

УДК 575.113:575.8

ПАТТЕРНЫ АМИНОКИСЛОТНЫХ ЗАМЕН В БЕЛКАХ Е6 И Е7 ВИРУСА ПАПИЛЛОМЫ ЧЕЛОВЕКА 16 ТИПА: ФИЛОГЕОГРАФИЯ И ЭВОЛЮЦИЯ

© 2024 г. Е. Е. Зеленова^{a, b, *}, А. А. Карлсен^{c, d}, Д. В. Авдошина^e, К. К. Кюрегян^{c, d},
М. Г. Беликова^{c, e, f}, И. Д. Троценко^{c, g}

^aИнститут молекулярной биологии им. В. А. Энгельгардта Российской академии наук, Москва, 119991 Россия

^bНациональный медицинский исследовательский центр онкологии им. Н. Н. Блохина, Москва, 115478 Россия

^cРоссийский университет дружбы народов, Москва, 117198 Россия

^dИнститут вакцин и сывороток им. И. И. Мечникова, Москва, 105064 Россия

^eФедеральный научный центр исследований и разработки иммунобиологических препаратов им. М. П. Чумакова
Российской академии наук, Москва, 117218 Россия

^fНаучный центр эпидемиологии и микробиологии им. Н. Ф. Гамалеи
Министерства здравоохранения Российской Федерации, Москва, 123098 Россия

^gМосковское онкологическое общество, Москва, 111123 Россия

*e-mail: zelenovayeye@gmail.com

Поступила в редакцию 05.10.2023 г.

После доработки 28.12.2023 г.

Принята к публикации 29.12.2023 г.

Белки Е6 и Е7 вируса папилломы человека (ВПЧ) играют ключевую роль в онкогенезе папилломавирусной инфекции. Данные по изменчивости этих онкобелков ограничены, а факторы, влияющие на их изменчивость, малоизучены. Нами проанализирована вариабельность известных к настоящему времени последовательностей белков Е6 и Е7 ВПЧ типа 16 (ВПЧ16) с учетом их географического происхождения и года сбора образцов, а также направления их эволюции в основных географических регионах мира. Из базы данных GenBank NCBI 6 октября 2022 года были выгружены все последовательности, относящиеся к фрагментам генома ВПЧ16, кодирующим онкобелки Е6 и Е7. Образцы отфильтровали по следующим параметрам: последовательность включает хотя бы одну из двух интересующих рамок считывания целиком, известны дата сбора и страна происхождения. Всего отобрали 3651 полногеномную нуклеотидную последовательность, кодирующую белок Е6, и 4578 полногеномных нуклеотидных последовательностей, кодирующих белок Е7. Полученные после выборки и выравнивания нуклеотидные последовательности переводили в аминокислотные и анализировали с применением программ MEGA11, R, RStudio, Jmodeltest 2.1.20, BEAST v1.10.4, Fastcov и Biostrings. Наибольшая изменчивость первичной структуры белка Е6 зарегистрирована в позициях 17, 21, 32, 85 и 90 а.о., а для Е7 – в позициях 28, 29, 51 и 77 а.о. Выборка разделилась по географическому признаку на 5 гетерогенных групп: африканскую, европейскую, американскую, регион Юго-Западной и Южной Азии и регион Юго-Восточной Азии. Для ряда стран были выявлены уникальные аминокислотные замены (АК-замены) в белках Е6/Е7 ВПЧ16, предположительно, характерные для отдельных этнических групп. В основном они локализованы в сайтах известных В- и Т-клеточных эпитопов и относительно редко в структурно-функциональных доменах. Выявленные различия в АК-заменах у различных этнических групп и их колокализация с кластерами В- и Т-клеточных эпитопов позволяют предположить их возможную взаимосвязь с географическим распределением аллелей и гаплотипов главного комплекса гистосовместимости (HLA). Это может приводить к распознаванию различного набора В- и Т-клеточных эпитопов вируса, что обуславливает региональные различия в направлении эпитопного дрейфа. В результате филогенетического анализа нуклеотидных последовательностей, кодирующих белок Е6 ВПЧ16, выявлен общий предок, подтверждена региональная кластеризация последовательностей гена белка Е6 по набору наиболее распространенных АК-замен, а также обнаружены случаи реверсии отдельных АК-замен при изменении региона распространения вируса. Для белка Е7 аналогичный анализ был невозможен из-за высокой гомологии последовательностей. Ковариационный анализ совокупной выборки выявил отсутствие взаимосвязи между аминокислотными остатками в белке Е6, в белке Е7, а также между Е6 и Е7. Полученные в работе данные важны для разработки терапевтических вакцин против ВПЧ высокого канцерогенного риска.

Ключевые слова: вирус папилломы человека, белок Е6, белок Е7, первичная структура, эволюция, филогенетический анализ, ковариационный анализ, В-клеточные эпитопы, Т-клеточные эпитопы

DOI: 10.31857/S0026898424040036, EDN: INFCZJ

Сокращения: АК-замены – замены аминокислотных остатков;
ВПЧ – вирус папилломы человека; РШМ – рак шейки матки.

Вирус папилломы человека (ВПЧ) — ДНК-содержащий вирус диаметром 45–55 нм, семейства Papillomaviridae. На сегодняшний день известно более 200 штаммов папилломавирусов, патогенных для человека, которые можно разделить на три основные группы по степени онкогенного риска [1]:

- неонкогенные типы (1, 2, 3 и 5);
- типы низкого канцерогенного риска (6, 11, 42, 43 и 44);
- типы высокого канцерогенного риска (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 73 и 82).

Штаммы ВПЧ высокого канцерогенного риска относятся к основным этиологическим агентам развития рака шейки матки (РШМ), а также других злокачественных новообразований: ротоглотки, полового члена, анальной области, вульвы и влагалища [2]. Продукты трансляции ранних генов: онкобелки Е5, Е6 и Е7 – вовлечены в пролиферацию и малигнизацию инфицированных клеток.

Белок Е6 принимает непосредственное участие в трансформации инфицированных клеток в неопластические. Он ингибирует остановку клеточного цикла и апоптоз путем блокады активности главного онкосупрессора — белка р53, — способствует потере полярности клеток и инициирует гиперплазию [3]. Белок Е6 также подавляет интерфероновый ответ организма через взаимодействие с IRF3 (interferon regulatory factor 3) [4].

Онкобелок Е7 способствует дестабилизации генома и накоплению хромосомных аберраций. Взаимодействие белка Е7 с IRF1 (interferon regulatory factor 1) приводит к уклонению инфицированных клеток от интерфероновой реакции иммунной системы организма и тем самым поддерживает длительную персистенцию ВПЧ [5], что считается ведущим звеном в ВПЧ-ассоциированном канцерогенезе. Также белок Е7 ингибирует один из главных онкосупрессоров — белок RB1 (retinoblastoma 1), — способствуя бесконтрольному делению инфицированных клеток [6].

В отличие от капсидных белков L1 и L2, онкобелки Е6 и Е7 высококонсервативны [7]. Тем не менее, согласно ряду исследований, определенные аминокислотные замены (АК-замены) в этих белках ассоциированы с хронической персистенцией ВПЧ и последующей малигнизацией инфицированных клеток [8–10].

В силу их роли в ВПЧ-ассоциированном канцерогенезе онкобелки Е6 и Е7 рассматривают как основу для разработки терапевтических вакцин против ВПЧ высокого канцерогенного риска, поэтому большое значение имеет информация относительно их вариативности и ее определяющих параметров.

Нами проведен анализ глобальной изменчивости первичной структуры онкобелков Е6 и Е7 ВПЧ типа 16 (ВПЧ16) и направлений их эволюции

с привязкой к основным географическим регионам мира. Выбор ВПЧ16 обусловлен тем, что он преобладает среди генотипов высокого канцерогенного риска и ассоциирован с развитием ряда злокачественных опухолей. Кроме того, полногеномные последовательности ВПЧ16 широко представлены в международных базах данных, что важно для корректного филогенетического анализа.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Отбор образцов из баз данных. Из международной базы GenBank NCBI (<https://www.ncbi.nlm.nih.gov/nucleotide>) 6 октября 2022 года были выгружены все имеющиеся на тот момент фрагменты генома ВПЧ16, кодирующие белки Е6 и Е7. С использованием программы MEGA11 (<https://www.megasoftware.net/>) были отобраны образцы, содержащие полную нуклеотидную последовательность в пределах хотя бы одной из двух интересующих нас рамок считывания (ORF). Полученные после выборки и выравнивания нуклеотидные последовательности были переведены в аминокислотные в программе MEGA11. Оценку стабильности генома проводили на основе сравнения с референсными последовательностями белков Е6 и Е7 ВПЧ16: NP_041325.1 (GeneID: 1489078) и NP_041326.1 (GeneID: 1489079) соответственно.

Филогенетический анализ. Для выявления популяционноспецифических АК-замен ВПЧ16 из исследуемой выборки были отобраны образцы с известной датой сбора биоматериала и страной происхождения. Всего была выгружена 3651 последовательность гена *Е6* и 4578 последовательностей гена *Е7*. Для филогенетического анализа базы образцов проверяли на избыточность и исключали все последовательности, идентичность которых превышала 99.8%. Для построения филогенетических деревьев для гена *Е6* было отобрано 356 нуклеотидных последовательности, для *Е7* — 143.

Для оценки временного сигнала полученных выборок использовали программу TempEst (<https://www.beast.community/tempest>). Параметры для расчета модели замен для построения филогенетических деревьев выбирали с помощью программы Jmodeltest 2.1.20 (<https://www.mybiosoftware.com/tag/jmodeltest>). Оптимальным оказался вариант GTR (General Time Reversible) с Gamma, равным 4. Для построения филогенетического дерева использовали программы из пакета BEAST v1.10.4 (https://www.beast.community/2018-11-14.10.4_released.html). Вначале были сгенерированы тестовые запуски, позволившие на основании АИС-критерия (Akaike Information Criterion) выбрать оптимальный вариант комбинации параметров, а именно: Fixed local clock, Skyline Growth. Длина MCMC (Markov chain Monte Carlo) составила 100 млн генераций. Для дальнейшей визуализации использована

программа FigTree-1.4.4 (<https://www.mybiosoftware.com/figtree-1-3-1-produce-figures-phylogenetic-trees.html>). Выделение кластеров для последующего анализа было основано на принципе гомологии нуклеотидных последовательностей.

Ковариационный анализ. Для анализа ковариаций использовали пакет Fastcov (<http://www.yanlilab.github.io/fastcov>), а также пакет Biostrings (<https://www.doi.org/10.7554/eLife.08932.001>).

Данные по локализации В- и Т-клеточных эпитопов белков Е6 и Е7 были получены из международной базы данных открытого доступа IEDB Analysis Resource (<http://www.tools.iedb.org>, даты загрузки — 02.07.2019 и 12.10.2022). Эпитопные карты строили в программе JalView 2.11.14 (<https://www.jalview.org/>).

Статистический анализ. Для выявления статистически значимых различий между группами сравнения применяли двусторонний тест Фишера. Различия считались значимыми при $p < 0.05$. Анализ проводили с помощью языка R с визуализацией через Rstudio (<https://www.posit.co/download/rstudio-desktop/>).

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Общая характеристика вариабельности онкобелков Е6 и Е7

В общей сложности нами проанализирована 3651 нуклеотидная последовательность, кодирующая белок Е6 ВПЧ16, из 29 стран и 4578 нуклеотидных последовательностей, кодирующих белок Е7, из 23 стран. Все страны оказались распределены по 5 регионам мира: Европе, Африке, Америке, Юго-Западной и Южной Азии, а также Юго-Восточной Азии (рис. 1).

Из 158 аминокислотных остатков (а.о.) Е6 39 оказались консервативными во всех анализируемых последовательностях, хотя в кодирующих их триплетах нуклеотидов встречались синонимичные замены. В остальных 119 а.о. белка Е6 выявлены АК-замены.

