

УДК 575.22

ОСОБЕННОСТИ ГЕНОМНОГО РАСПРЕДЕЛЕНИЯ РЕГИОНОВ ВЫСОКОЙ ГОМОЗИГОТНОСТИ У КОРЕННОГО НАСЕЛЕНИЯ СЕВЕРНОЙ ЕВРАЗИИ НА ИНДИВИДУАЛЬНОМ И ПОПУЛЯЦИОННОМ УРОВНЯХ НА ОСНОВЕ АНАЛИЗА SNP ВЫСОКОЙ ПЛОТНОСТИ

© 2021 г. Н. А. Колесников^{1, *, **}, В. Н. Харьков¹, А. А. Зарубин¹, М. О. Раджабов²,
М. И. Воевода³, М. А. Губина⁴, Э. К. Хуснутдинова⁵, С. С. Литвинов⁵, Н. В. Екомасова⁵,
О. В. Штыгашева⁶, Н. Р. Максимова⁷, А. Л. Сухомясова⁷, В. А. Степанов¹

¹Научно-исследовательский институт медицинской генетики, Томский национальный исследовательский
медицинский центр Российской академии наук, Томск, 634050 Россия

²Институт физики им. Х.И. Амирханова, Дагестанский федеральный исследовательский центр
Российской академии наук, Махачкала, 367012 Россия

³Федеральный исследовательский центр фундаментальной и трансляционной медицины, Новосибирск, 630060 Россия

⁴Федеральный исследовательский центр Институт цитологии и генетики
Сибирского отделения Российской академии наук, Новосибирск, 630090 Россия

⁵Институт биохимии и генетики, Уфимский федеральный исследовательский центр
Российской академии наук, Уфа, 450054 Россия

⁶Хакасский государственный университет им. Н.Ф. Катанова, Абакан, 655000 Россия

⁷Северо-Восточный федеральный университет им. М.К. Аммосова, Якутск, 677000 Россия

*e-mail: nikita.kolesnikov@medgenetics.ru

**e-mail: nik.fleming@mail.ru

Поступила в редакцию 29.01.2021 г.

После доработки 11.05.2021 г.

Принята к публикации 20.05.2021 г.

Анализ геномных данных по 886 889 аутомных SNP позволил оценить уровень инбридинга на основе исследования регионов высокой гомозиготности (F_{RON}) в выборке из 1836 человек, включающей 76 популяций коренного населения Восточной Европы, Кавказа, Волго-Уральского региона, Сибири и Средней Азии. Данные были получены с помощью биочипов Infinium Multi-Ethnic Global-8 Kit. Среди популяций Дагестана, внутри нахско-дагестанской языковой семьи, можно выделить представителей языковых групп с максимальным уровнем F_{RON} : дидойская (0.0727) и андийская (0.0378) группы. Последний показатель близок к таковому в популяциях, говорящих на чукотско-камчатском и нивхском языках Сибири (0.0360). Для популяций Сибири характерно наибольшее значение суммарной длины коротких и средних регионов высокой гомозиготности на человека, по сравнению с другими исследованными популяциями. Популяции коренных народов Дагестана и Сибири в целом характеризуются большей степенью гомозиготизации генома и большей вариативностью регионов с повышенной гомозиготностью по сравнению с другими популяциями Кавказа, Средней Азии, Европы и Волго-Уральского региона.

Ключевые слова: популяционная генетика, регионы высокой гомозиготности, популяции человека, инбридинг, Северная Евразия.

DOI: 10.31857/S0016675821110059

Существенная часть любого индивидуального генома человека представлена регионами высокой гомозиготности (англ. RON – runs of homozygosity). Общая длина и распределение таких участков связаны с генетико-демографической историей популяции и отражают эффекты основателя, уровень инбридинга, а также могут быть следствием действия сильного направленного отбора [1, 2]. Современные методы анализа данных полных геномов или высокоплотных наборов SNP позволя-

ют детально выявлять структуру и распределение RON в индивидуальных геномах и популяциях с целью уточнения генетической истории популяций, оценки уровня инбридинга, поиска сигналов естественного отбора [1, 3]. Данные по распределению RON в популяциях могут служить основой для идентификации участков генома, связанных с заболеваниями, в контексте картирования гомозиготности [4]. Общая длина и структура RON характеризуются существенной межиндивиду-

альной и межпопуляционной вариабельностью. Как правило, сумма длин коротких и средних ROH у индивида возрастает от восточно-африканских к европейским и азиатским популяциям в силу накопления рекомбинационных событий. Напротив, общий суммарный размер длинных ROH показывает большие межиндивидуальные вариации, отражая, вероятно, эффекты недавнего инбридинга [5, 6]. Кроме того, показано, что частота распределения ROH в геноме коррелирует со скоростью рекомбинации, а также с сигналами недавнего положительного отбора [7].

Предыдущие исследования показали, что у неродственных людей широко распространены короткие ROH, измеряемые несколькими десятками тысяч пар нуклеотидов (тпн) [8]. Регионы высокой гомозиготности средних размеров – от сотен тысяч до нескольких миллионов пн – являются, как правило, результатом недавнего общего происхождения индивидов внутри популяции. Длинные ROH, измеряющиеся десятками миллионов пн, характерны для популяций с высоким уровнем близкородственных браков [5].

Среди многочисленных групп населения Северной Евразии, имеющих различную генетико-демографическую историю, особый интерес для исследования ROH представляют популяции коренных народов Северного Кавказа и Сибири, характеризующиеся, как правило, небольшой численностью и значительной географической изоляцией от других популяций. В частности, ранее для популяций Дагестана, говорящих на языках нахско-дагестанской семьи, были показаны высокие коэффициенты инбридинга, очень большое количество и большая суммарная длина гомозиготных блоков, по сравнению с окружающими популяционными группами Кавказа, Ближнего Востока, Европы, Центральной и Южной Азии [9]. Для популяций Сибири мы ранее на низкоплотных чипах обнаружили значительно большее число и суммарную длину ROH по сравнению с популяциями Волго-Уральского региона, Кавказа, Европы и Средней Азии [10].

Настоящая работа посвящена оценке уровня инбридинга, структуры и вариабельности ROH в популяциях Северной Евразии на основе анализа SNP высокой плотности.

МАТЕРИАЛЫ И МЕТОДЫ

Данные были получены с помощью биочипов “Infinium Multi-Ethnic Global-8 Kit”, содержащих 1779819 SNP, в выборке из 1836 человек. Популяции коренного населения Сибири ($N = 477$) представлены алтайцами (Б – с. Бешпельтир Чемальского муниципального района, $N = 24$ и К – с. Кулада Онгудайского района, $N = 25$), бурятами (А – п. Агинское Агинского района, $N = 23$ и К – с. Курумкан Курумканского района, $N = 28$), калмыками ($N = 29$), кетами ($N = 15$), коряками ($N = 20$), нивхами ($N = 13$), татарами (Т – г. Томск, $N = 20$),

тувинцами ($N = 28$), удэгейцами ($N = 15$), хантами (К – с. Казым Белоярского района, $N = 30$ и Р – д. Русскинская Сургутского района, $N = 26$), хакасами (Т – сагайцами Таштыпского р-на, $N = 29$ и Ш – качинцами Ширинского р-на, $N = 26$), чукчами ($N = 25$), чулымцами ($N = 22$), эвенками (З – забайкальские (п. Чара Каларского района, с. Моклан и с. Тупик Тунгиро-Олёкминского района), $N = 25$ и Я – якутские, $N = 28$) и якутами ($N = 26$).

Коренные популяции Дагестана ($N = 518$) включают аварцев ($N = 24$), даргинцев ($N = 28$), лакцев ($N = 24$), табасаран ($N = 21$), лезгин ($N = 28$), агулов ($N = 24$), рутульцев ($N = 22$), цахуров ($N = 24$), арчинцев ($N = 24$), андийцев ($N = 17$), ахвацев ($N = 24$), тиндинцев ($N = 18$), каратинцев ($N = 24$), багулалов ($N = 23$), ботлихцев ($N = 16$), хваршин ($N = 15$), цезов ($N = 24$), бежтинцев ($N = 22$), чамалинцев ($N = 24$), гунизбцев ($N = 17$), гинухцев ($N = 19$), кумыков ($N = 37$) и караногайцев ($N = 19$).

Волго-Уральский географический регион и популяции Европы ($N = 419$) представлены коми ($N = 30$), марийцами ($N = 30$), вепсами ($N = 30$), удмуртами (удмурты $N = 30$, Б – п. Балезино, $N = 28$ и Ш – с. Шаркан, $N = 18$), чувашами ($N = 26$), башкирами (Б – Бурзянский р-н, $N = 34$, П – пермские, $N = 15$ и С – Салаватский р-н, $N = 15$), бесермянами ($N = 16$), карелами ($N = 29$), мордвой (Э – эрзя, $N = 16$, М – мокша, $N = 30$, Ш – шокша, $N = 14$), татарами (К – казанские, $N = 33$) и русскими ($N = 33$).

