Renfrew C., Wells R.S., Tyler-Smith Ch., Balanovska E., The Genographic Consortium 2011 Parallel evolution of genes and languages in the Caucasus region. *Mol. Biol. Evol.*, 2011, vol. 28, no. 10, pp. 2905–2920. doi: 10.1093/molbev/msr126.
5. Lansing S., Cox M.P., Downey S.S., Gabler B.M., Hallmark B., Karafet T.M., Norquest P., Schoenfelder J.W., Sudoyo H., Watkins J.C., Hammer M.F. Coevolution of languages and genes on the island of Sumba, eastern Indonesia. *Proc. Natl. Acad. Sci. U.S.A.*, 2007, vol. 104, no. 41, pp. 16022–16026. doi: 10.1073/pnas.0704451104.
6. Balanovskii O.P., Pshenichnov A.S., Sychev R.S., Evseeva I.V., Balanovskaya E.V. Y-base: Y chromosome haplogroup frequencies among the peoples of the world. 2010. Available at: <http://www.genofond.ru/genofond.ru/default22a2e.html?s=0&p=711>. (In Russian)
7. Zaporozhchenko V.V., Balanovskii O.P., Pshenichnov A.S., Balanovskaya E.V. Database “Frequencies of mtDNA haplogroups in Western Eurasia”, version 1.0. 2007. Available at: <http://www.genofond.ru/genofond.ru/default26226.html?s=0&p=333>. (In Russian)
8. Purdie M.P., Johansson M., Zelenika D., Toro J.R., Scelo G., Moore L.E., Prokhortchouk E., Wu X., Kiemeny L.A., Gaborieau V., Jacobs K.B., Chow W.H., Zaridze D., Matveev V., Lubinski J., Trubicka J., Szeszenia-Dabrowska N., Lissowska J., Rudnai P., Fabianova E., Bucur A., Bencko V., Foretova L., Janout V., Boffetta P., Colt J.S., Davis F.G., Schwartz K.L., Banks R.E., Selby P.J., Harnden P., Berg C.D., Hsing A.W., Grubb R.L. 3rd, Boeing H., Vineis P., Clavel-Chapelon F., Palli D., Tumino R., Krogh V., Panico S., Duell E.J., Quirós J.R., Sanchez M.J., Navarro C., Ardanaz E., Dorronsoro M., Khaw K.T., Allen N.E., Bueno-de-Mesquita H.B., Peeters P.H., Trichopoulos D., Linseisen J., Ljungberg B., Overvad K., Tjønneland A., Romieu I., Riboli E., Mukeria A., Shangina O., Stevens V.L., Thun M.J., Diver W.R., Gapstur S.M., Pharoah P.D., Easton D.F., Albanes D., Weinstein S.J., Virtamo J., Vatten L., Hveem K., Njølstad I., Tell G.S., Stoltenberg C., Kumar R., Koppova K., Cussenot O., Benhamou S., Oosterwijk E., Vermeulen S.H., Aben K.K., van der Marel S.L., Ye Y., Wood C.G., Pu X., Mazur A.M., Boulygina E.S., Chekanov N.N., Foglio M., Lechner D., Gut I., Heath S., Blanche H., Hutchinson A., Thomas G., Wang Z., Yeager M., Fraumeni J.F. Jr., Skryabin K.G., McKay J.D., Rothman N., Chanock S.J., Lathrop M., Brennan P. Genome-wide association study of renal cell

- carcinoma identifies two susceptibility loci on 2p21 and 11q13.3. *Nat. Genet.*, 2011, vol. 43, no. 1, pp. 60–65. doi: 10.1038/ng.723.
9. Excoffier L., Smouse P.E., Quattro J.M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, 1992, vol. 131, no. 2, pp. 479–491.
 10. Polyakov V.N., Solovyev V.D. *Komp'yuternye modeli i metody v tipologii i komparativistike* [Computer Models and Methods in Typology and Comparative Studies]. Kazan, Kazan. Gos. Univ., 2006. 210 p. (In Russian)
 11. Nei M., Tajima F. Genetic drift and estimation of effective population size. *Genetics*, 1981, vol. 98, no. 3, pp. 625–640.
 12. Moritz C., Dowling T. E., Brown W. M. Evolution of animal mitochondrial DNA: Relevance for population biology and systematics. *Annu. Rev. Ecol. Syst.*, 1987, vol. 18, no. 1, pp. 269–292.
 13. Seielstad M., Minch E., Cavalli-Sforza L. Genetic evidence for a higher female migration rate in humans. *Nat. Genet.*, 1998, vol. 20, no.3, pp. 278–280.
 14. Tallerman M., Gibson K. *The Oxford Handbook of Language Evolution*. Oxford, Oxford Univ. Press, 2012. 763 p.

⟨ **Для цитирования:** Булыгина Е.А., Бочкарев В.В., Соловьев В.Д. Сопоставление генетических и лингвистических данных как подход к изучению миграций народов // Учен. зап. Казан. ун-та. Сер. Естеств. науки. – 2018. – Т. 160, кн. 2. – С. 227–239. ⟩

⟨ **For citation:** Boulygina E.A., Bochkarev V.V., Solovyev V.D. Comparison of genetic and linguistic data as an approach to studying human migrations. *Uchenye Zapiski Kazanskogo Universiteta. Seriya Estestvennye Nauki*, 2018, vol. 160, no. 2, pp. 227–239. (In Russian) ⟩