Из 99 а.о. Е7 43 оказались консервативными (100%) во всех анализируемых последовательностях, при этом в соответствующей им нуклеотидной последовательности встречались синонимичные замены. Остальные 56 кодонов содержали замены нуклеотидов, приводящие к АК-заменам. Частота наиболее распространенных АК-замен в белках Е6 и Е7 ВПЧ16 представлена на рис. 2а и 2б соответственно.

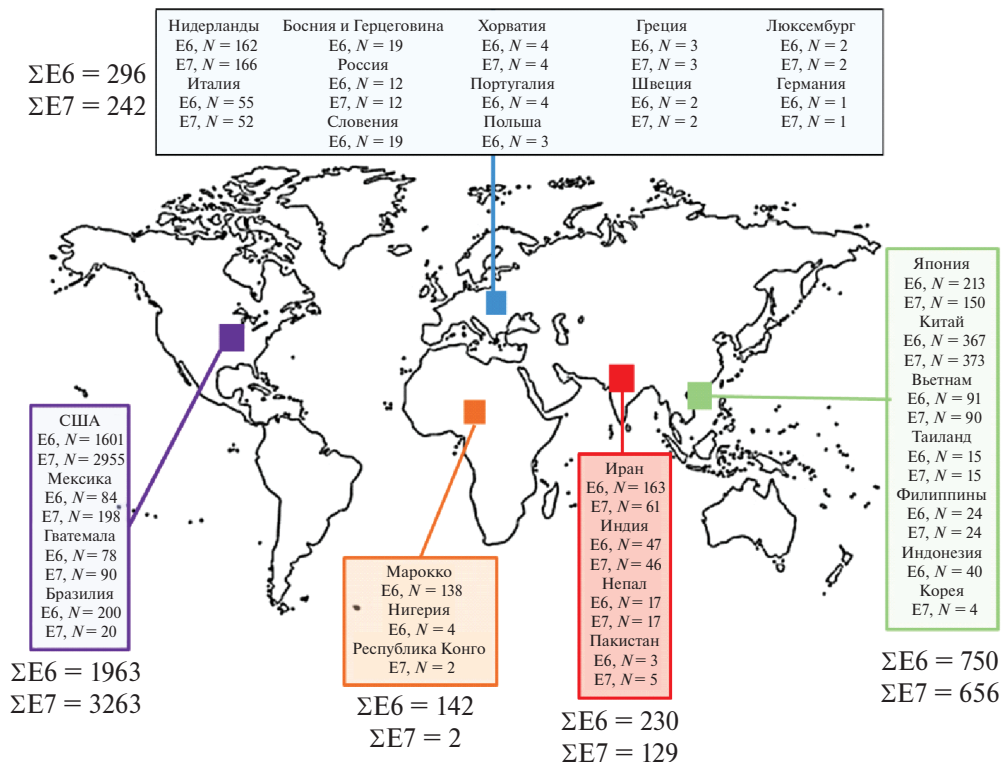


Рис. 1. Распределение отобранных образцов последовательностей генов белков Е6 и Е7 ВПЧ16 по географическим регионам: Европе (серая рамка), Африке (оранжевая рамка), Америке (фиолетовая рамка), Юго-Восточной Азии (зеленая рамка), Южной и Юго-Западной Азии (красная рамка). N – число последовательностей Е6 и Е7, депонированных в базе GenBank, с указанием представительства отдельных стран в каждом из регионов.

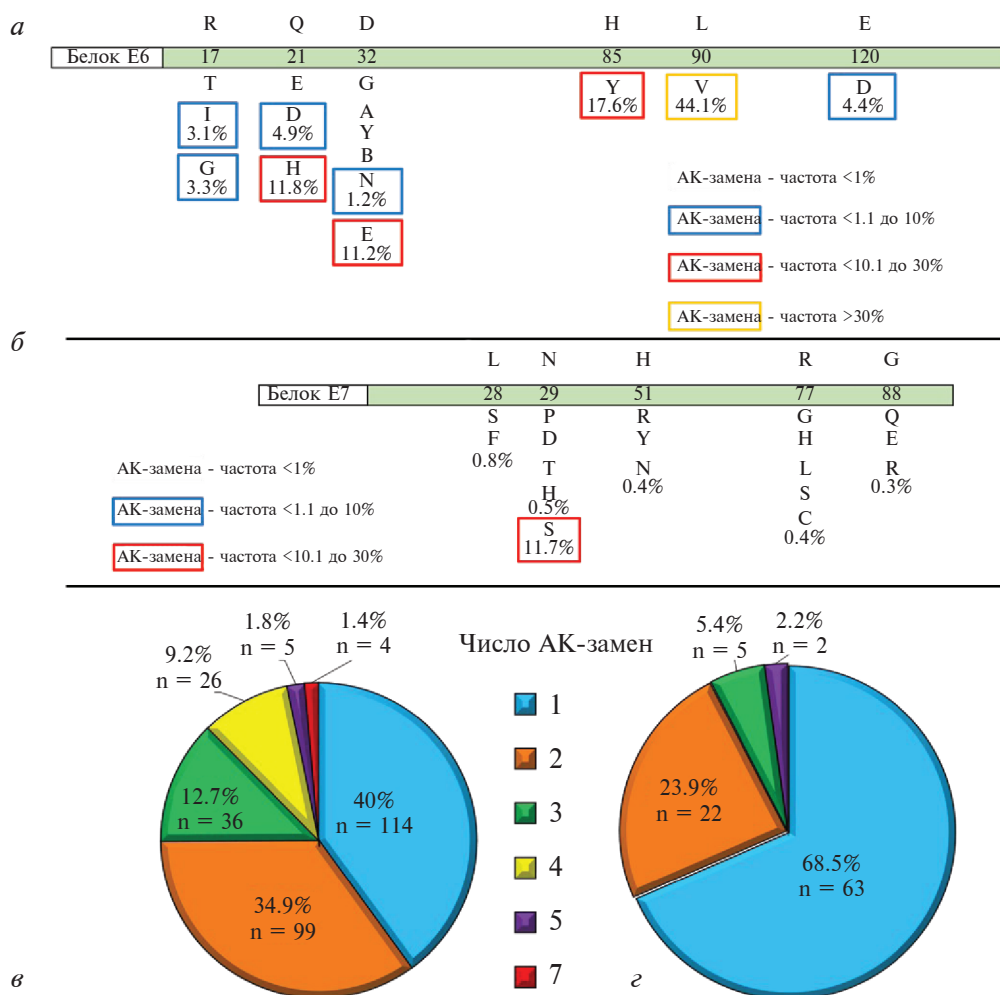


Рис. 2. Изменчивость онкобелков E6 и E7 ВПЧ16. Аминокислотные замены, встречающиеся в >1% последовательностей E6 (а) и E7 (б); распределение образцов по числу АК-замен в обогащенных выборках белка E6 (284 образца из 356 последовательностей) (в) и белка E7 (92 образца из 132 последовательностей) (г). Обогащенные выборки исключали последовательности, идентичность которых превышала 99.8%. Цветовое обозначение числа АК-замен общее для панелей (в) и (г).

В 170 из 356 (47.8%) анализируемых образцов E6 выявлено не менее двух АК-замен (рис. 2в). Заметим, что среди последовательностей одновременно с двумя АК-заменами наиболее частым было их сочетание в позициях 21 и 85, а также различные тандемы с аминокислотой валин в позиции 90. Стоит отметить, что сочетание замен в позициях 85 и 90 встречалось крайне редко. Среди трех и более замен чаще встречалась комбинация Q21H, N85Y и L90V.

Для последовательностей белка E7 подобных сочетаний не выявлено, в большинстве образцов АК-замены были единичные (рис. 2г).

Вариабельность по отдельным географическим регионам

На втором этапе был проведен анализ нуклеотидных последовательностей для выявления географических и/или популяционных особенностей

белков E6 и E7 ВПЧ16 на основании образцов, для которых была известна страна и год сбора. Распределение АК-замен в популяции указывало на кластеризацию последовательностей белков E6 и E7 ВПЧ16 по географическому признаку. Выявлено, что характер вариабельности различался в 5 географических группах: страны Европы (EU), Африки (AF), Америки (AM), Юго-Западной и Южной Азии (AS1) и Юго-Восточной Азии (AS2). Значимыми АК-заменами для региона считали те, частота встречаемости которых в данной позиции достоверно превышала таковую в других позициях этой выборки.

Так, в странах Африки (AF) в белке E6 в 52.8% случаев в позиции 90 преобладала замена L→V ($p = 0.0306$). Частыми в этой географической группе были также замены N85Y (48.6%), E120D (30.3%), Q21D (30.9%) и R17I (23.9%) ($p < 0.0001$ для всех вариантов; табл. 1, рис. 3).

Таблица 1. Частота встречаемости АК-замен в белке Е6 изолятов ВПЧ16 из стран Африки (выборка АF, N = 142)

АК-позиция ^а	АК-замены ^б					
	Вариант 1		Вариант 2		Вариант 3	
	АК	n (%)	АК	n (%)	АК	n (%)
R17	G	9 (6.3%)	I	34 (23.9%)	T	2 (1.4%)
Q21	D	44 (31%)	E	2 (1.4%)	H	9 (6.3%)
A68	G	3 (2.1%)	–	–	–	–
D71	E	11 (7.7%)	–	–	–	–
H85	Y	69 (48.6%)	–	–	–	–
L90	V	75 (52.8%)	–	–	–	–
E120	D	43 (30.3%)	–	–	–	–

Здесь и далее: ^аВ качестве референсной аминокислотной последовательности белка Е6 выбрана NCBI Reference Sequence NP_041325.1.

^бТаблица не включает АК-замены, встречающиеся один раз. Приведены варианты АК-замен, их число (n) и частота встречаемости (%). АК-замены, встречающиеся в >10% последовательностей, выделены жирным шрифтом.

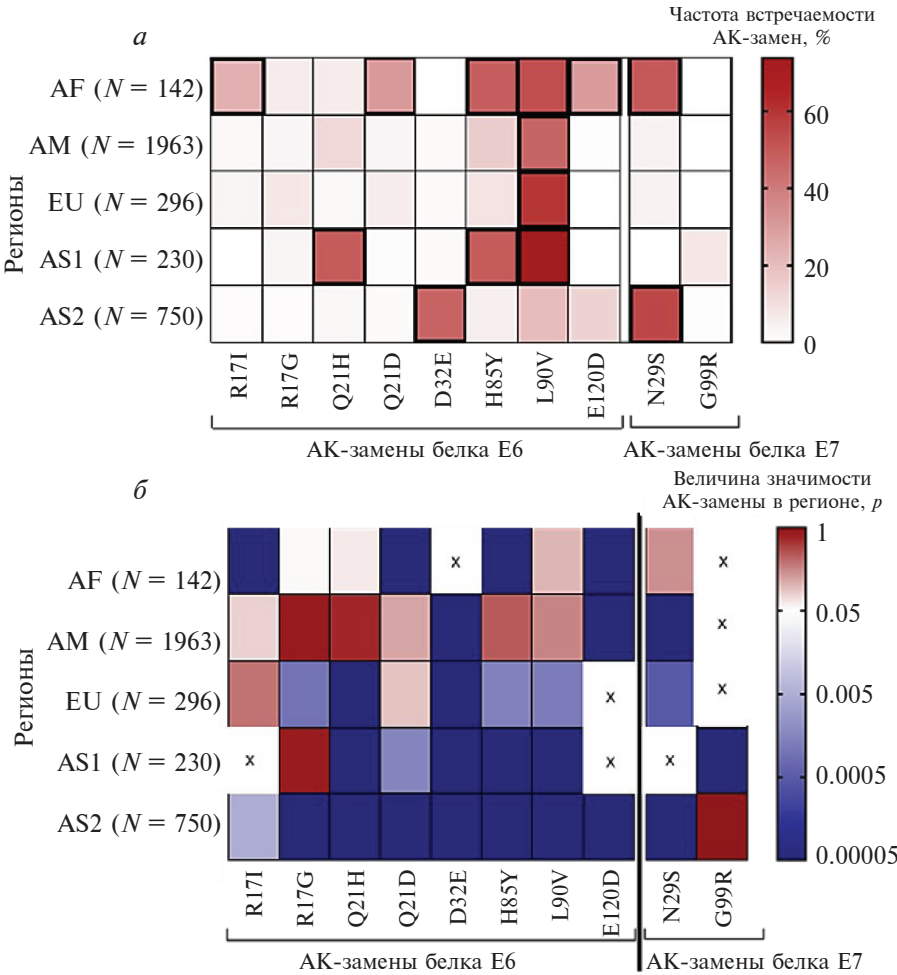


Рис. 3. Тепловые карты встречаемости АК-замен в пяти основных географических регионах мира. Карта распределения частот встречаемости: паттерн АК-замен в белках Е6 и Е7, цветовая шкала справа отражают частоту встречаемости замены для региона; обведены ячейки с частотой встречаемости >20% (а). Карта значимости АК-замены в белках Е6 и Е7 для региона при попарном сравнении с частотой встречаемости данной замены в общей популяции. Цветовая шкала справа отражает величину значимости, *p*. Сравнение проводили с использованием критерия Фишера (б). Тепловые карты сделаны с помощью языка R с визуализацией через Rstudio. Отметка “х” означает отсутствие АК-замены в регионе, поэтому сравнение не проводили.