Коренные популяции Кавказского региона ($N = 294$) включают абхазов ($N = 30$), адыгейцев ($N = 10$), балкарцев ($N = 50$), ингушей ($N = 30$), карачаевцев ($N = 22$), мегрелов ($N = 28$), ногайцев ($N = 20$), осетин (С – северные, $N = 30$, Ю – южные, $N = 17$), черкесов ($N = 30$) и чеченцев ($N = 27$). Популяции северных киргизов ($N = 28$), узбеков ($N = 22$), казахов (М – младший жуз, $N = 29$, С – старший жуз, $N = 26$) и дунган ($N = 23$) относятся к популяциям Средней Азии ($N = 128$).

При биоинформатической обработке полученных данных из анализа были исключены SNP с более чем 10% отсутствующих генотипов. Были исключены все позиции делеций и инсерций. Данные были предварительно отфильтрованы по минимальной частоте редкого аллеля (MAF, minor allele frequency > 0.01). В итоге после фильтрации исходных данных по 1779819 маркерам в финальный массив данных вошло 886889 аутосомных SNP. Кроме этого, были исключены образцы, имеющие более 5% пропущенных SNP, а также метисы первого, второго и третьего поколений. Для нахождения участков, которые с большей вероятностью представляют аутозиготность, перед поиском ROH в программе PLINK v1.9 [11] мы также удалили SNP с высоким неравновесием по сцеплению (LD) (англ. LD – linkage disequilibrium). Для этого использовали сканирование генома в рамке из 50 SNP со скользящим окном 5 SNP и

порогом попарного коэффициента сцепления (r^2 порог = 0.8). В результате контроль прошли 585715 SNP. Для анализа ROH в PLINK указали следующие параметры: `-homozyg-snp 50`, `-homozyg-kb 300`. Все остальные параметры использовались по умолчанию. Анализ был выполнен с учетом трех различных наборов ROH с учетом минимальной длины 0.5, 1.5 и 5 млн пн (малые, средние и большие ROH).

Для расчета коэффициента инбридинга F_{ROH} у каждого индивидуума размеры аутосомных областей ROH, превышающих 1.5 млн пн, суммировали и затем делили на общую длину аутосом, за исключением центромер (2708.613716 млн пн для сборки генома hg19). Как показано ранее, F_{ROH} для 1.5 млн пн лучше коррелирует с оценками инбридинга, полученными из анализа родословных [1].

Для оценки возраста ROH применяли программу GARLIC [2], использующую метод, основанный на вероятностной модели для ROH в классах с “коротким” (А, до 0.5 млн пн), “средним” (В, 0.5–1.5 млн пн) и “длинным” (С, более 1.5 млн пн) размерами. Короткие ROH отражают гомозиготность по древним гаплотипам, средние ROH являются результатом недавних демографических изменений в популяции (например, эффект “бутылочного горлышка”) и длинные ROH отражают гомозиготность возникшую в результате недавних кровнородственных браков. Для анализа ROH в GARLIC указали следующие параметры: `-build hg19`, `-error 0.001`, `-winsize 40`, `-auto-winsize`, `-auto-winsize-step 2`, `-kde-subsample 0`, `-ld-subsample 0`, `-resample 40`.

Для межпопуляционных сравнений были рассчитаны средние значения количества и совокупной длины ROH на геном для каждой популяции. Статистический анализ проводился в программной среде R. Статистическую значимость межпопуляционных различий оценивали при помощи теста Вилкоксона. Для коррекции уровня значимости по причине множественных сравнений использовалась поправка Бонферрони (уровень значимости = 0.05).

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Коэффициент инбридинга F_{ROH}

Коэффициенты инбридинга F_{ROH} для всех индивидов из 76 популяционных выборок оценивали отдельно и группировали по географическим регионам и языковым группам для трех классов ROH (длина больше трех пороговых значений 0.5, 1.5 и 5 млн пн). Коэффициенты F_{ROH} для трех диапазонов длин ROH, среднее суммарное количество и средняя суммарная длина гомозиготных участков в исследованных популяциях представлены в табл. 1. Примечательно, что ROH > 1.5 млн пн не было выявлено лишь у четырех индивидов (кумыка, лезгина, эвенка и узбека).

Максимальные значения по $F_{ROH > 1.5}$ (0.0376) и $F_{ROH > 5}$ (0.0239) наблюдаются в популяциях Дагестана (табл. 1). Наибольшим коэффициентом геномного инбридинга для длин ROH > 1.5 млн пн обладают популяции гинухцев и гунзибцев ($F_{ROH} = 0.1039$, $F_{ROH} = 0.0887$). Этот показатель для гинухцев является одним из самых высоких в мире и схож лишь с уровнем инбридинга в изолированной популяции индейцев каритиана из бразильской Амазонии с общей численностью в 320 человек ($F_{ROH > 1.5} = 0.1067$) [3]. Также по уровню гомозиготизации генома значительно выделяется и ряд других малочисленных народов Дагестана — бежтинцы, ахвахцы и цезы (F_{ROH} 0.0689, 0.0678 и 0.0571 соответственно). Эти результаты хорошо соотносятся с предыдущими данными по популяциям Дагестана [9], где для народов нахско-дагестанской языковой семьи были выявлены более высокие коэффициенты инбридинга по сравнению с популяциями из других географических регионов и не нахско-дагестанскими популяциями Дагестана.

Внутри нахско-дагестанской языковой семьи по F_{ROH} этносы разделяются на три группы, в полном соответствии с их языковой классификацией. Максимальный средний уровень F_{ROH} выявлен для популяций дидойской (цезской) группы (0.0727), проживающих в высокогорных регионах в течение сотен поколений в относительной изоляции и имеющих также значительно более низкий уровень разнообразия по гаплогруппам Y-хромосомы, по сравнению с другими северокавказскими популяциями [12]. Для андийской группы показано гораздо меньшее значение (0.0378), находящееся на одном уровне с выборками чукчей, коряков и нивхов (0.0360). Наименьший коэффициент инбридинга выявлен для лезгинской группы (0.0193), занимающей в основном низменные территории и предгорья южной части Дагестана и отличающейся гораздо большей по сравнению с дидойской и андийской группами численностью представляющих ее этносов. Эти результаты полностью согласуются с данными об очень высокой частоте инбредных браков среди малочисленных и более изолированных популяций высокогорных районов по сравнению с равнинными [13].

Среди коренного населения Сибири максимальные значения $F_{ROH > 1.5}$ обнаружены у дальневосточных коряков, чукчей и нивхов (0.0422, 0.0409, 0.0363), что вполне закономерно в связи с их относительной малочисленностью и территориальной изолированностью. Высокое значение показано и для выборки хакасов-сагайцев из предгорного Таштыпского района (0.0318), вдвое превосходящее хакасов-качинцев равнинного Ширинского района. Минимальное значение характерно для томских татар (0.0071). Для остальных популяций наблюдается разброс $F_{ROH > 1.5}$ от 0.0112 до 0.0292. Несмотря на то что F_{ROH} в популяциях Дагестана значительно превышает по средним и длинным

Таблица 1. Значения коэффициентов $F_{\text{РОН}}$ для трех диапазонов длин РОН , среднее суммарное количество (А, В, С) и средняя суммарная длина (а, б, с) для трех классов длин исследуемых популяций