Таблица 2. Частота встречаемости АК-замен в белке Е6 изолятов ВПЧ16 из США и стран Центральной и Южной Америки (выборка АМ, N = 1963)

АК-позиция	АК-замены									
	Вариант 1		Вариант 2		Вариант 3		Вариант 4		Вариант 5	
	АК	n (%)	АК	n (%)	АК	n (%)	АК	n (%)	АК	n (%)
M1	L	23 (1.17%)	–	–	–	–	–	–	–	–
K4	E	2 (0.1%)	N	1 (0.05%)	Q	1 (0.05%)	–	–	–	–
T6	P	4 (0.2%)	–	–	–	–	–	–	–	–
Q10	E	2 (0.1%)	–	–	–	–	–	–	–	–
R17	G	66 (3.36%)	I	43 (2.2%)	T	19 (0.97%)	–	–	–	–
K18	T	4 (0.2%)	–	–	–	–	–	–	–	–
Q21	H	228 (11.6%)	D	66 (3.36%)	–	–	–	–	–	–
Q27	R	3 (0.15%)	L	1 (0.05%)	–	–	–	–	–	–
H31	N	5 (0.25%)	Y	4 (0.2%)	–	–	–	–	–	–
D32	E	27 (1.4%)	N	22 (1.12%)	H	3 (0.15%)	A	1 (0.05%)	G	1 (0.05%)
I34	R	33 (1.68%)	L	10 (0.5%)	T	2 (0.1%)	K	1 (0.05%)	–	–
L35	V	3 (0.15%)	–	–	–	–	–	–	–	–
E36	Q	18 (0.9%)	K	1 (0.05%)	G	1 (0.05%)	A	1 (0.05%)	–	–
R47	H	2 (0.1%)	C	1 (0.05%)	–	–	–	–	–	–
E48	K	3 (0.15%)	–	–	–	–	–	–	–	–
D51	N	6 (0.3%)	–	–	–	–	–	–	–	–
I59	V	11 (0.56%)	–	–	–	–	–	–	–	–
D63	H	1 (0.05%)	Y	1 (0.05%)	–	–	–	–	–	–
N65	S	2 (0.1%)	–	–	–	–	–	–	–	–
D71	N	6 (0.3%)	E	1 (0.05%)	–	–	–	–	–	–
S78	C	5 (0.25%)	–	–	–	–	–	–	–	–
H85	Y	298 (15.2%)	–	–	–	–	–	–	–	–
L90	V	904 (46.1%)	–	–	–	–	–	–	–	–
E120	D	3 (0.15%)	–	–	–	–	–	–	–	–
K128	E	2 (0.1%)	–	–	–	–	–	–	–	–
R136	K	2 (0.1%)	T	1 (0.05%)	–	–	–	–	–	–
R138	L	2 (0.1%)	–	–	–	–	–	–	–	–
S145	C	1 (0.05%)	F	1 (0.05%)	–	–	–	–	–	–
C147	S	2 (0.1%)	–	–	–	–	–	–	–	–
R148	T	3 (0.15%)	I	1 (0.05%)	–	–	–	–	–	–
Q157	R	1 (0.05%)	H	1 (0.05%)	–	–	–	–	–	–

В США и странах Центральной и Южной Америки (АМ) наиболее распространенными в белке Е6 были замены L90V (46.1%, $p = 0.0020$), H85Y (15.2%, $p = 0.0582$), а также Q21H, хотя частота ее встречаемости достоверно не превышала этот показатель в других позициях (11.6%, $p = 0.2425$; табл. 2, рис. 3).

В европейских странах (ЕU) наиболее распространенной была замена L90V, встречающаяся в 60.6% образцов ($p < 0.0001$; табл. 3, рис. 3).

В странах Юго-Западной и Южной Азии (AS1) в онкобелке Е6 замены L90V, H85Y и Q21H встречались в 73.9, 49.1 и 49.1% образцов соответственно ($p < 0.0001$ для всех вариантов; табл. 4, рис. 3).

В странах Юго-Восточной Азии (AS2) в белке Е6 довольно часто встречались замены D32E (46.7%), L90V (20.5%) и E120D (13.5%), а в позиции 29 белка Е7 преобладал серин (55%) ($p < 0.0001$ для всех вариантов; табл. 5, 6; рис. 3).

При анализе полученных данных замечено, что в разных регионах в одних и тех же позициях

Таблица 3. Частота встречаемости АК-замен в белке Е6 ВПЧ16-изолятов из стран Европы (выборка EU, $N = 296$)

АК-позиция	АК-замены					
	Вариант 1		Вариант 2		Вариант 3	
	АК	n (%)	АК	n (%)	АК	n (%)
R5	K	1 (0.3%)	S	1 (0.3%)	T	1 (0.3%)
T6	P	1 (0.3%)	S	1 (0.3%)	–	–
Q10	H	1 (0.3%)	E	1 (0.3%)	–	–
E14	D	1 (0.3%)	K	1 (0.3%)	–	–
R15	P	2 (0.6%)	–	–	–	–
R17	G	22 (7%)	I	11 (3.7%)	T	2 (0.6%)
Q21	D	17 (6%)	H	8 (2.4%)	–	–
D32	E	4 (1.2%)	N	1 (0.3%)	–	–
I34	L	1 (0.3%)	R	1 (0.3%)	–	–
E36	Q	1 (0.3%)	K	1 (0.3%)	G	1 (0.3%)
H85	Y	26 (8.8%)	–	–	–	–
L90	V	176 (59.5%)	–	–	–	–
R131	T	2 (0.6%)	–	–	–	–

встречаются различные аминокислоты. Так, в белке Е6 в африканской группе, а также в странах Юго-Восточной Азии в позиции 17 преобладала АК-замена R17I, в то время как в американской и европейской группах – R17G. В странах Юго-Западной и Южной

Азии в этой позиции встречался только вариант R17G (табл. 1–5).

Аналогично для позиции 21 белка Е6 в африканской и европейской группе преобладала АК-замена Q21D, в то время как в Америке (AM) и Азии (AS1 и AS2) – Q21H (табл. 1–5).

Таблица 4. Частота встречаемости АК-замен в белке Е6 ВПЧ-изолятов из стран Юго-Западной и Южной Азии (выборка AS1, $N = 230$)

АК-позиция	АК-замены					
	Вариант 1		Вариант 2		Вариант 3	
	АК	n (%)	АК	n (%)	АК	n (%)
D11	H	2 (0.87%)	–	–	–	–
R17	G	8 (3.5%)	–	–	–	–
Q21	H	113 (49.1%)	D	1 (0.4%)	–	–
D32	E	3 (1.3%)	H	1 (0.4%)	N	4 (1.7%)
E48	G	2 (0.87%)	–	–	–	–
A53	G	2 (0.87%)	–	–	–	–
F54	V	2 (0.87%)	–	–	–	–
S78	C	5 (2.2%)	–	–	–	–
R84	S	2 (0.87%)	–	–	–	–
H85	Y	113 (49.1%)	–	–	–	–
L90	V	170 (73.9%)	–	–	–	–
P102	L	2 (0.87%)	A	1 (0.4%)	–	–
D105	N	2 (0.87%)	–	–	–	–
L106	V	2 (0.87%)	–	–	–	–
Q114	K	2 (0.87%)	–	–	–	–
R131	S	1 (0.4%)	T	1 (0.4%)	–	–
E155	K	2 (0.87%)	–	–	–	–

Таблица 5. Частота встречаемости АК-замен в белке E6 ВПЧ16-изолятов в странах Юго-Восточной Азии (выборка AS2, N = 750)

АК-позиция	АК-замены							
	Вариант 1		Вариант 2		Вариант 3		Вариант 4	
	АК	n (%)	АК	n (%)	АК	n (%)	АК	n (%)
M1	K	2 (0.26%)	–	–	–	–	–	–
R5	G	22 (2.93%)	–	–	–	–	–	–
M8	R	2 (0.26%)	K	3 (0.4%)	L	1 (0.13%)	–	–
Q10	E	2 (0.26%)	–	–	–	–	–	–
R17	G	5 (0.67%)	I	9 (1.2%)	T	1 (0.13%)	–	–
K18	N	1 (0.13%)	Q	1 (0.13%)	–	–	–	–
L19	F	1 (0.13%)	V	1 (0.13%)	–	–	–	–
Q21	H	22 (2.93%)	D	9 (1.2%)	–	–	–	–
T29	S	3 (0.4%)	–	–	–	–	–	–
D32	E	350 (46.67%)	H	1 (0.13%)	N	8 (1.07%)	Y	1 (0.13%)
I34	R	5 (0.67%)	L	2 (0.26%)	T	2 (0.26%)	M	2 (0.26%)
L35	V	9 (1.2%)	–	–	–	–	–	–
E36	Q	6 (0.8%)	–	–	–	–	–	–
R46	Q	2 (0.26%)	–	–	–	–	–	–
D51	E	2 (0.26%)	–	–	–	–	–	–
R62	K	4 (0.53%)	I	2 (0.26%)	–	–	–	–
N65	S	28 (3.7%)	–	–	–	–	–	–
P66	A	2 (0.26%)	S	1 (0.13%)	–	–	–	–
D71	N	1 (0.13%)	E	6 (0.8%)	–	–	–	–
F76	L	3 (0.4%)	–	–	–	–	–	–
H85	Y	41 (5.5%)	–	–	–	–	–	–
L90	V	154 (20.5%)	–	–	–	–	–	–
D105	H	2 (0.26%)	–	–	–	–	–	–
N112	Y	2 (0.26%)	–	–	–	–	–	–
C118	G	2 (0.26%)	–	–	–	–	–	–
E120	D	101 (13.5%)	–	–	–	–	–	–
R151	T	2 (0.26%)	–	–	–	–	–	–

Следует заметить, что при сравнении частот встречаемости наиболее значимых замен по регионам замена N29S в белке E7, широко представленная в AS2-регионе (55%), обнаружена также в американской и европейской группах (4.1 и 4.5% образцов соответственно; табл. 6), в то время как в регионе AS1, расположенном географически намного ближе к AS2, эта замена не встречалась вообще. Это позволяет предположить, что некоторые АК-замены в региональных образцах ВПЧ16 могут быть связаны с глобальными процессами миграции населения.

Частота встречаемости АК-замен значительно различалась внутри регионов (табл. 1–6). Региональное распределение замен проиллюстрировано на рис. 4, анализ частот распространения замен в регионах – тепловыми картами на рис. 3 (см. также табл. S1,

рис. S1, рис. S2, Дополнительные материалы на сайте http://www.molecbio.ru/downloads/2024/4/supp_Zelenova_rus.pdf, а также доступны по DOI: 10.31857/S0026898424040036). В белке E6 ВПЧ16 замена R17I достоверно чаще встречалась в AF-выборке ($p < 0.0001$), Q21H – в AS1, Q21D – в AF (по сравнению с другими выборками $p < 0.0001$; рис. 3, рис. 4). Замена D32E чаще всего встречалась в группе AS2, частоты ее встречаемости в других выборках не различались. Замена H85Y была одинаково распространена в группах AF и AS1, немного реже в выборке AM и достоверно реже в EU и AS2. Замена L90V была широко распространена во всех выборках, кроме AS2. Замена E120D была характерна только для изолятов AF и AS2 (рис. 3). Анализ региональных выборок последовательностей высококонсервативного белка E7 выявил высокую частоту встречаемости

Таблица 6. Частота встречаемости АК-замен в белке E7 ВПЧ-изолятов в различных регионах мира

АК-позиция ^a	АК-замены ^b							
	Вариант 1		Вариант 2		Вариант 3		Вариант 4	
	АК	n (%)	АК	n (%)	АК	n (%)	АК	n (%)
Африканская группа (AF, N = 2)								
N29	S	1 (50%)	–	–	–	–	–	–
Американская группа (AM, N = 3263)								
D4	N	2 (0.06%)	–	–	–	–	–	–
N29	S	133 (4.1%)	P	1 (0.03%)	H	1 (0.03%)	–	–
E33	K	3 (0.09%)	–	–	–	–	–	–
A42	T	2 (0.06%)	V	1 (0.03%)	–	–	–	–
H51	R	1 (0.03%)	N	13 (0.4%)	–	–	–	–
D62	H	1 (0.03%)	N	1 (0.03%)	–	–	–	–
R77	C	2 (0.06%)	G	1 (0.03%)	S	1 (0.03%)	–	–
P92	L	2 (0.06%)	S	1 (0.03%)	–	–	–	–
Европейская группа (EU, N = 242)								
N29	S	11 (4.5%)	–	–	–	–	–	–
D75	N	2 (0.83%)	–	–	–	–	–	–
Азиатская группа 1 типа (AS1, N = 129)								
L15	S	2 (1.6%)	–	–	–	–	–	–
T20	S	2 (1.6%)	–	–	–	–	–	–
G88	R	10 (7.8%)	–	–	–	–	–	–
I93	T	2 (1.6%)	–	–	–	–	–	–
C94	S	2 (1.6%)	–	–	–	–	–	–
Азиатская группа 2 типа (AS2, N = 656)								
D14	H	2 (0.3%)	–	–	–	–	–	–
L28	F	29 (4.4%)	–	–	–	–	–	–
N29	S	361 (55%)	H	20 (3%)	D	1 (0.15%)	T	1 (0.15%)
D39	N	1 (0.15%)	H	1 (0.15%)	–	–	–	–
H51	N	2 (0.3%)	Y	1 (0.15%)	–	–	–	–
S63	A	2 (0.3%)	F	6 (0.9%)	–	–	–	–
D75	Y	2 (0.3%)	–	–	–	–	–	–
R77	C	17 (2.6%)	L	1 (0.15%)	H	1 (0.15%)	–	–
G88	Q	1 (0.15%)	E	1 (0.15%)	R	1 (0.15%)	–	–
P92	S	2 (0.3%)	–	–	–	–	–	–
Q96	R	1 (0.15%)	H	1 (0.15%)	–	–	–	–

^aВ качестве референсной аминокислотной последовательности белка E7 выбрана NCBI Reference Sequence NP_041326.1.

^bТаблица не включает АК-замены, встречающиеся один раз. Приведены варианты АК-замен, их число (n) и частота встречаемости (%). АК-замены, встречающиеся в >10% последовательностей, выделены жирным шрифтом.

замены N29S в выборке AS2, причем достоверно чаще, чем в AM и EU; в то время как в группе AS1 эта замена не обнаружена (рис. 3, рис. 4). Замена G88R в белке E7 наблюдалась только в выборках AS2 и AS1, причем в AS1 чаще, чем в AS2 ($p < 0.0001$; рис. 3, рис. 4).

Распределение образцов белка E6 ВПЧ16 внутри одного региона отличалось неравномерностью

(характерно для AM, AF, EU). Несмотря на это, сравнение спектров АК-замен в стране, лидирующей по числу образцов, и в других странах каждого из этих регионов достоверно значимых различий не выявило. Заметим, что для ряда стран, в том числе с немногочисленной выборкой, в дополнение к регионспецифичным АК-заменам находили 1–2 уникальные (табл. 7, рис. 3). Различия в частоте

распространения АК-замен обнаружены и между отдельными провинциями Китая. Так, в центральных и восточных преобладали замены, характерные для региона AS2 (N29S), с минимальной представленностью замены L90V. Для северо-западных провинций паттерн АК-замен оказался близким к европейскому: высокая частота L90V с низкой встречаемостью “азиатской” N29S (рис. 4, табл. 7). В целом, на основании полученных результатов можно сделать вывод о циркуляции в различных географических локациях регионспецифичных вариантов ВПЧ16, различающихся по паттернам АК-замен в онкобелках E6 и E7.

Ковариационный анализ нуклеотидных и аминокислотных замен

Присутствие АК-замен в определенных сочетаниях позволяет предположить наличие ковариационных связей (сцепления) между аминокислотными остатками и/или кодирующими их нуклеотидами. Для проверки этой гипотезы проведен ковариационный анализ нуклеотидных и аминокислотных последовательностей E6 и E7.

При анализе выборки из 356 нуклеотидных последовательностей, кодирующих белок E6 ВПЧ16, выявлено два паттерна сцепления четырех

нуклеотидов: 63T-204A-207G-253T с вероятностью 0.9312 и 63G-204T-207A-253C с вероятностью 0.9795 (этот паттерн входит в состав референсной последовательности). Замены нуклеотидов в позициях 204 и 207 были синонимичными. Замена нуклеотидов G63T определяла две самые встречаемые АК-замены в позиции 21: Q21H и Q21D, – а замена 253 нуклеотида – одну из распространенных АК-замен – H85Y. На основании вышеизложенного можно предположить сцепление аминокислотных остатков в позициях 21 и 85. Однако анализ ковариаций полного набора аминокислотных последовательностей белка E6 не подтвердил наличия групп сцепления по отдельным АК-заменам, в частности в положениях 21 и 85; сцепленность сохранилась только на уровне E6-кодирующей ДНК.

При проведении анализа ковариаций для полной выборки нуклеотидных последовательностей, кодирующих белок E6, помимо двух вышеописанных паттернов выявлены еще два дополнительных – со сцеплением трех нуклеотидов: 50G-61C-321A с вероятностью 0.9937 (этот паттерн входит в состав референсной последовательности) и 50T-61G-321G с вероятностью 0.8377. Важно заметить, что сцепленность зарегистрирована

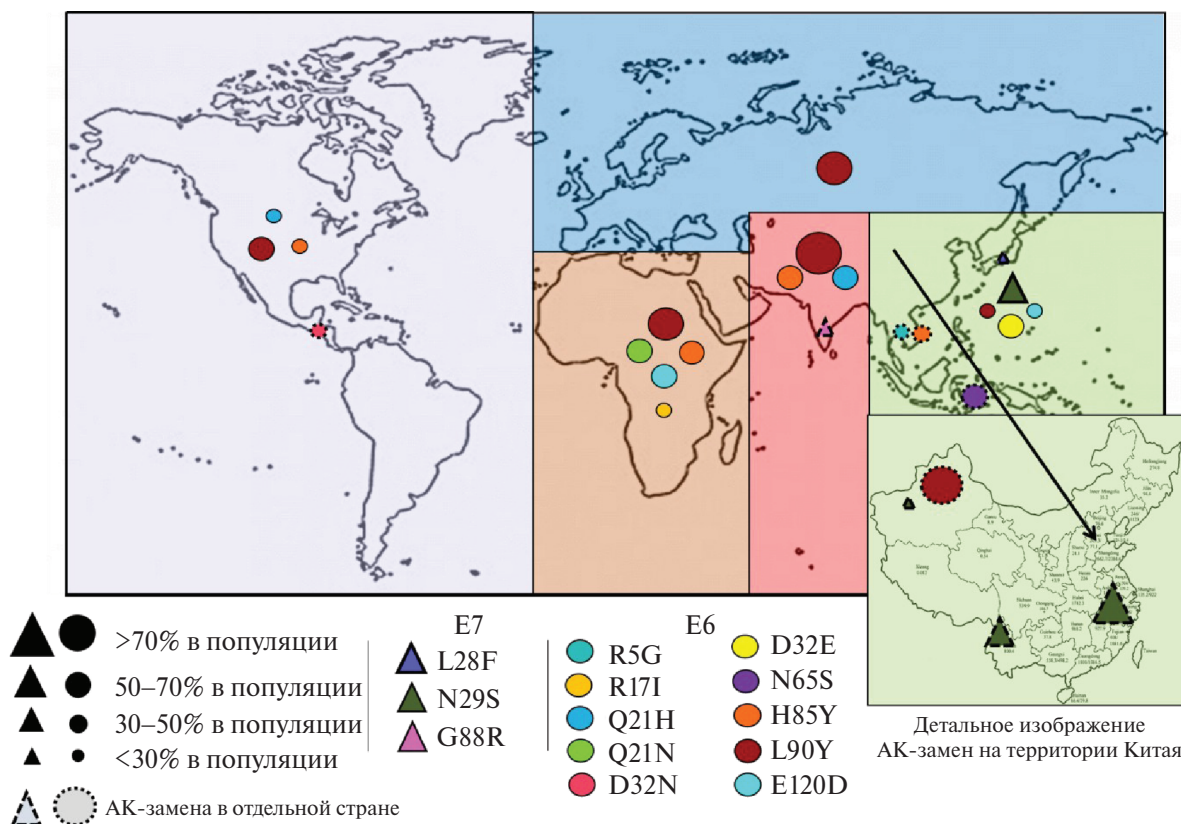


Рис. 4. Распределение АК-замен, встречающихся в >10% последовательностей выборок по пяти географическим регионам, включая АК-замены в отдельных странах. Фиолетовый блок – США, страны Центральной и Южной Америки; синий блок – страны Европы; оранжевый блок – страны Африки; красный блок – страны Юго-Западной и Южной Азии; зеленый блок – страны Юго-Восточной Азии.

Таблица 7. Распределение АК-замен в белках E6 и E7 в отдельных странах*

Страна	Варианты АК-замен в белке E6		Варианты АК-замен в белке E7	
	АК	n (%)	АК	n (%)
Регион AS1			G88R	10 (7.8%)
Индия (N = 46)			G88R	10 (21.7%) $p < 0.0001$
Регион AM	D32N	22 (1.12%)		
Гватемала (N = 78)	D32N	11 (14.1%) $p < 0.0001$		
Регион AS2			L28F	29 (4.4%)
Япония (N = 150)			L28F	26 (17.3%) $p < 0.0001$
Регион AS2	N65S	28 (3.7%)		
Индонезия (N = 40)	N65S	13 (32.5%) $p < 0.0001$		
Регион AS2	R5G H85Y	22 (2.93%) 41 (5.5%)		
Вьетнам (N = 91)	R5G H85Y	18 (19.7%) $p < 0.0001$ 10 (10.9%) $p = 0.0233$		
Регион AS2	D32E L90V	350 (46.67%) 154 (20.5%)	N29S	361 (55%)
Китай, центральные и восточные регионы (N = 154)			N29S	115 (74.7%)
Китай, юго-западный регион (N = 47)			N29S	25 (53.6%)
Китай, северный регион (E6 N = 46) (E7 N = 50)	D32E L90V	2 (4.3%) 36 (78.2%)	N29S	3 (6%)

*Указаны АК-замены, уникальные для страны в сравнении с регионом. По умолчанию все остальные АК-замены в указанной стране встречаются с той же частотой, что и в других странах данного региона (пустая строка).

на уровне исключительно нуклеотидной последовательности (для аминокислот ковариаций не выявлено).

Анализ ковариаций выполнен также для нуклеотидных и аминокислотных последовательностей E7 ВПЧ16. В результате не выявлено групп сцепления ни по нуклеотидным, ни по отдельным АК-заменам.