Популяция	Регистр	Группа	PLINK						GARLIC								
			$F_{\text{РОН}}$			$F_{\text{РОН}}$			NROH			SROH			Σ		
			0.5	1.5	5	0.5	1.5	5	А	В	С	а	б	с	NROH	SROH	Σ
Башкиры (Б)	В	1	0.0438	0.0122	0.0043	0.0388	0.0094	0.0035	544	189	14	95	95	31	746	221	
Башкиры (П)	В	1				0.0450	0.0147	0.0065	561	188	17	95	95	45	765	235	
Башкиры (С)	В	1				0.0368	0.0083	0.0034	439	157	10	69	74	24	606	167	
Татары (К)	В	1				0.0334	0.0061	0.0023	575	192	9	100	97	20	776	216	
Чуваши	В	1				0.0406	0.0103	0.0032	443	136	15	79	70	31	594	180	
Бесермяне	В	2	0.0498	0.0169	0.0063	0.0509	0.0199	0.0105	678	202	16	118	110	55	896	283	
Коми	В	2				0.0468	0.0164	0.0084	453	111	13	84	63	43	577	191	
Удмурты	В	2				0.0504	0.0158	0.0047	476	143	19	89	81	46	637	216	
Удмурты (В)	В	2				0.0494	0.0155	0.0037	576	197	24	107	107	52	796	266	
Удмурты (Ш)	В	2				0.0514	0.0171	0.0039	574	205	28	98	105	57	806	259	
Вепсы	В	3	0.0445	0.0107	0.0027	0.0458	0.0126	0.0031	432	145	18	76	75	38	594	188	
Карелы	В	3				0.0432	0.0089	0.0024	453	153	16	80	78	31	622	189	
Марийцы	В	4	0.0437	0.0117	0.0040	0.0496	0.0162	0.0044	450	159	23	81	83	51	632	215	
Мордва (М)	В	4				0.0396	0.0082	0.0026	595	179	10	109	97	24	784	231	
Мордва (Ш)	В	4				0.0460	0.0141	0.0067	635	189	13	109	98	39	837	246	
Мордва (Э)	В	4				0.0396	0.0083	0.0026	596	204	16	100	99	30	816	229	
Русские	В	5	0.0377	0.0061	0.0019	0.0377	0.0061	0.0019	607	189	8	111	102	19	804	231	
Агулы	Д	6	0.0490	0.0193	0.0119	0.0464	0.0171	0.0104	617	163	11	114	93	46	790	253	
Арчинцы	Д	6				0.0782	0.0459	0.0315	547	161	25	106	94	123	733	323	
Лезгины	Д	6				0.0412	0.0124	0.0076	568	156	8	105	88	33	733	226	
Рутульцы	Д	6				0.0433	0.0131	0.0072	445	131	10	78	70	34	586	182	
Табасараны	Д	6				0.0350	0.0066	0.0028	573	196	9	101	99	20	778	220	
Цахуры	Д	6				0.0499	0.0206	0.0120	463	132	14	81	70	54	609	206	
Андийцы	Д	7	0.0696	0.0378	0.0240	0.0578	0.0279	0.0185	577	167	17	101	90	76	761	266	
Ахвахцы	Д	7				0.1020	0.0678	0.0430	424	114	35	90	79	174	572	343	
Багулалы	Д	7				0.0721	0.0400	0.0253	621	171	22	119	103	105	814	328	
Ботлихцы	Д	7				0.0512	0.0196	0.0102	673	230	20	114	117	59	924	290	
Каратинцы	Д	7				0.0775	0.0460	0.0321	547	164	24	101	93	122	735	316	
Тиндинцы	Д	7				0.0569	0.0253	0.0139	536	164	16	95	91	66	715	253	
Чамалалы	Д	7				0.0697	0.0377	0.0246	501	151	23	91	82	102	675	276	

Таблица 1. Продолжение

Популяция	Регин	Группа	PLINK						GARLIC							
			F_{ROH}			F_{ROH}			NROH			SROH			Σ	
			0.5	1.5	5	0.5	1.5	5	A	B	C	a	b	c	NROH	SROH
Бежтинцы	Д	8	0.1126	0.0727	0.0460	0.1105	0.0689	0.0417	423	113	28	101	97	158	563	356
Гинухцы	Д	8				0.1437	0.1039	0.0733	259	71	27	80	85	226	357	391
Гунзибцы	Д	8				0.1305	0.0887	0.0549	318	98	33	85	95	204	449	383
Хваршины	Д	8				0.0773	0.0448	0.0309	594	183	24	108	102	121	801	331
Цезы	Д	8				0.1011	0.0571	0.0293	492	141	31	109	108	137	664	354
Караногайцы	Д	9	0.0318	0.0037	0.0010	0.0324	0.0039	0.0011	582	198	13	99	94	20	793	213
Кумыки	Д	9				0.0312	0.0036	0.0009	548	183	9	100	95	16	739	210
Даргинцы	Д	10	0.0696	0.0369	0.0243	0.0696	0.0369	0.0243	594	143	19	114	87	97	756	297
Лакцы	Д	11	0.0666	0.0347	0.0230	0.0666	0.0347	0.0230	587	179	20	107	100	93	786	300
Аварцы	Д	12	0.0761	0.0442	0.0320	0.0761	0.0442	0.0320	589	178	24	109	100	119	791	328
Абхазы	К	13	0.0335	0.0044	0.0006	0.0350	0.0047	0.0004	519	224	22	86	101	28	764	215
Адыгейцы	К	13				0.0339	0.0048	0.0004	552	193	14	91	91	21	759	204
Черкесы	К	13				0.0317	0.0037	0.0010	589	173	7	106	92	13	769	211
Балкарцы	К	14	0.0345	0.0069	0.0028	0.0357	0.0071	0.0027	514	151	8	100	85	21	674	205
Карачаевцы	К	14				0.0382	0.0104	0.0055	589	172	10	106	92	30	771	228
Ногайцы	К	14				0.0295	0.0030	0.0003	528	239	21	83	102	24	788	209
Ингуши	К	15	0.0396	0.0099	0.0030	0.0392	0.0094	0.0026	521	196	16	91	98	33	733	222
Чеченцы	К	15				0.0399	0.0105	0.0034	550	177	15	99	92	33	741	223
Осетины (С)	К	16	0.0363	0.0066	0.0017	0.0359	0.0063	0.0015	515	220	20	86	101	31	755	217
Осетины (Ю)	К	16				0.0367	0.0069	0.0018	570	212	20	92	97	31	801	220
Мегрелы	К	17	0.0369	0.0049	0.0006	0.0369	0.0049	0.0006	529	202	16	93	100	23	747	216
Алтайцы (Б)	С	18	0.0533	0.0171	0.0089	0.0519	0.0173	0.0086	517	163	14	93	86	46	695	224
Алтайцы (К)	С	18				0.0552	0.0168	0.0100	632	177	13	121	101	47	821	269
Татары (Т)	С	18				0.0338	0.0071	0.0037	483	127	5	86	68	17	615	171
Тувинцы	С	18				0.0591	0.0151	0.0061	731	242	19	139	136	46	992	322
Хакасы (Т)	С	18				0.0664	0.0318	0.0183	585	173	23	111	97	87	781	295
Хакасы (Ш)	С	18				0.0534	0.0146	0.0070	652	199	12	122	112	38	862	273

Таблица 1. Окончание

Популяция	Регион	Группа	PLINK						GARLIC							
			F_{ROH}			F_{ROH}			NROH			SROH			Σ	
			0.5	1.5	5	0.5	1.5	5	A	B	C	a	b	c	NROH	SROH
Буряты (А)	С	19а	0.0656	0.0181	0.0076	0.0599	0.0112	0.0027	691	317	31	116	152	50	1040	318
Буряты (К)	С	19а				0.0615	0.0166	0.0081	733	235	16	135	130	47	983	312
Калмыки	С	19а				0.0465	0.0058	0.0008	688	260	23	121	126	33	970	280
Эвенки (Э)	С	19б				0.0632	0.0195	0.0080	608	195	19	111	110	55	822	276
Эвенки (Я)	С	19б				0.0786	0.0256	0.0105	705	221	23	138	133	69	949	339
Удэгейцы	С	19б				0.0774	0.0246	0.0134	764	219	14	152	136	59	997	347
Якуты	С	19в				0.0723	0.0233	0.0097	725	240	25	137	137	68	990	343
Кеты	С	20а	0.0707	0.0274	0.0127	0.0720	0.0259	0.0121	753	221	21	139	132	70	995	341
Ханты (К)	С	20б				0.0716	0.0280	0.0119	620	201	24	123	125	75	844	323
Ханты (Р)	С	20б				0.0716	0.0266	0.0102	633	212	24	120	126	70	869	316
Чулымцы	С	20в				0.0676	0.0292	0.0167	541	165	22	99	92	79	727	270
Нивхи	С	21а	0.1139	0.0398	0.0133	0.1073	0.0363	0.0119	699	239	28	147	161	84	965	393
Коряки	С	21б				0.1209	0.0422	0.0142	806	301	30	172	210	98	1136	480
Чукчи	С	21б				0.1136	0.0409	0.0137	703	257	32	149	174	100	992	424
Казахи (М)	С. А.	22а	0.0367	0.0055	0.0018	0.0360	0.0040	0.0006	575	198	13	100	98	19	786	217
Казахи (С)	С. А.	22а				0.0358	0.0039	0.0002	511	232	25	83	101	28	768	212
Киргизы	С. А.	22а				0.0414	0.0067	0.0022	589	194	8	106	103	19	791	227
Узбеки	С. А.	22а				0.0334	0.0073	0.0043	549	170	6	97	88	20	725	205
Дунтгане	С. А.	22б	0.0562	0.0116	0.0070	0.0562	0.0116	0.0070	808	218	6	151	126	27	1032	304
	В		0.0438	0.0122	0.0043				535	173	16	94	90	37	723	221
	Д		0.0704	0.0377	0.0239				525	156	20	100	93	96	701	289
	К		0.0357	0.0065	0.0018				543	196	15	94	96	26	755	215
	С		0.0702	0.0229	0.0099				663	218	21	127	127	62	902	316
	С. А.		0.0406	0.0067	0.0028				606	202	12	107	103	23	820	233

Примечание 1. В – Волго-Уральский регион и Европа, К – Кавказ, Д – Дагестан, С – Сибирь, С. А. – Средняя Азия.