Анализ проводили на всех образцах – без разделения по регионам мира. Отсутствие групп сцепления может быть обусловлено различиями в сцеплениях между отдельными региональными выборками, с потерей достоверности сцеплений при объединении выборок. Для проверки этой гипотезы мы провели анализ ковариации (сцеплений) отдельно для каждой из 5 выборок (страны Европы, Африки, Америки, Юго-Западной и Южной Азии

и Юго-Восточной Азии). Анализ проводили с использованием второго биоинформатического метода (пакет Biostrings) с калькулятором MI (<http://www.yanlilab.github.io/fastcov>). Для образцов E6 американского происхождения (выборка AM) выявлена ковариация аминокислотных остатков в позициях 17, 21, 85 и 90 (рис. 5). Интересно отметить, что сцепленными оказались 4 из 6 вариабельных позиций E6 (рис. 2a, табл. 2).

Таким образом, хотя в целом по всей выборке позиции подверженных замен аминокислотных остатков в составе как E6, так и E7 не сцеплены между собой, при ограничении выборки одним географическим регионом удалось выявить сцепление аминокислотных остатков в 4 из 6 вариабельных позиций белка E6.

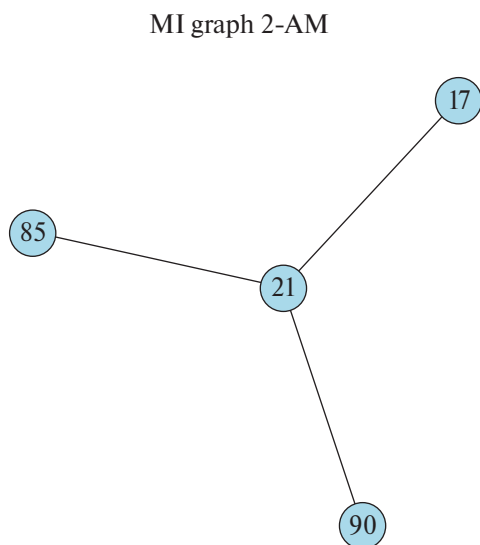


Рис. 5. Ковариационный анализ АК-замен белка Е6 для образцов американского региона (АМ). Анализ проведен с использованием второго биоинформатического метода с применением пакета Biostrings с калькулятором MI (<http://www.yanlilab.github.io/fastcov>). В окружностях указаны позиции аминокислотных остатков, линии между ними — сцепление двух АК-замен, вероятность сцепления между которыми $p < 0.05$ (MI).

Реконструкция истории возникновения и последующего распространения вариантов гена Е6, приводящих к АК-заменам в белке Е6

Для реконструкции истории возникновения и последующего распространения АК-замен ВПЧ16 был проведен филогенетический анализ 356 нуклеотидных последовательностей гена *Е6* (рис. 6). Из анализа были исключены все последовательности, идентичность которых превышала 99.8%. Аналогичное исследование предполагалось выполнить и для гена *Е7*, однако из-за высокой гомологии последовательностей временной сигнал не был зарегистрирован. Принцип кластеризации заключался в выделении отдельных клад на основании близости нуклеотидных последовательностей (традиционный принцип сродства последовательностей) с добавлением принципа географического единства после завершения построения дерева (“разрезание дерева”) для интерпретации полученных результатов.

В ходе анализа установлено, что общий предок известных современных вариантов ВПЧ16 имеет американское происхождение и к настоящему моменту разделился на 9 кластеров (рис. 6, табл. 8), различающихся филогенетически и/или по региону циркуляции. Первое крупное разделение на две принципиально разные ветви (группа образцов, отходящая от узла на основании возникновения ключевой замены нуклеотидов) произошло только в 1840–1850 гг. (HPD 95%: 1702–1948 гг.).

Первая ветвь, преимущественно американской выборки в 1880–1890 гг. (HPD 95%: 1796–1965 гг.), разделилась на 2 группы, при этом в каждой из них с интервалом примерно в 30 лет произошло независимое отделение африканских линий: в 1895 году (HPD 95%: 1852–1973 гг.) и в 1925 году (HPD 95%: 1873–1979 гг.) (рис. 6, кластеры VII и IX соответственно). Это разделение на клады (конечная группа образцов, не подлежащая дальнейшему делению, объединенная на основании максимального сходства нуклеотидной последовательности) подтвердили при оценке состава АК-замен в данных группах: для кластера IX это H85Y, в то время как изоляты кластера VII, кроме вышеуказанной, несут дополнительную – Q21D (табл. 8). Важно подчеркнуть, что изоляты этих кластеров сформировались в результате независимой эволюции ВПЧ16 разных линий. Интересны первичные структуры анализируемого фрагмента белка Е6 из кластера VIII, в котором большинство образцов содержит более одной АК-замены, при этом наиболее часто встречается сочетание трех замен: Q21H-H85Y-L90V (табл. 8; рис. 6, кластер VIII). Описывая эволюцию образцов ВПЧ16 первой ветви, стоит также заметить, что в дальнейшем из разных линий исходной американской группы со схожим набором АК-замен (Q21H, H85Y, L90V) примерно в 1950 году произошло одновременное выделение двух групп AS1 в кластере V и кластере VIII (HPD 95%: 1917–1984 гг. и 1930–1983 гг.) (рис. 6).

Для изолятов второй ветви, образованной от общего предка в 1840–1850 гг., характерна более четкая кластеризация по географическому признаку (рис. 6). Обращает внимание раннее выделение небольшой группы китайского происхождения из 7 образцов в 1905–1910 гг. (HPD 95%: 1822–1972 гг.) (рис. 6, кластер VI), что может быть обусловлено завозом штаммов ВПЧ16 из Китая и дальнейшим распространением в американской популяции. Эта гипотеза подтверждена при анализе аминокислотного состава этой группы изолятов, а именно сохранением замены H85Y; в то время как в большинстве образцов, происходящих от второй ветви, произошла реверсивная замена. В отличие от H85Y, замена L90V, возникшая еще у гипотетического предка, сохранилась как в американском, так и в азиатском кластере, что позволяет считать ее не связанной с этносом.

Также в 1920 году (HPD 95%: 1860–1975 гг.) произошло знаковое разделение изолятов второй ветви, что сопровождалось изменением аминокислотного состава – появлением L90V вне зависимости от географического распределения. Позднее, в 1930 году (HPD 95%: 1893–1978 гг.), произошло выделение группы AS2 (рис. 6). Примерно в это же время, в 1930 г. (HPD 95%: 1943–1964 гг.), среди обширной группы изолятов ВПЧ16 с белком Е6 без замены L90V четко выделилась группа AS2

Регионы:

- AF
- AM
- AS1
- AS2
- EU

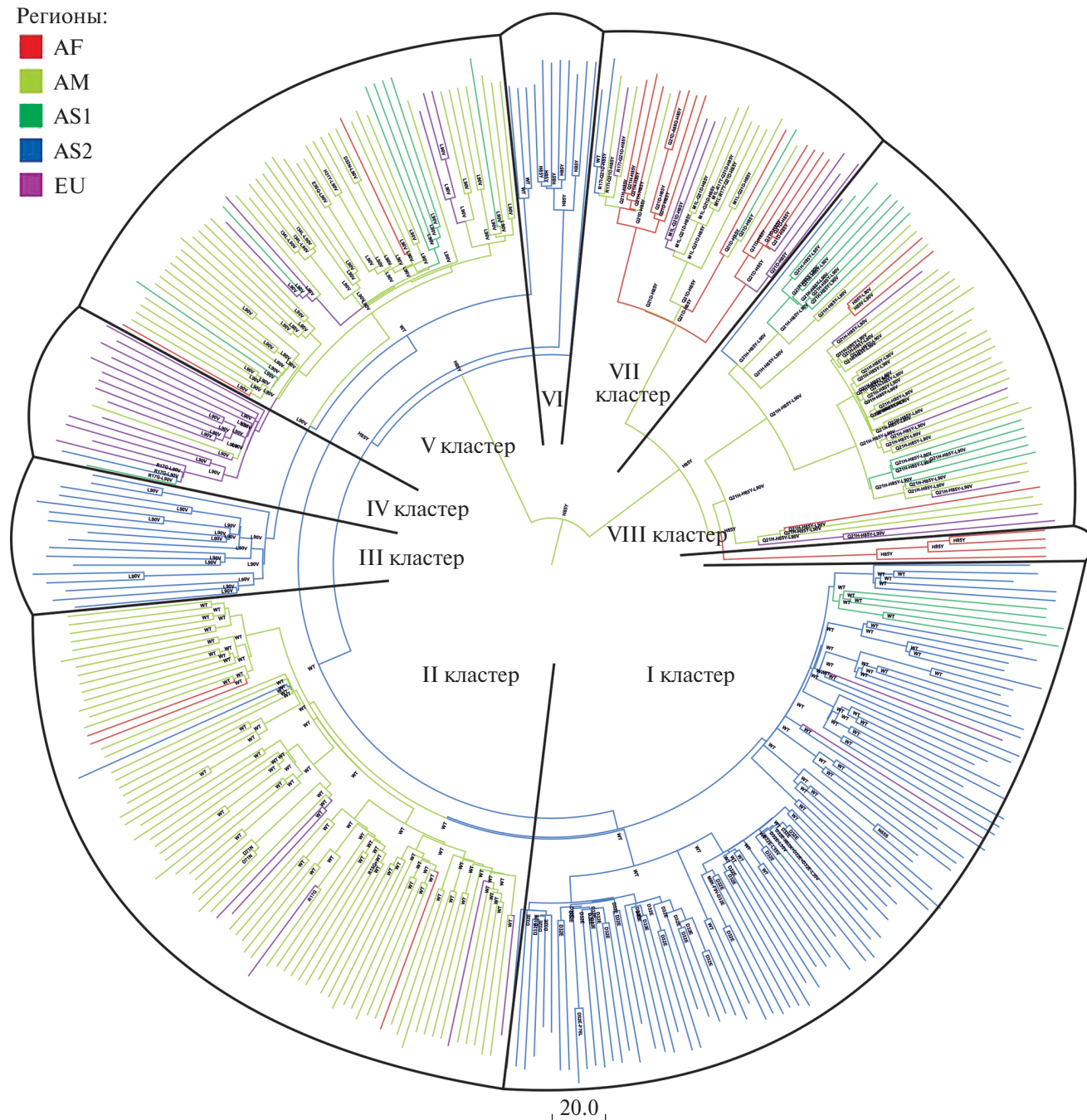


Рис. 6. Филогенетическое дерево с использованием байесовского подхода с временной шкалой для последовательностей гена *E6*, длина фрагмента – 477 п.н. (координаты по эталонному образцу NC_001526.4:7125–7601). Для каждой последовательности указан номер в базе данных GenBank, страна и год выделения. В узлах дерева приведены аминокислотные остатки, предсказанные для гипотетического предка каждой филогруппы. Обозначения: EU – страны Европы, AF – страны Африки, AM – страны Америки, AS1 – страны Юго-Западной и Южной Азии, AS2 – страны Юго-Восточной Азии.

Регионы:

- AF
■ AM
■ AS1
■ AS2
■ EU

Кластер I

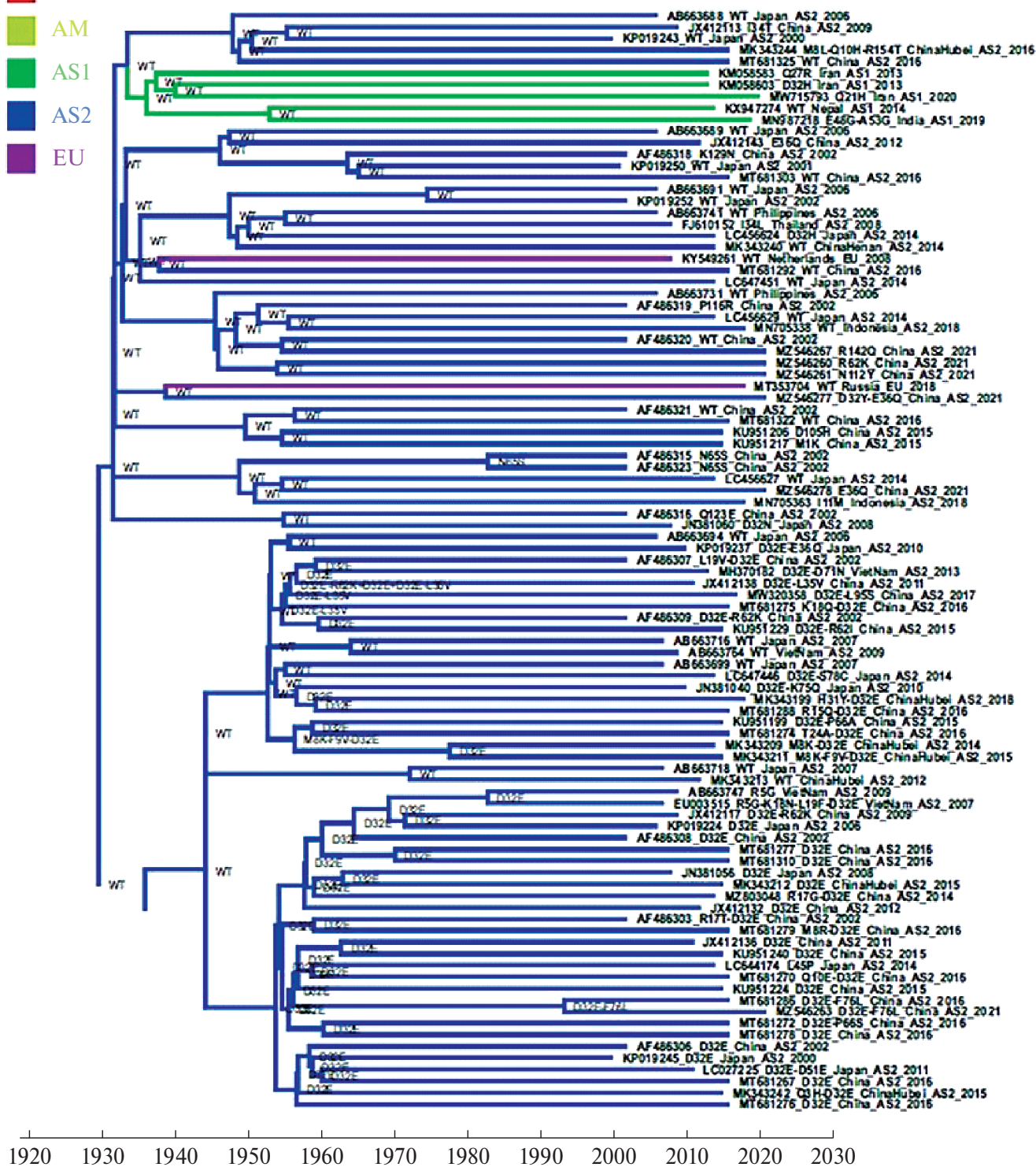


Рис. 6. Продолжение.



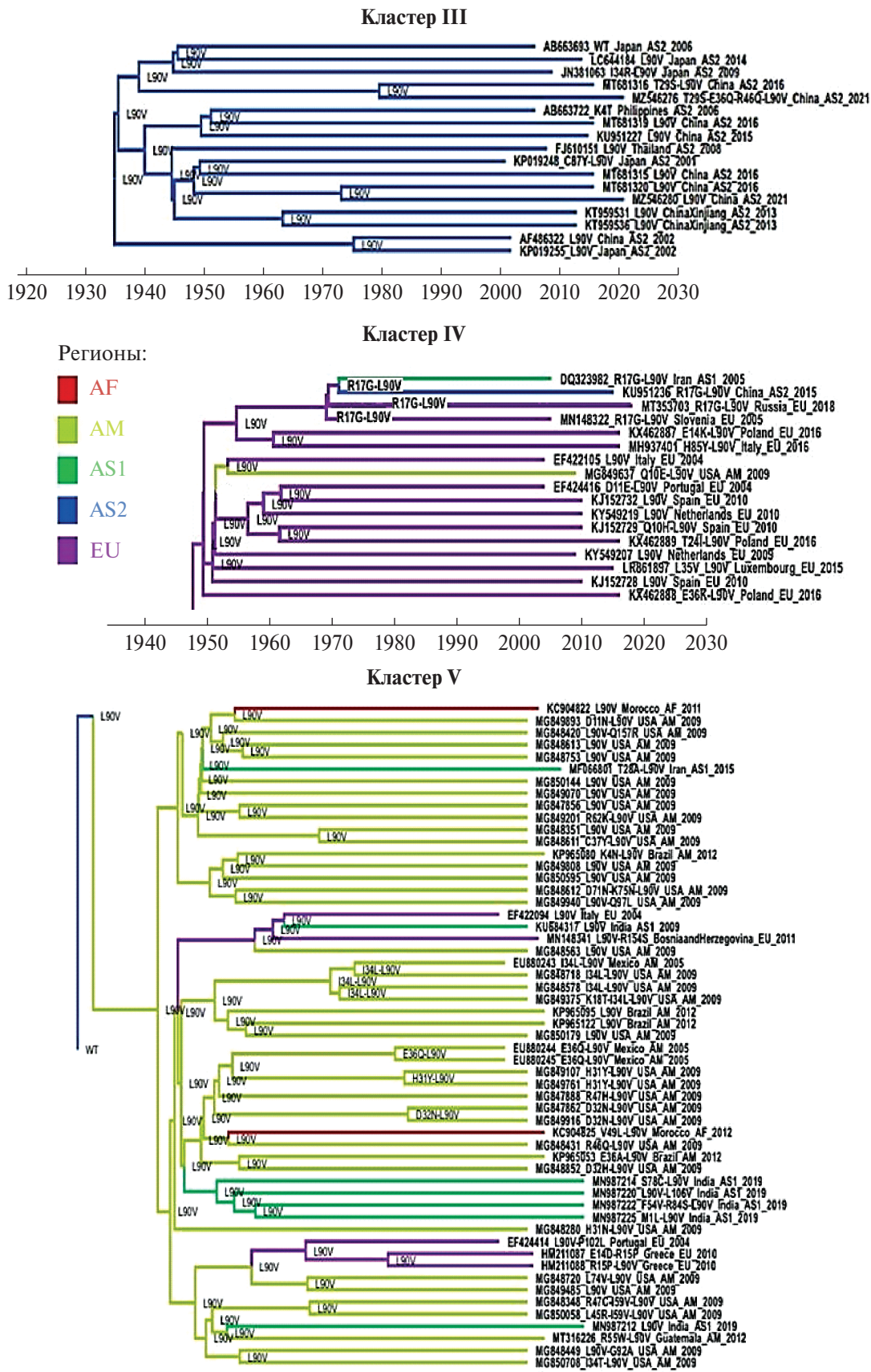


Рис. 6. Продолжение.

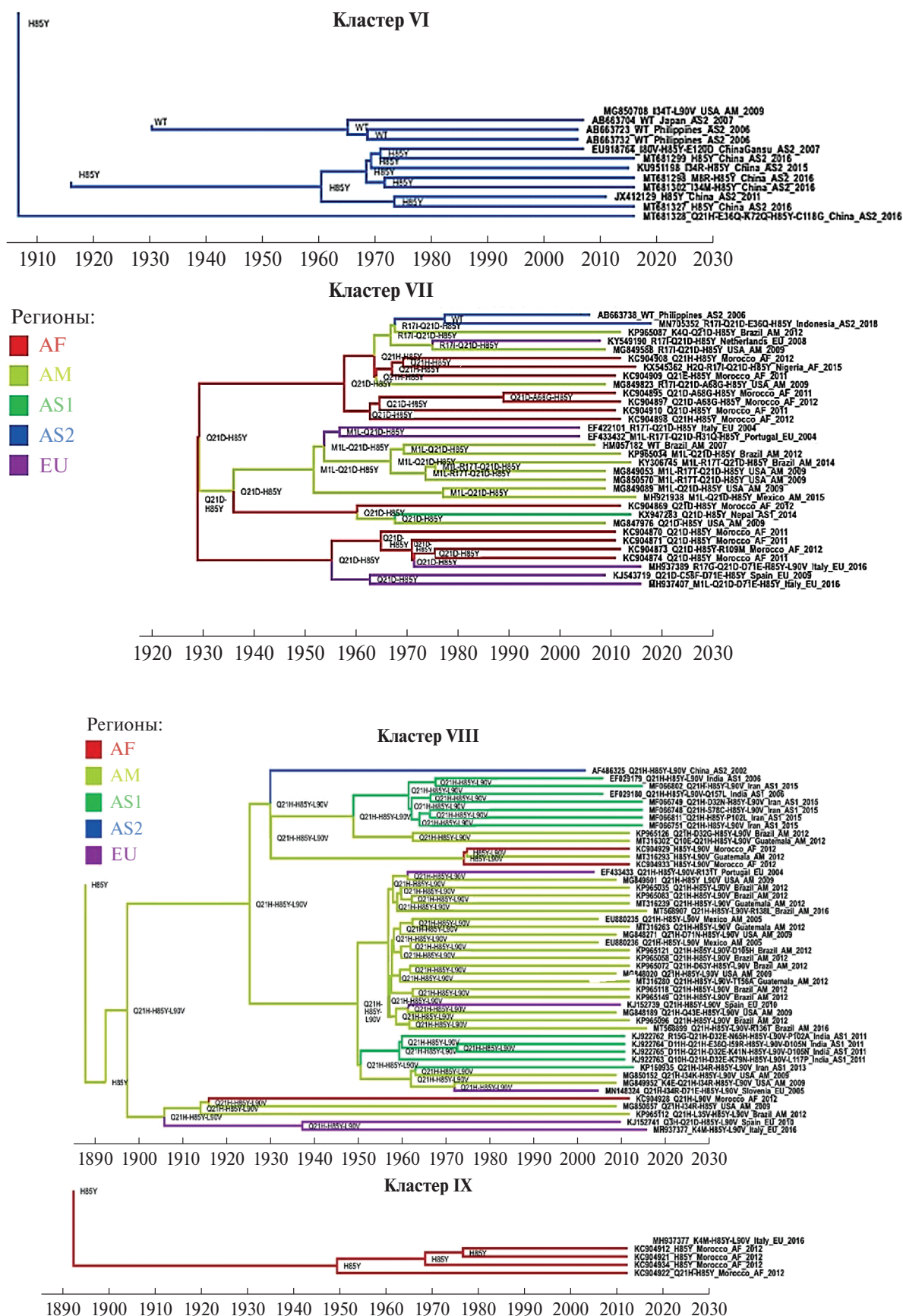


Рис. 6. Окончание.

с группоспецифической заменой D32E и американская ветвь (рис. 6, кластеры I и II).

Параллельно с этим, среди изолятов, происходящих от общего предка с АК-заменой L90V, в 1930–1940 гг. (HPD 95%: 1894–1979 гг.) произошло четкое разделение на образцы американской группы и азиатской типа 2 (рис. 6, кластеры III и V). Как и среди изолятов первой ветви, из образцов американского происхождения второй ветви в 1950 году отделилась группа AS2 (рис. 6, кластер V). Примерно в это же время отделилась и европейская группа (HPD 95%: 1967–1969 гг.; рис. 6, кластер IV).

Описываемая нами эволюция ВПЧ16 на основе анализа фрагмента его генома, кодирующего белок Е6, включает относительно небольшой временной интервал – продолжительностью всего 150 лет. По результатам исследования, АК-замена H85Y филогенетически оказалась самой древней, а скорость накопления АК-замен в исследованном фрагменте генома ВПЧ16 составляла 10^{-5} замен на сайт в год. Основные АК-замены в белке Е6 штаммов ВПЧ16, циркулирующих в человеческой популяции на современном этапе, возникли в течение последних 50 лет; причем самая “молодая” – F76L, обнаруженная в китайских изолятах конца XX века (1990 год, HPD 95%: 1966–2016 гг.).