Примечание 2. 1 – алтайская/тюркская; 2 – пермская подгруппа; 3 – прибалтийско-финская подгруппа; 4 – волжско-финская подгруппа; 5 – восточнославянская группа; 6 – нахско-дагестанская/лезгинская ветвь; 7 – нахско-дагестанская/андийская группа; 8 – нахско-дагестанская/лидийская группа; 9 – алтайская/тюркская; 10 – нахско-дагестанская/даргинская ветвь; 11 – нахско-дагестанская/лакская ветвь; 12 – нахско-дагестанская/аваро-андийская группа; 13 – абхазо-адыгские языки; 14 – алтайская/тюркская; 15 – нахско-дагестанская/нахская ветвь; 16 – индоевропейская/иранская группа; 17 – картвельская семья; 18 – алтайская/тюркская; 19 – а) алтайская/монгольская; б) алтайская/тунгусо-маньчжурские языки; в) алтайская/тюркская; 20 – а) енисейская; б) уральская/обско-угорская; в) алтайская/тюркская; 21 – а) палеоазиатский язык; б) чукотско-камчатские языки; 22 – а) алтайская/тюркская; б) сино-тибетские языки/китайские языки.

РОН (значения $F_{\text{РОН} > 5}$, $F_{\text{РОН} > 1.5}$), для популяций Сибири по более коротким участкам гомозиготности ($F_{\text{РОН} > 0.5}$) наблюдаются схожие значения. Это может свидетельствовать о значительных эффектах инбридинга для этих популяций Сибири в прошлом с постепенным уменьшением доли кровнородственных браков к настоящему времени.

Популяции Волго-Уральского региона занимают промежуточное положение среди других регионов по коэффициенту инбридинга для всех трех групп гомозиготных регионов. Это можно объяснить как его территориальным положением на месте пересечения путей различных миграционных потоков, так и давней традицией заключения браков лишь с родственниками не ближе пятого–шестого колена, что характерно для башкир, казанских татар, чувашей, коми, марийцев и мордвы. Внутри Волго-Уральского региона для популяций с низкой эффективной численностью и проживающих на периферии наблюдаются более высокие значения уровня гомозиготности генома. Максимальное значение обнаружено у бесермян (0.0199), обособленных территориально и имеющих численность около 2 тыс. человек, минимальное значение — у казанских татар и русских (0.0061).

Популяции Кавказа и Средней Азии обладают минимальными значениями коэффициентов инбридинга, оцененных по РОН, — от 0.0030 до 0.0116 для $F_{\text{РОН} > 1.5}$. Среди популяций Средней Азии максимальные значения наблюдаются у дунган (0.0116) что, вероятно, объясняется их недавней миграцией из Китая с сохранением их языкового, культурного и демографического единства на фоне практически полного отсутствия межнациональных браков. Ранее было показано, что дунгане Узбекистана генетически намного ближе к популяциям Китая и Северо-Восточной Индии, чем к соседним с ними узбекским популяциям, и не имеют с ними недавнего обмена генами [14].

Оценка РОН с помощью PLINK

В отличие от модельного подхода для поиска РОН, этот анализ не зависит от оценки частоты популяционных аллелей. Общая длина и количество РОН для каждого индивидуального генома колеблется от 41.2 до 687.7 млн пн и от 61 до 394 соответственно. Медианное число РОН на индивидуальный геном в исследованных образцах составило 135.4, граница первого квартиля 121.8, граница третьего квартиля 151.5. Медианная длина РОН 128.4 млн пн, граница первого квартиля 103.7 млн пн, третьего 139.3 млн пн. Среднее значение совокупной длины РОН на индивидуума варьирует от 79.9 млн пн в выборке ногайцев до 289.1 млн пн у гинухцев. Минимальное среднее популяционное значение числа РОН на человека наблюдается у узбеков (103.7), максимальное — среди коряков (312).

Доля РОН различной длины значительно колеблется в зависимости от географического реги-

она и языковой группы. Выборки из Дагестана и Сибири, особенно носители языков нахско-дагестанской группы, чукотско-камчатских и нивхского языков, разительно отличаются от таковых из других географических регионов. Среднее значение общей длины РОН на человека в нашей выборке, говорящей на языках дидойской подгруппы, 305.03 млн пн сравнимо со значением 308.6 млн пн у представителей чукотско-камчатских и нивхского языков. Это полностью согласуется с оценкой геномного инбридинга в этих популяциях.

По доле образцов с одним и более РОН длиной больше 5 млн пн по регионам выделяются три различные группы, которые явно связаны с эндогамией и изоляцией: большая часть дагестанских (78.1%) и сибирских (75.5%) образцов имеет длинные РОН, далее идут кавказский, Волго-Уральский регион и Европа (57.9 и 38% соответственно). Для Среднеазиатского региона наблюдается минимальная доля геномов, имеющих длинные РОН (15.6%). Это полностью соответствует полученным ранее данным о наибольшем генетическом разнообразии и многокомпонентности генофонда среднеазиатских популяций в сравнении с другими этносами Северной Евразии. При этом наблюдаются значительные различия между популяциями внутри каждого региона. Например, у казахов старшего жуза всего 7% образцов имеют РОН больше 5 млн пн. Для двух человек из эвенкийской (З) и казахской (С) популяций не было найдено РОН длиннее 1.5 млн пн.

Оценка РОН с помощью GARLIC

Общая длина и количество РОН, идентифицируемые с помощью пакета GARLIC, для каждого индивида колеблется от 124 до 788.6 млн пн и от 319 до 1237 соответственно. Медианное число РОН на индивидуальный геном в исследованных образцах составило 774.5, граница первого квартиля 754.1, граница третьего квартиля 798.5. Медианная длина РОН — 261.4 млн пн, граница первого квартиля 243.9 млн пн, третьего 282.8 млн пн. Среднее значение совокупной длины РОН на индивидуума варьирует от 167.3 млн пн в выборке башкир (С) до 480.2 млн пн у коряков. Минимальное среднее популяционное значение числа РОН на человека наблюдается у гинухцев (356.8) из-за большей длины РОН за счет сильного инбридинга, максимальное — среди коряков (1136).

Средние размеры РОН классов А и В одинаковы среди популяций одного географического региона, за исключением того, что популяции Дагестана и Сибири имеют большую изменчивость (табл. 2). Максимальные значения наблюдаются у дагестанских популяций (гинухцев, гунзибцев, ахвахцев, цезов).

Наибольшие средние значения класса С также наблюдаются у популяций Дагестана и Сибири,

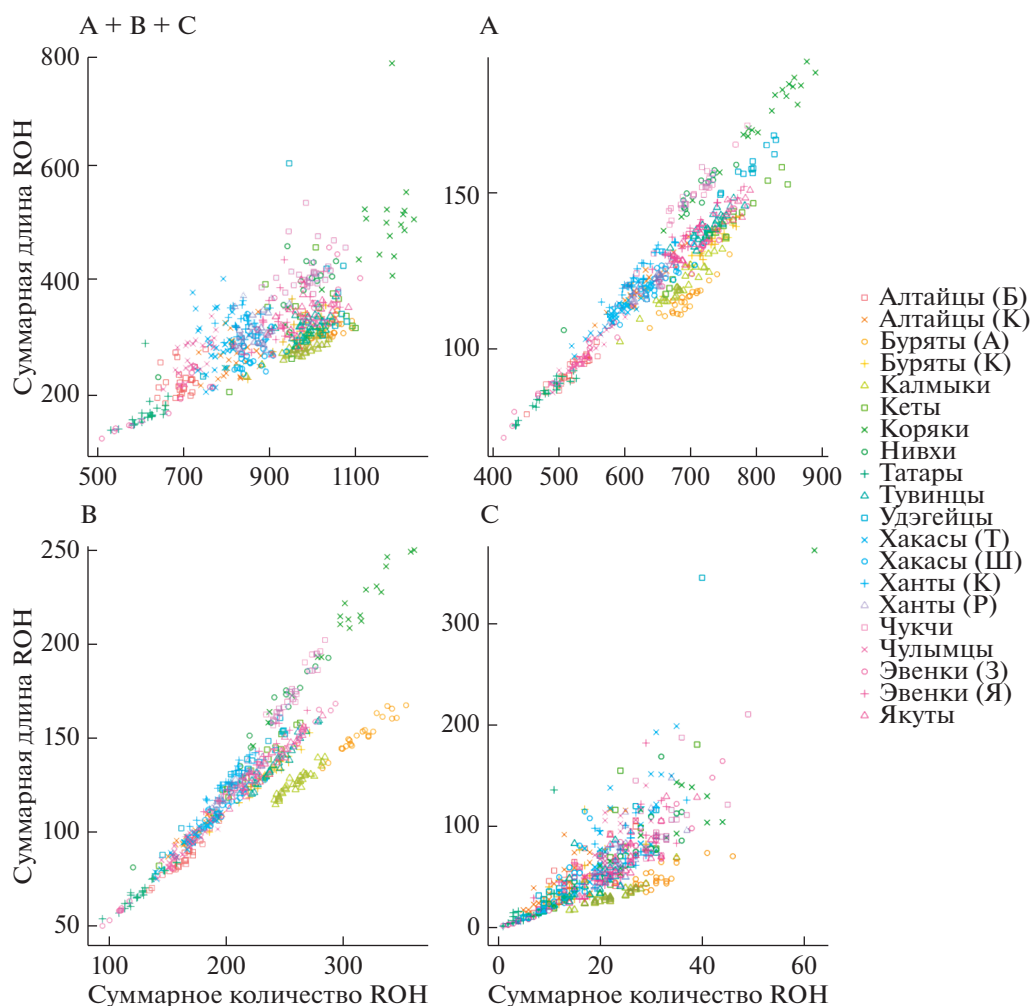


Рис. 1. Отношение суммарной длины ROH (млн пп) к суммарному количеству ROH в исследуемых популяциях Сибири.