Аминокислотные замены в белках Е6 и Е7 ВПЧ16 и противовирусный иммунный ответ

Различия в вариантах Е6 и Е7 ВПЧ16, циркулирующих в различных географических регионах (рис. 2), а также реверсии АК-замен при переходе ВПЧ16 в другой географический регион (рис. 6) могут быть связаны с этническими особенностями населения, в частности с различиями в экспрессии молекул главного комплекса гистосовместимости (HLA) и их аллелей. Последние ассоциированы с различной специфичностью В- и Т-клеточного противовирусного иммунного ответа. Это приводит к селекции различных вариантов вируса в той или иной человеческой популяции, которые успешно избегают иммунного ответа хозяина.

Для проверки этой гипотезы выполнено картирование эпитопов Е6 и Е7, взятых из международной базы данных открытого доступа. Для онкобелка Е6 ВПЧ16 описано 126 эпитопов, для Е7 – 142. В ходе анализа установлено, что практически все значимые АК-замены, описанные выше, входят в состав более 10 эпитопов (рис. 7, 8).

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Известно, что распространенность генотипов ВПЧ в разных регионах мира варьирует. В настоящее время условно выделяют “европейские” типы ВПЧ (16, 18, 31, 33, 45 и др.) и “азиатские” (52, 58, 59, 65 и др.) [11]. В результате секвенирования геномов ВПЧ выявлены популяционные различия

даже внутри одного генотипа вируса. В литературе описано 4 генотипические группы ВПЧ16: А, В, С и D [12], а также подгруппы А1–4, В1–2, С и D1–3, различающиеся по последовательности капсидного белка L1. Интересно, что эти подгруппы имеют четкую региональную представленность [13]. Вышеописанную классификацию активно используют в фундаментальных и прикладных научных исследованиях, так как именно на белках капсида основаны разработки профилактических ВПЧ-вакцин [14].

В проведенном нами исследовании акцент сделан на онкобелках Е6 и Е7 ВПЧ16. По сравнению с капсидными, белки Е6 и Е7 отличаются большей консервативностью. Именно эти вирусные белки реализуют канцерогенный потенциал вируса [7], который зависит, в том числе, от их первичной структуры и со временем появляющихся мутаций [8–10]. Именно с этими особенностями белков Е6 и Е7 связан тот факт, что секвенирование и филогенетический анализ этих фрагментов генома ВПЧ сфокусированы в основном на выявлении корреляций между полиморфизмами в генах Е6 и Е7, частотой случаев хронической персистенции вируса и последующей малигнизацией. И хотя полногеномные исследования ВПЧ проводят во многих странах, до сих пор не было сравнительного анализа выявленных вариаций в первичной структуре вирусных онкобелков по разным географическим регионам.

Нами проведен анализ результатов многочисленных исследований и идентифицированы АК-замены в белках ВПЧ16, характерные для различных регионов мира. Важно отметить, что ген Е6 имеет два старт-кодона (AUG) и две открытые рамки считывания: 1) ORF-E6(1) между нуклеотидами 83 и 559, которая кодирует белок длиной 158 а.о.; и 2) ORF-E6(2) между нуклеотидами 104 и 559, которая кодирует белок, состоящий из 151 а.о. По сравнению с AUG в позиции 104–106 старт-кодон в позиции 83 слабее, так как в его последовательности Козак отсутствует пурин в положении –3 AUG и G в позиции +4 AUG, что ослабляет связывание мРНК Е6 рибосомой. В силу этого старт-кодон AUG в позиции 83 реже используется для инициации трансляции, поэтому преимущественно синтезируется белок длиной 151 а.о. [15]. Тем не менее обе формы белка Е6 активно изучают, в связи с чем нумерация остатков может различаться. Например, замена R17G может быть представлена как R20G, L90V – как L83V и т.д., что осложняет анализ последовательностей. В проведенном исследовании в качестве референсной мы использовали нуклеотидную последовательность гена Е6 ВПЧ16 (NC_001526, GeneID: 1489078) и белковый продукт, кодируемый ORF-E6(1) (83–559 н.), то есть длиной 158 а.о.

Таблица 8. Кластеризация изолятов Е6*

Позиция	Кластер I N = 93		Кластер II N = 77		Кластер III N = 17		Кластер IV N = 17		Кластер V N = 53		Кластер VI N = 13		Кластер VII N = 28		Кластер VIII N = 47		Кластер IX N = 4	
	AK	n (%)	AK	n (%)	AK	n (%)	AK	n (%)	AK	n (%)	AK	n (%)	AK	n (%)	AK	n (%)	AK	n (%)
M1	-		-		-		-		-		-		L	7 (25%)	-		-	
R17	T G	1 (1.1%) 1 (1.1%)	G	2 (2.6%)	-		G	4 (23.5%)	-		I	1 (7.7%)	T I G	5 (17.9%) 3 (10.7%) 1 (3.6%)	-		-	
Q21	H	1 (1.1%)	E	1 (1.3%)	-		-		-		D H	1 (7.7%) 1 (7.7%)	D H E	24 (85.7%) 4 (14.3%) 1 (3.6%)	H D	40 (85.1%) 1 (2.1%)	H	1 (25%)
D32	E N H	42 (45.2%) 1 (1.1%) 2 (2.2%)	A	1 (1.3%)	-		-		N H	2 (3.8%) 1 (1.9%)	-		-		E N G	3 (6.4%) 1 (2.1%) 1 (2.1%)	-	
H85	-		-		-		Y	1 (5.9%)	-		Y	9 (69%)	Y	28 (100%)	Y	44 (93.6%)	Y	4 (100%)
L90	-		V	2 (2.6%)	V	15 (88.2%)	V	17 (100%)	V	52 (98%)	-		V	3 (10.7%)	V	43 (91.5%)	-	
E120	-		-		-		-		-		D	1 (7.7%)	-		-		-	

*Указано число образцов (N) в каждом кластере, варианты замен (AK), их число (n) и частота встречаемости (%). АК-замены, встречающиеся в >10% последовательностей выделены жирным шрифтом.

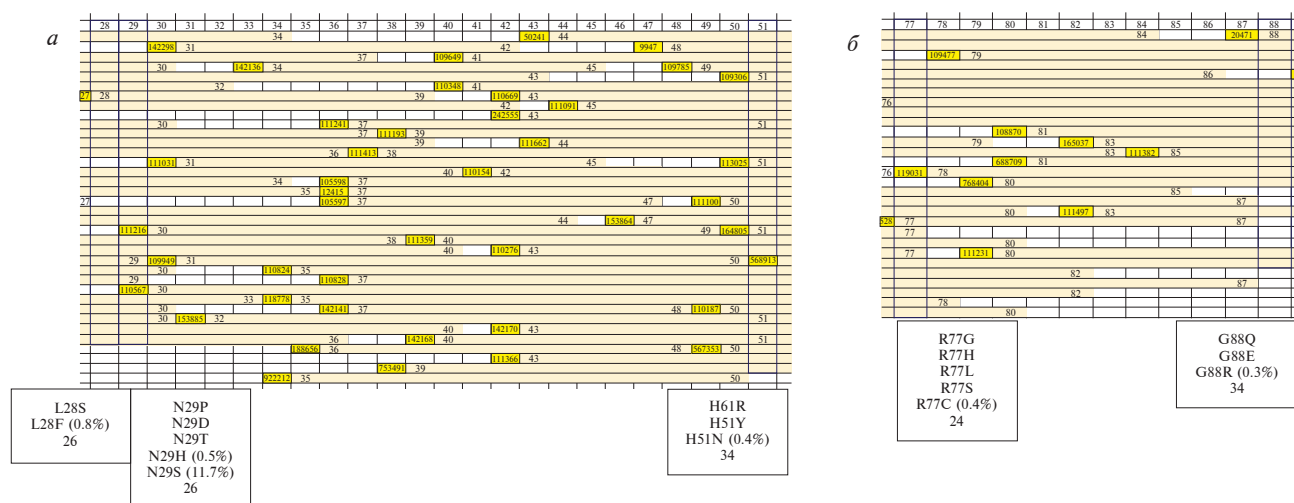


Рис. 8. Эпитопная карта онкобелка E7. Представлены фрагменты 28–51 (а) и 52–99 (б) аминокислотной последовательности. Выборка включала последовательности, идентичность которых превышала 99.8%. Желтым цветом обозначены идентификационные номера (ID) эпитопов в базе данных IEDB, светло-желтым — последовательность картированного эпитопа, цифры внутри блока — позиции первых и последних аминокислотных остатков. В рамке внизу каждого блока указаны частоты встречаемости АК-замен в варибельной позиции белка E7 в составе эпитопного кластера. Цифры внизу — суммарное число эпитопов, затрагиваемых АК-заменами в указанных позициях.

образцах ВПЧ16, собранных в греческой популяции, выявили высокую частоту замены L90V в белке E6: 78.2% при высокодифференцированных опухолях шейки матки и несколько ниже, 64.2%, при низкодифференцированных. Корректных полногеномных данных секвенирования генов E6 и E7 из стран Западной Европы на сегодняшний день крайне мало, в связи с чем сравнить АК-замены в странах Восточной и Западной Европы на настоящий момент не представляется возможным.

В исследовании J. Zhao и др. [16], проведенном на образцах, собранных на территории Шанхая (Китай), описаны АК-замены H85Y и E120D в белке E6 и N29S в белке E7 и их связь с канцерогенезом (частота выявления данных замен была выше в группе пациенток с дисплазией шейки матки). Результаты проведенного нами анализа подтверждают эти данные: замена N29S в белке E7 широко распространена на территории Китая (57.2%, 210/367 образцов). Что касается замены E120D в белке E6, нами показано, что она встречается в Юго-Восточной Азии с частотой 13.5% (101/750 образцов), в то время как на территории Китая — всего в 7.9% образцов; при этом замена H85Y вообще не характерна для региона Юго-Восточной Азии — ее наибольшая частота приходится на вьетнамскую популяцию (10.9%) (рис. 4, табл. 7).

В то же время в Синьцзян-Уйгурском автономном районе Китая преобладает европейский вариант ВПЧ16 — с характерной заменой L90V в белке E6 и минимальной представленностью “азиатской” замены N29S в белке E7. Согласно нашим данным, частота “европейской” АК-замены L90V в этом

регионе составляет 78.2% (36/46 образцов), а “азиатская” (тип 2) D32E встречается редко (4.3%, 2/46 образцов) (рис. 4, табл. 7). Эти результаты согласуются с данными Н. Не и др. [21], полученными при анализе образцов ВПЧ16 из провинции Синьцзян. Преобладание определенных вариантов онкобелков E6 и E7 характерно не только для отдельных провинций Китая, но и для ряда других стран и/или их отдельных областей, в частности в Японии, в странах Африки (рис. 4, табл. 7). Интересно, что эти страны и их отдельные области некоторые исследователи считают моноэтническими [22, 23].

V. Pimenoff с соавт. [13] исследовали эволюции линий (субгенотипов) ВПЧ16, типированных на основании анализа полногеномных последовательностей вируса с последующим типированием на основе гена капсидного белка L1. Авторы составили карту географического распределения субгенотипов с выделением регионов с доминантным представлением субтипов A1–3 (Европа), A4 (Юго-Восточная Азия), В и С (Африка), D (Центральная и Южная Америка). Интересно, что эти регионы в целом совпадают с выявленной нами географической распространенностью вариантов белков E6 и E7. В дальнейшем мы планируем идентифицировать ВПЧ16-субгенотипспецифичные паттерны АК-замен в белках E6 и E7.

Анализ встречаемости АК-замен в онкобелках E6 и E7 представляется крайне важным ввиду предположения о связи определенных а.о. с канцерогенезом. Именно поэтому ряд исследований был нацелен на сравнение встречаемости АК-замен при РШМ и хронической персистенции ВПЧ

с частотой аналогичных замен в случаях спонтанной элиминации вируса. Таким образом, в настоящее время идет поиск корреляций между наборами АК-замен и течением, а также исходом хронической ВПЧ-инфекции. P. van der Weele с соавт. [24] не выявили в образцах ВПЧ16 из Нидерландов четкой корреляции между полиморфизмами генов *E6/E7* и персистенцией вируса и сделали вывод о том, что особенности генома ВПЧ не играют существенной роли в процессе элиминации/хронизации инфекции, а решающий фактор — особенности организма хозяина. Напротив, Noé Escobar-Escamilla и др. [9] продемонстрировали связь замены L90V в белке E6 ВПЧ16 с длительной персистенцией вируса. В Китае с высоким риском малигнизации и развитием РШМ ассоциированы полиморфизмы D32E и H85Y в белке E6 ВПЧ16 [8, 25].