далее у популяций Средней Азии, Волго-Уральского региона и Европы, минимальные средние значения – у кавказских популяций. У народов, имеющих смешанное происхождение, – казахов (М, С), бурятов (А), калмыков, караногайцев и кумыков средние размеры ROH аналогичны кавказским популяциям.

Для сибирских чукчей, коряков, нивхов для всех классов длин наблюдается значительное увеличение суммарной длины и количества ROH на человека по сравнению с другими популяциями, что свидетельствует о продолжительном инбридинге, а также эффекте “бутылочного горлышка” (рис. 1). Для трех этих этносов также наблюдается

Таблица 2. Суммарная статистика для значений границ классов длин (тпн) ROH по регионам

Статистика	Сибирь	Средняя Азия	Дагестан	Кавказ	Волго-Уральский регион и Европа	Все
Классы А и В						
Мин.	305.54	287.59	315.23	279.41	301.07	279.41
Макс.	417.37	369.87	668.22	370.51	370.79	668.22
Сред. знач.	366.96	333.15	390.07	321.02	338.75	358.55
SD	7.28	15.00	17.94	8.89	4.35	6.67
Классы В и С						
Мин.	942.65	777.099	981.114	787.342	1010.64	777.10
Макс.	1553.83	1402.82	2598.01	1200.04	1257.98	2598.01
Сред. знач.	1260.15	1115.65	1380.02	999.93	1112.28	1214.99
SD	38.33	120.00	77.19	44.52	20.96	31.39

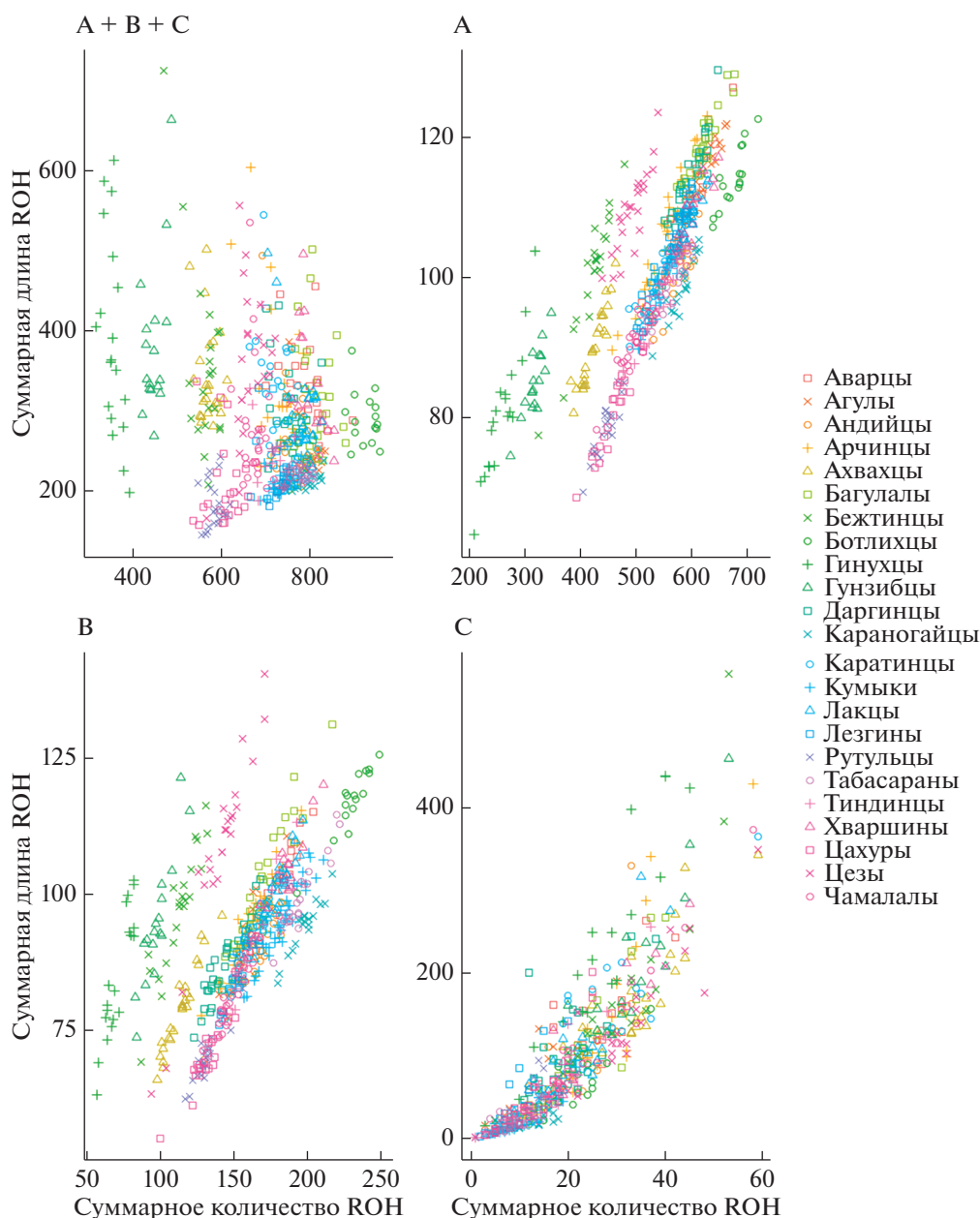


Рис. 2. Отношение суммарной длины ROH (млн пн) к суммарному количеству ROH в исследуемых популяциях Дагестана.

значительное увеличение суммарной длины среднего класса ROH на человека как по сравнению с остальными популяциями, так и по сравнению с классом коротких ROH внутри популяции, что подтверждает продолжительность инбридинга и небольшую эффективную численность популяций. Далее по коротким и средним длинам ROH выделяются кеты, якуты, эвенки (Я), удэгейцы, тувинцы и буряты (К). Для якутов и тувинцев, несмотря на большой эффективный размер популяции, большое количество и длина коротких и

средних ROH вероятно связаны с сильным эффектом основателя, который выявляется на маркерах мтДНК и Y-хромосомы [15, 16]. В популяции эвенков (З) обнаружены образцы с признаками метисации, что может значительно занижать значения по длинам и количеству ROH. При анализе Y хромосомных линий у них наблюдаются европейские гаплогруппы, не характерные для коренных сибирских популяций. При анализе аутосомных блоков, идентичных по происхождению (IBD — англ. identical by descent), у эвенков также наблю-

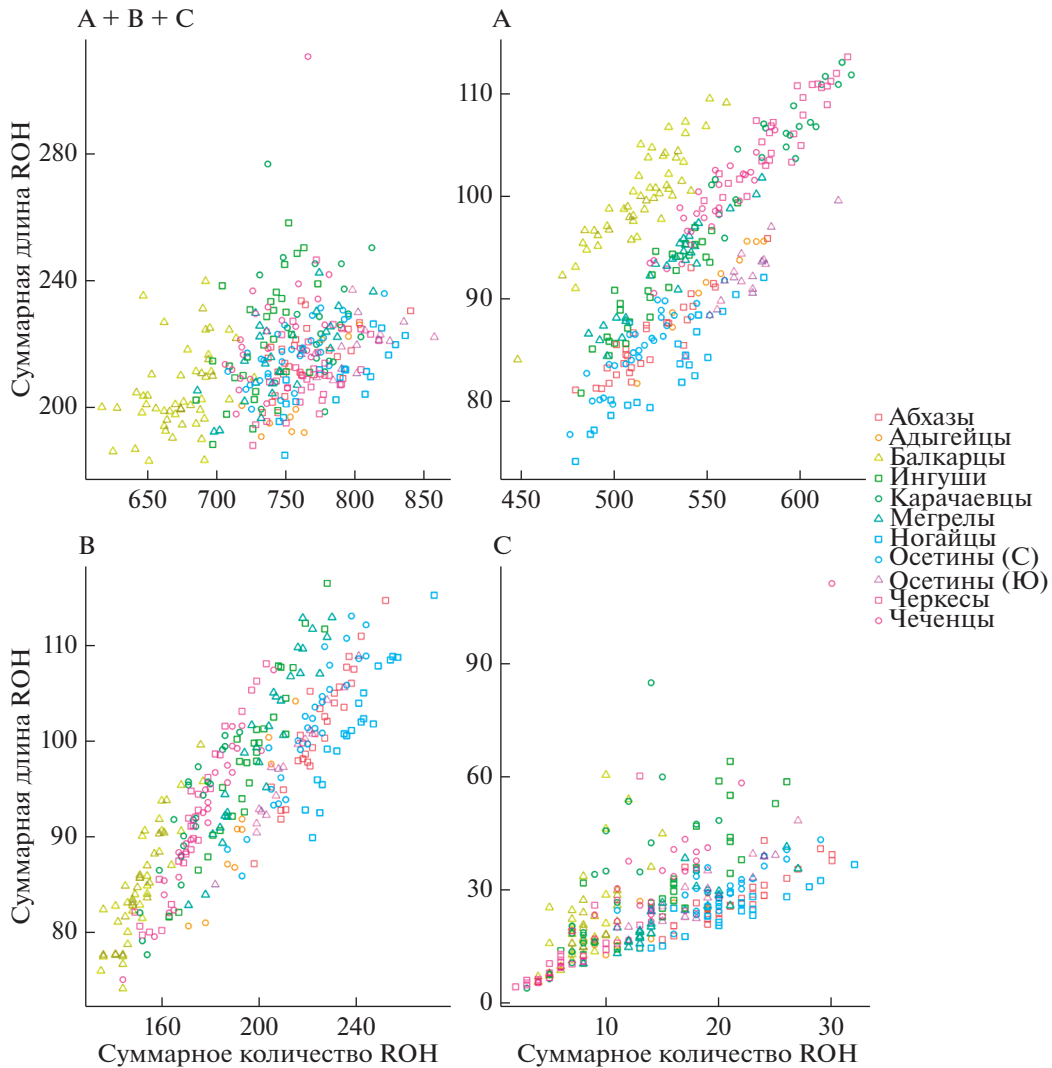


Рис. 3. Отношение суммарной длины ROH (млн пн) к суммарному количеству ROH в исследуемых популяциях Кавказа.

дается значительное количество длинных блоков (4 млн пн), общих с европейскими популяциями, что свидетельствует об относительно недавней метисации этой группы людей с пришлыми европейцами. Для бурятов (А) также характерно увеличение суммарной длины среднего класса ROH, но в большей степени не из-за длины фрагментов, а из-за их количества, что свидетельствует об эффекте “бутылочного горлышка” в прошлом (рис. 1). У бурятов (А) и калмыков наблюдается увеличение количества ROH, но без увеличения суммарной длины, наиболее ярко выражено для ROH средней длины, что свидетельствует о том, что они происходили из довольно небольших полуизолированных сообществ, где эндогамия не была редкостью. По средней суммарной длине и количеству для средних и коротких ROH максимальные

значения наблюдаются у сибирских популяций (табл. 1).

Для популяций гинухцев, гунзибцев, бежтинцев, ахвахцев и цезов для всех классов ROH наблюдается распределение соотношения суммарной длины и количества, характерное для популяций, подверженных сильному инбридингу (рис. 2). Распределение ROH связано с размером популяции и их эффективной численностью, при этом относительно небольшие популяции Дагестана и Сибири имеют ROH большего количества и размера, чем большие по численности популяции этих регионов. Инбридинг добавляет небольшое количество очень длинных ROH для тех, кто является потомком кровнородственных браков, увеличивая дисперсию суммы ROH, видимую как сдвиг влево на графике длин ROH по сравнению с количеством ROH.

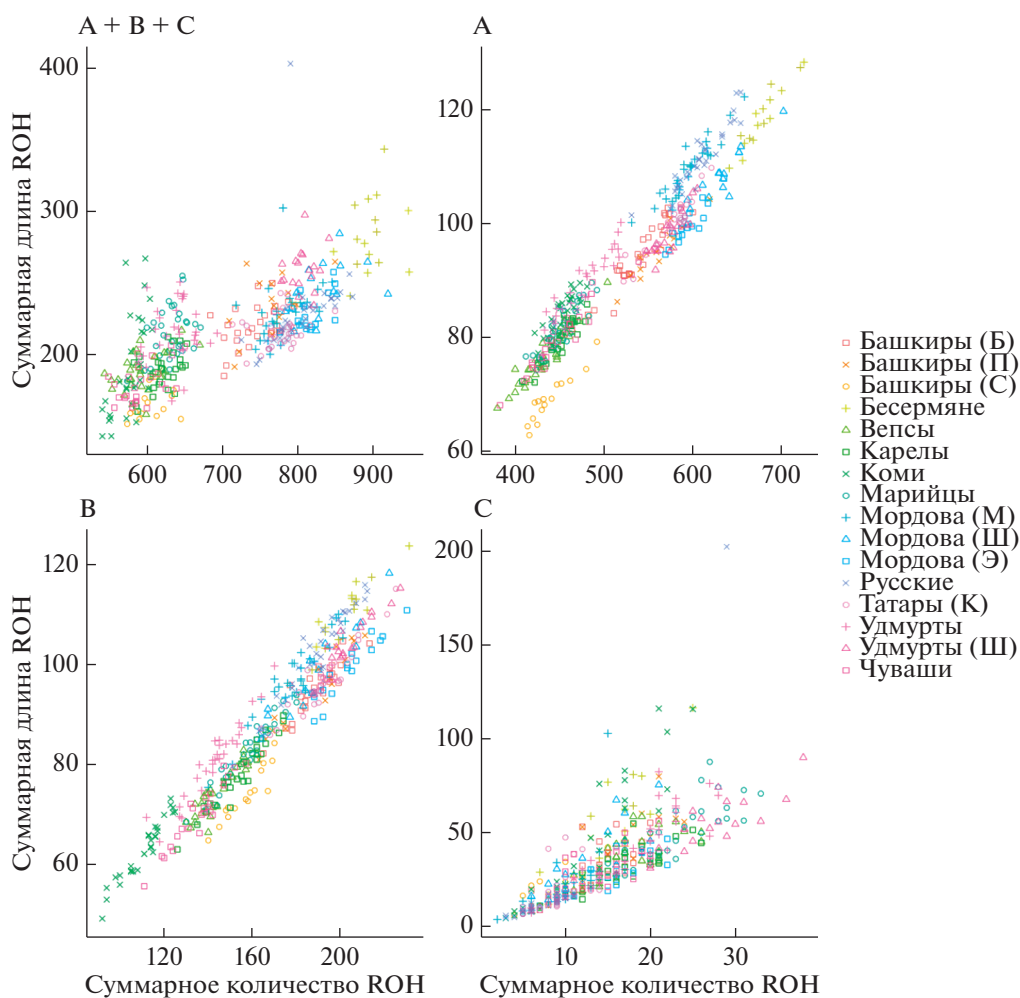


Рис. 4. Отношение суммарной длины ROH (млн пн) к суммарному количеству ROH в исследуемых популяциях Волго-Уральского региона и Европы.

Кавказский регион характеризуется меньшим разбросом по суммарной длине и количеству регионов высокой гомозиготности на индивида, чем в остальных географических регионах. Для длинных ROH наблюдается тенденция к увеличению количества, но не длины, что вероятно объясняется сильной метисацией населения. Среди кавказских популяций выделяется популяция балкарцев, являющаяся одним из самых высокогорных народов Европы и имеющая более длинные короткие ROH, чем у остальных популяций Кавказа, что может свидетельствовать о сильном инбридинге в прошлом (рис. 3).

В популяциях, смешанных по происхождению и характеризующихся высоким уровнем генетического разнообразия, объединяются различные гаплотипы, что обычно значительно снижает количество и длину ROH. Это обнаружено у томских и казанских татар, узбеков и башкир. Для Волго-Уральского региона и Европы наблюдается разделение на две группы по количеству ROH,

где основной вклад вносят короткие и средние ROH. В первую группу входят марийцы, коми, вепсы, чувашы, карелы, удмурты и башкиры (С), во вторую – бесермяне, мордва (М, Ш, Э), удмурты (Ш, Б), татары (К), башкиры (Б, П) (рис. 4).

Эффект “бутылочного горлышка” в демографической истории конкретного этноса увеличивают количество ROH, которые обычно все еще относительно короткие, например популяции калмыков и бурятов (А). У некоторых популяций, имеющих как эффект “бутылочного горлышка”, так и распространенность близкородственных браков, наблюдаются большее количество и длина ROH с увеличением дисперсии суммы ROH (коряки, чукчи и нивхи).

Наибольшими отличиями среди популяций Средней Азии характеризуется популяция дунган, в ней наблюдаются высокие значения по количеству и суммарной длине для средних и коротких ROH и минимальные значения для длинных (рис. 5).

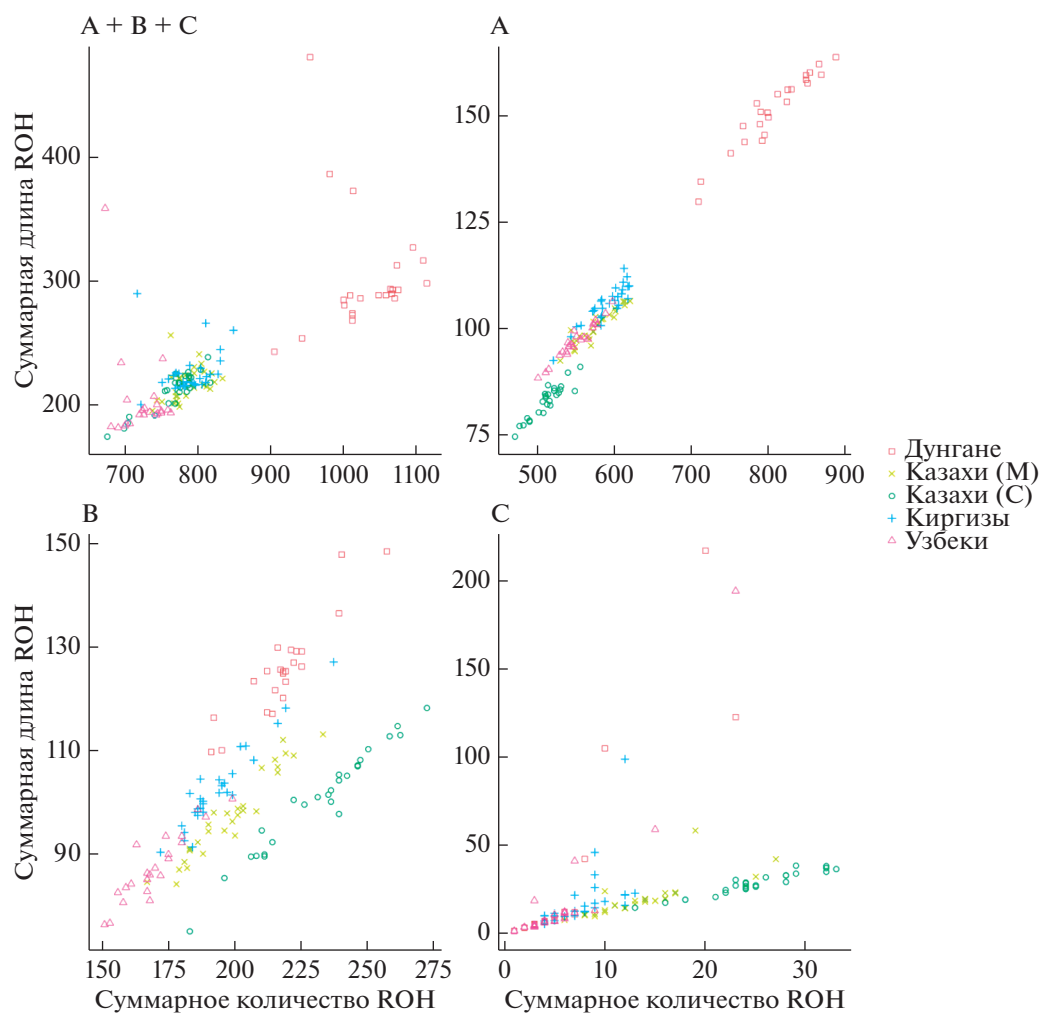


Рис. 5. Отношение суммарной длины ROH (млн пн) к суммарному количеству ROH в исследуемых популяциях Средней Азии.

Доля генома в ROH разной длины и разного количества сильно варьирует по регионам (рис. 6). Наши результаты согласуются с рядом недавних исследований, проведенных с использованием высокоплотных чипов. Уровень гомозиготности по всему геному показывает самый высокий уровень инбридинга в дагестанских и сибирских группах, говорящих на дидойских, чукотко-камчатских и нивхском языках, по сравнению с остальными северо-евразийскими популяционными выборками. Дагестанские и сибирские популяции обладают самыми длинными геномными участками гомозиготности для всех категорий длины ROH и имеют более высокую долю людей с длинными областями аутозиготности.

Высокий коэффициент инбридинга и наличие длинных аутозиготных сегментов согласуются с предыдущими выводами для коэффициентов инбридинга Дагестана [17–19], рассчитанными на основе данных родословных и высокоплотных чипов,

а также с данными по Y-хромосомным маркерам указывающими на эффекты основателя и высокую подразделенность по языковым группам [12]. Аналогично для работ по популяционной генетике сибирских этносов с использованием нейтральных маркерных систем Y-хромосомной, митохондриальной ДНК, показано, что во многих популяциях наблюдается эффект основателя по отдельным гаплогруппам с низким генетическим разнообразием, которое свидетельствует о событиях экспансии численности в относительно недавнее время [15, 20, 21]. Интересно, что полученные нами в настоящей работе оценки существенно не отличаются от тех, которые были получены в предыдущем исследовании нашей группы [10] с использованием гораздо меньшего числа маркеров на меньших выборках сибирских популяций.

Для популяций Сибири характерно большее значение суммарной длины классов А и В на человека, что соответствует данным об увеличении

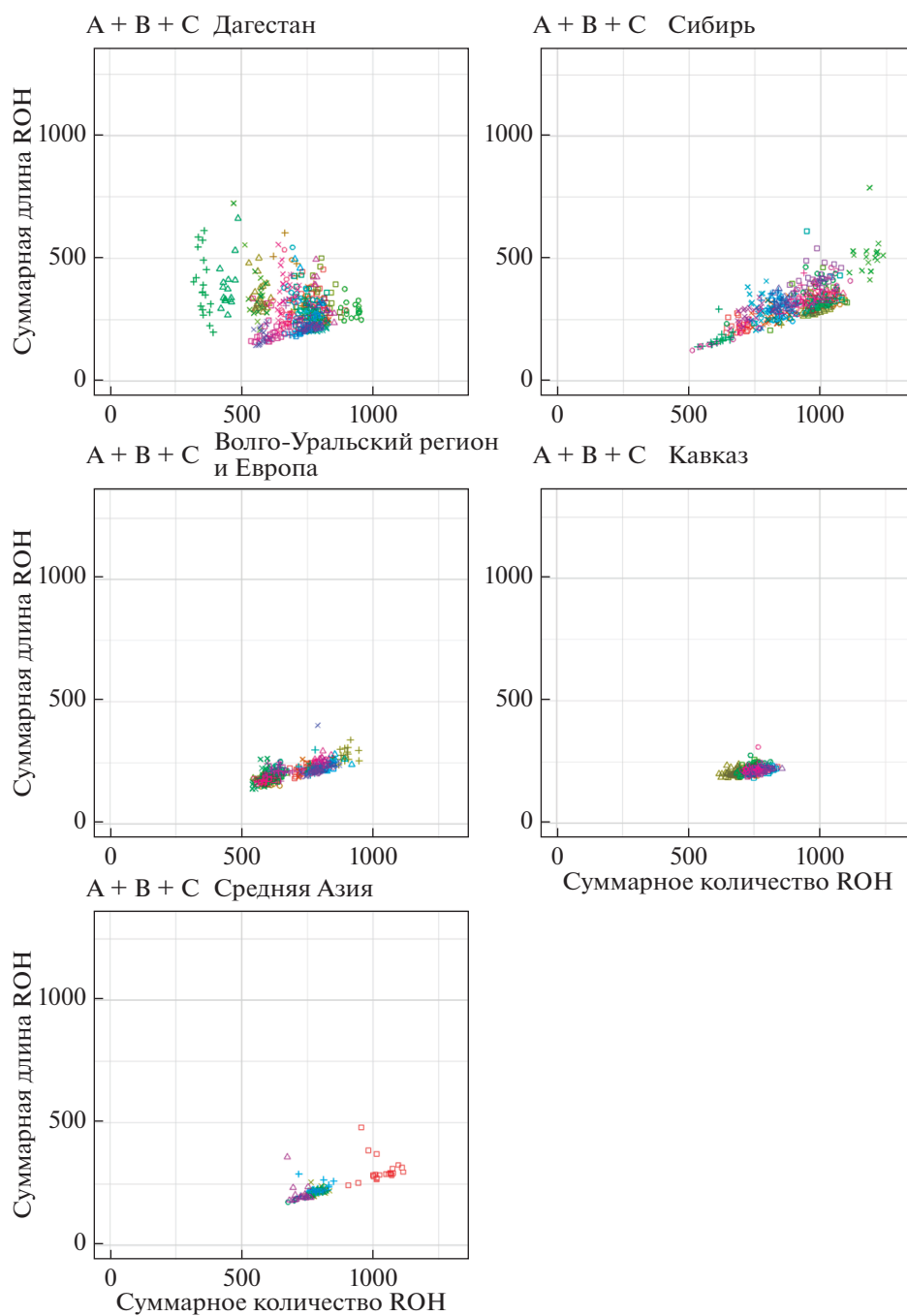


Рис. 6. Отношение суммарной длины ROH (млн пн) к суммарному количеству ROH в исследуемых популяциях для каждого географического региона.

длин классов А и В с удалением популяции от Африки. Эта тенденция похожа на наблюдаемое сокращение разнообразия гаплотипов с увеличением расстояния от Африки. Суммарная длина класса С ROH более изменчива и значительно больше в большинстве популяций Дагестана и Сибири по сравнению с популяциями Кавказа, Средней Азии, Европы и Волго-Уральского региона. Это также свидетельствует о том, что большая доля

популяций из Дагестана и Сибири имеют более высокий уровень кровного родства, это аналогично данным, полученным по значениям коэффициента инбридинга (F_{ROH}).

Результаты проведенного анализа показывают, что генофонд населения Северной Евразии демонстрирует значительные различия в распределении суммы длин и количества по всем классам ROH как между исследуемыми географическими

регионами, этносами, так и внутри локальных популяций. Особенно это характерно для территориально удаленных популяций чукчей и коряков, а также культурно и географически изолированных горных популяций Дагестана. Коренные этносы высокогорных районов Дагестана показывают максимальные значения инбридинга, сохранившегося на высоком уровне до настоящего времени. Большее количество и длина коротких и средних РОН у сибирских популяций указывают на более древний и более продолжительный инбридинг. Получено подтверждение значительной генетической гетерогенности среднеазиатских популяций и относительно низком уровне инбридинга у большинства народов Кавказа и Волго-Уральского региона. Полученные результаты позволяют существенно уточнить имеющиеся данные об особенностях структуры генофонда всех исследованных этносов. Данные о аутозиготных блоках сцепления в конкретных популяциях также могут быть использованы при изучении их компонентного состава предрасположенности к различным наследственным заболеваниям.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-34-90101.

Все процедуры, выполненные в исследовании с участием людей, соответствуют этическим стандартам институционального и/или национального комитета по исследовательской этике и Хельсинкской декларации 1964 г. и ее последующим изменениям или сопоставимым нормам этики.

От каждого из включенных в исследование участников было получено информированное добровольное согласие.

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. *McQuillan R., Leutenegger A.L., Abdel-Rahman R. et al.* Runs of homozygosity in European populations // *Am. J. Hum. Genet.* 2008. V. 83. № 3. P. 359–372. <https://doi.org/10.1016/j.ajhg.2008.08.007>
2. *Szpiech Z.A., Blant A., Pemberton T.J.* GARLIC: Genomic autozygosity regions likelihood-based inference and classification // *Bioinformatics.* 2017. V. 33. № 13. P. 2059–2062. <https://doi.org/10.1093/bioinformatics/btx102>
3. *Lemes R.B., Nunes K., Carnavalli J.E.P. et al.* Inbreeding estimates in human populations: Applying new approaches to an admixed Brazilian isolate // *PLoS One.* 2018. V. 13. № 4. P. e0196360. <https://doi.org/10.1371/journal.pone.0196360>
4. *Ropers H.* New perspectives for the elucidation of genetic disorders // *Am. J. Hum. Genet.* 2007. V. 81. № 2. P. 199–207. <https://doi.org/10.1086/520679>
5. *Kirin M., McQuillan R., Franklin C.S. et al.* Genomic runs of homozygosity record population history and consanguinity // *PLoS One.* 2010. V. 5. № 11. e13996. <https://doi.org/10.1371/journal.pone.0013996>
6. *Pemberton T.J., Absher D., Feldman M.W. et al.* Genomic patterns of homozygosity in worldwide human populations // *Am. J. Hum. Genet.* 2012. V. 91. № 2. P. 275–292. <https://doi.org/10.1016/j.ajhg.2012.06.014>
7. *Ceballos F.C., Joshi P.K., Clark D.W. et al.* Runs of homozygosity: Windows into population history and trait architecture // *Nat. Rev. Genet.* 2018. V. 19. № 4. P. 220–234. <https://doi.org/10.1038/nrg.2017.109>
8. *Frazer K.A., Ballinger D.G., Cox D.R. et al.* International HapMap Consortium A second generation human haplotype map of over 3.1 million SNPs // *Nature.* 2007. V. 449(7164). P. 851–861. <https://doi.org/10.1038/nature06258>
9. *Karafet T.M., Bulayeva K.B., Bulayev O.A. et al.* Extensive genome-wide autozygosity in the population isolates of Daghestan // *Eur. J. Hum. Genet.* 2015. V. 23. № 10. P. 1405–1412. <https://doi.org/10.1038/ejhg.2014.299>
10. *Колесников Н.А., Харьков В.Н., Зарубин А.А., Степанов В.А.* Особенности распределения регионов высокой гомозиготности в геномах представителей коренного населения Северной Евразии // *Генетика.* 2019. Т. 55. № 10. С. 1231–1236. <https://doi.org/10.1134/S0016675819100072>
11. *Purcell S., Neale B., Todd-Brown K. et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses // *Am. J. Hum. Genet.* 2007. V. 81. P. 559–575. <https://doi.org/10.1086/519795>
12. *Глазунова Е.О., Харьков В.Н., Раджабов М.О. и др.* Генофонд коренных народов Дагестана цезковой группы по маркерам Y-хромосомы // *Мед. генетика.* 2016. Т. 15. № 4. С. 29–31. <https://doi.org/10.1234/XXXX-XXXX-2016-4-29-31>
13. *Гаджиев А.Г.* Антропология малых популяций Дагестана. Махачкала: Даг. филиал АН СССР, 1971. 368 с.
14. *Zhabagin M., Balanovska E., Sabitov Z. et al.* The connection of the genetic, cultural and geographic landscapes of transoxiana // *Scientific Reports.* 2017. V. 7. № 1. P. 3085. <https://doi.org/10.1038/s41598-017-03176>
15. *Харьков В.Н., Степанов В.А., Медведева О.Ф. и др.* Происхождение якутов: анализ гаплотипов Y-хромосомы // *Мол. биология.* 2008. Т. 42. № 2. С. 226–237.
16. *Харьков В.Н., Хамина К.В., Медведева О.Ф. и др.* Структура генофонда тувинцев по маркерам Y-хромосомы // *Генетика.* 2013. Т. 49. № 12. С. 1416–1425. <https://doi.org/10.7868/S0016675813120035>
17. *Колесников Н.А., Харьков В.Н., Раджабов М.О. и др.* Различия по уровню инбридинга в популяциях Да-

- гестана: анализ регионов высокой гомозиготности // Мед. генетика. 2020. Т. 19. № 7(216). С. 21–22.
18. *Bulaeva K.B., Davudov O.M., Pavlova T.A. et al.* Genetic subdivision of Dagestan ethnic populations // *Genetika*. 2003. V. 39. № 1. P. 83–92.
19. *Bulayeva K.B.* Overview of genetic-epidemiological studies in ethnically and demographically diverse isolates of Dagestan, Northern Caucasus, Russia // *Croat. Med. J.* 2006. V. 47. № 4. P. 641–648.
20. *Деренко М.В., Малярчук Б.А., Возняк М. и др.* Разнообразие линий Y-хромосомы у коренного населения Южной Сибири // Докл. РАН. 2006. Т. 411. № 2. С. 273–277.
21. *Харьков В.Н., Хамина К.В., Медведева О.Ф. и др.* Генотип бурят: климатическая изменчивость и территориальная подразделенность по маркерам Y-хромосомы // *Генетика*. 2014. Т. 50. № 2. С. 203–213. <https://doi.org/10.7868/S0016675813110088>

Features of the Genomic Distribution of Runs of Homozygosity in the Indigenous Population of Northern Eurasia at the Individual and Population Levels Based on High Density SNP Analysis

N. A. Kolesnikov^{a,*,**}, V. N. Kharkov^a, A. A. Zarubin^a, M. O. Radzhabov^b, M. I. Voevoda^c, M. A. Gubina^d, E. K. Khusnutdinova^e, S. S. Litvinov^e, N. V. Ekomasova^e, O. V. Shtygasheva^f, N. R. Maksimova^g, A. L. Sukhomyasova^g, and V. A. Stepanov^a

^a*Tomsk National Research Medical Center of the Russian Academy of Sciences, Research Institute of Medical Genetics, Tomsk, 634050 Russia*

^b*Amirhanov Institut of Fysics, Dagestan Federal Research Center, Russian Academy of Sciences, Makhachkala, 367012 Russia*

^c*Federal State Budget Scientific Institution "Federal Research Center of Fundamental and Translational Medicine", Novosibirsk, 630060 Russia*

^d*Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, 630060 Russia*

^e*Institute of Biochemistry and Genetics, Ufa Federal Research Center of the Russian Academy of Sciences, Ufa, 450054 Russia*

^f*Katanov Khakass State University, Abakan, 655000 Russia*

^g*Ammosov North-Eastern Federal University, Yakutsk, 677000 Russia*

*e-mail: nikita.kolesnikov@medgenetics.ru

**e-mail: nik.fleming@mail.ru

The genomic data analysis (~886 889 autosomal SNPs) allowed us to estimate the inbreeding level based on the analysis of runs of homozygosity (F_{ROH}) in a sample of 1806 people, including 76 indigenous populations of Siberia, Dagestan, the Caucasus, Volga-Ural region, Eastern Europe and Central Asia. Data was obtained using biochips Infinium Multi-Ethnic Global-8 Kit. In the populations of Dagestan, within the Nakh-Dagestani language family, representatives of several language groups can be distinguished, with the maximum F_{ROH} level for the Dido group (0.0727) and the Andean group (0.0378), which is similar in values to the populations speaking Chukchi-Kamchatka and Nivkh languages of Siberia (0.0360). The Siberian populations are characterized by a greater value of the total length of short and medium regions of runs of homozygosity per person. The total length of long ROHs is more variable and much larger in the most of Dagestan and Siberia populations, compared to the populations of the Caucasus, Central Asia, Europe, and the Volga-Ural region.

Keywords: population genetics, ROH, runs of homozygosity, human populations, inbreeding, North Eurasia.