Описанные АК-замены мало затрагивают ключевые структуры белков E6 и E7: два “цинковых пальца”, образованных мотивами CXXC (63–66, 106–109, 139–142 а.о.) и PBM (PDZ-binding motif; 148–151 а.о.), в белке E6 и ключевые участки в доменах CD2 (22–26, 31–32 а.о.) и CD3 (58–61, 91–94 а.о.) белка E7 [26]. Эти последовательности необходимы для корректного функционирования онкобелков и, как правило, высококонсервативны. Исключение составил участок 30–33 а.о. E6, несущий один из четырех мотивов CXXC, затронутый полиморфизмом D32E. Интересно, что в клетках, зараженных *in vitro* ВПЧ16 с заменой D32E в белке E6, экспрессия p53 была снижена по сравнению с клетками, инфицированными вирусом с E6 дикого типа [8]. T. Mesplède T и др. [10] показали, что замены R17G, L90V и их сочетание R17G/L90V в онкобелке E6 ВПЧ16 ассоциированы со значимым снижением уровня p53 по сравнению с его экспрессией в клетках, инфицированных ВПЧ16 с E6 дикого типа. Таким образом, хотя большинство полиморфизмов гена *E6* не приводит к изменениям в структурно-функциональных доменах кодируемого белка, некоторые несинонимичные замены все-таки влияют на функционирование E6, как правило, усиливая его онкогенный потенциал. Что касается онкобелка E7, то его неизменная первичная структура, по-видимому, важна для сохранения онкогенных свойств [27].

В этой связи имеет смысл отдельно рассмотреть коэволюционные изменения: когда возникновение АК-замены в определенной позиции стимулирует появление вторичных адаптивных замен в других положениях, что необходимо для сохранения функциональности белка. Коэволюцию выявляют методом ковариационного анализа. Такой анализ, проведенный нами для общей совокупности аминокислотных последовательностей, не позволил выявить наличие ковариационных связей ни в белке E6, ни в E7. Ковариационный паттерн,

охватывающий переменные позиции 17, 21, 85 и 90 для белка E6, обнаружен только для последовательностей американского происхождения. Возможно, наличие разных паттернов сцепления для генетически отдаленных наборов последовательностей маскирует существующие ковариационные связи. Учитывая это, в дальнейшем мы планируем исследовать наличие групп сцепления аминокислотных остатков на генетически однородных выборках образцов с использованием различных алгоритмов биоинформатической обработки. Успешность такого подхода была продемонстрирована нами ранее при проведении ковариационного анализа АК-замен в белке NS5A вируса гепатита С [28].

Влияние первичной структуры белка на его функции и их изменения в ходе эволюции вполне согласуются с появлением естественных мутаций в белках длительно циркулирующего в человеческой популяции ВПЧ, хотя трудно понять их филогеографию: географические различия распространности отдельных АК-замен и их сочетаний. Возможно, причина кроется в этнически обусловленных различиях хозяев, например при процессинге вирусных белков и ответе иммунной системы. Уже были сообщения о полиморфизмах системы процессинга антигенов и рецепторов иммунных клеток организма-хозяина (TLR, FAS, Tap1/2 и др.), об этнических различиях в распространности аллелей и гаплотипов HLA [29–31]. Совокупность этих факторов обуславливает этнические и, как следствие, географические различия в распознавании В- и Т-клеточных эпитопов. Однонуклеотидные замены в геноме ВПЧ, в том числе несинонимичные, происходят случайным образом. Однако, если АК-замена затрагивает структуру эпитопа, то это может нарушить распознавание соответствующего антигена иммунными клетками, что приводит к ускользанию вируса-мутанта от иммунного ответа хозяина. Конечно, такой вирус получает преимущества над родительским и продолжает циркуляцию в этой популяции, вытесняя другие штаммы. Так, по результатам биоинформатического анализа J. He и др. [25] выявили, что замены R17G, D32N, D32E, H85Y и L90V в белке E6 затрагивают Т-клеточные эпитопы, а это означает появление новых эпитопов или изменение аффинности связывания молекулами HLA [25]. К высокоиммуногенной относится область E14–I34 а.о. онкобелка E6 [32]. Экспериментально показано, что замены в позиции 17 белка E6 влияют на распознавание Т-клеточного эпитопа молекулой HLA-B7 [33]. Достоверно большее число АК-замен в иммуногенных областях E14–I34 и L45–A68 у пациентов с инвазивной цервикальной карциномой по сравнению с пациентами с раком *in situ* считают отражением факта ухода вируса от иммунного ответа/эпитопного дрейфа, вызывающего потерю организмом контроля за его репликацией [32]. Такого

рода замены могут носить сочетанный характер. Так, для гемагглютинина вируса гриппа А выявлено сочетанное изменение аминокислотных остатков на поверхности белка в В-клеточных эпитопах или вблизи, что определяет ускользание вируса от иммунного ответа хозяина [34].

Учитывая различную частоту встречаемости АК-замен в географических популяциях, выявленную филогенетическим анализом их реверсию, а также тот факт, что наиболее частые АК-замены Е6 ВПЧ затрагивали структуры эпитопов, мы пришли к выводу, что распознавание вирусного антигена клетками иммунной системы хозяина – один из ключевых факторов в закреплении новой мутации или, наоборот, элиминации мутировавшего варианта вируса из данного региона.

Изменчивость вирусов и ее направленность, особенно при хронических инфекциях, – результат действия многих факторов. Случайные нуклеотидные замены в геноме вируса, вызванные ошибками в работе вирусной полимеразы, приводят к мутациям в белках, влияя на их функционирование и, в итоге, на преобладание определенных вариантов вируса [35]. Однако не только репликативная активность и эффективность сборки вирионов определяют успешное распространение определенных вариантов вирусов. Давление со стороны иммунной системы хозяина относится к важнейшим факторам, определяющим эволюцию вирусов, так как АК-замены в В- и Т-клеточных эпитопах вирусных белков позволяют мутантному варианту ускользнуть от иммунного ответа. Так, А. Кумаг и др. [36] продемонстрировали важность 16–60 и 76–94 а.о. белка Е6 ввиду их влияния на иммуногенность вируса, так как эти области содержат большинство известных эпитопов. Иммунное давление зависит от представленности в популяции организма-хозяина некоторых иммунных гаплотипов, которые и определяют различия в генетике человека, в том числе из разных этносов. Именно поэтому в результате коэволюции вирусного патогена и человека в определенных этнических группах преобладают разные варианты одного и того же вируса, несущие те или иные АК-замены в иммунных эпитопах. Ранее мы выдвигали гипотезу разнонаправленной эволюции вируса гепатита С в 7 этнически разных популяциях под действием различного иммунного давления [28, 37]. Адаптация вирусов к HLA-ассоциированному иммунному давлению описана для других вирусов, включая ВИЧ-1 [38] и вирус гепатита В [39].

В результате проведенного филогенетического анализа гена *E6* продемонстрировано четкое распределение АК-замен в кодируемом белке по популяциям, что подтверждает кластеризацию нуклеотидных последовательностей по географическому признаку. Ускользание ВПЧ16 от иммунного давления, приводящее к разнонаправленной

эволюции вируса в разных этносах (то есть направление эволюции определяется преобладающими в популяции иммунными гаплотипами), до сих пор описано не было. Этноспецифические особенности иммунного ответа на ВПЧ и его адаптивную изменчивость – тема дальнейших исследований.

В результате проведенного анализа нами идентифицирован общий предок ВПЧ16 – вирус американского происхождения. Однако остался нерешенным ряд вопросов:

- образовавшиеся варианты ВПЧ16 перешли из Америки в другие регионы и сформировали там очаги или американские штаммы приобрели те же “региональные” нуклеотидные замены, потому что попали в те же этнические группы, что в Африке и в Азии;

- связано ли появление в Америке этноспецифических вариантов ВПЧ16 (по белкам Е6 и Е7) с массовой миграцией в конце XIX–начале XX вв.

Ограниченность выборки нуклеотидных последовательностей по времени (самые ранние образцы были собраны в конце 1980 гг.) не позволяет ответить на эти вопросы, что определяет необходимость дальнейшего изучения этой темы.

К существенным ограничениям для экстраполяции полученных данных на популяции всех описанных географических регионов относится недостаточный охват секвенированием генома ВПЧ16, поэтому спектр идентифицированных АК-замен и реконструкция их географического распространения носит мозаичный характер. В дальнейших исследованиях мы планируем охватить секвенированием гены *E6* и *E7* изолятов ВПЧ16, циркулирующих на европейской и азиатской территориях Российской Федерации, а также провести полногеномный анализ этих штаммов, что позволит подтвердить или опровергнуть результаты, полученные на фрагментах генома (гены *E6* и *E7*).

Нами продемонстрировано, что циркулирующие штаммы ВПЧ16 имеют общего предка, который, попав в разные географические регионы, эволюционировал по-разному, что подтверждено различным спектром АК-замен в белке Е6 в разных популяциях. Мы считаем, что различия в паттернах мутаций обусловлены этноспецифическими особенностями иммунного ответа. Полученные результаты могут быть использованы при надзоре за распространением ВПЧ-инфекции, разработке диагностических тест-систем, а также вакцин против ВПЧ-ассоциированных онкологических заболеваний.

Авторы декларируют отсутствие внешнего финансирования для проведения исследования и публикации статьи.

Настоящая статья не содержит каких-либо исследований с участием людей или животных в качестве объектов исследований.

Авторы сообщают об отсутствии конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Тихомиров А.Л., Сарсания С.И., Филатова Г.А. (2018) Вирус папилломы человека: от понимания иммунопатогенеза к рациональной тактике ведения. *Гинекология*. **3**, 5–11. doi: 10.26442/2079-5696-2018.3.5-11
2. Волгарева Г.М. (2020) Папилломавирусный канцерогенез. Основные достижения и некоторые проблемы Часть 1. Общие представления о папилломавирусах. Формы рака, ассоциированные с вирусами папилломы человека. *Российский биотерапевтический журнал*. **19**(1), 6–12. doi: 10.17650/1726-9784-2019-19-1-6-12
3. Huibregtse J.M., Scheffner M., Howley P.M. (1991) A cellular protein mediates association of p53 with the E6 oncoprotein of human papillomavirus types 16 or 18. *EMBO J.* **10**(13), 4129–4135. doi: 10.1002/j.1460-2075.1991.tb04990.x
4. Hebner C., Beglin M., Laimins L.A. (2007) Human papillomavirus E6 proteins mediate resistance to interferon-induced growth arrest through inhibition of p53 acetylation. *J. Virol.* **81**(23), 12740–12747. doi: 10.1128/JVI.00987-07
5. Um S.J., Rhyu J.W., Kim E.J., Jeon K.C., Hwang E.S., Park J.S. (2002) Abrogation of IRF-1 response by high-risk HPV E7 protein *in vivo*. *Cancer Lett.* **179**(2), 205–212. doi: 10.1016/s0304-3835(01)00871-0
6. Курмышкина О.В., Волкова Т.О., Ковчур П.И., Бахлав И.Е., Немова Н.Н. (2011) Гены раннего ответа в патогенезе рака шейки матки: обзор. *Опухоли женской репродуктивной системы*. **1**, 96–105. doi: 10.17650/1994-4098-2011-0-1-96-105
7. Duensing A., Spardy N., Chatterjee P., Zheng L., Parry J., Cuevas R., Korzeniewski N., Duensing S. (2009) Centrosome overduplication, chromosomal instability, and human papillomavirus oncoproteins. *Environ. Mol. Mutagen.* **50**(8), 741–747. doi: 10.1002/em.20478
8. Ai W., Wu C., Jia L., Xiao X., Xu X., Ren M., Xue T., Zhou X., Wang Y., Gao C. (2022) Deep sequencing of HPV16 E6 region reveals unique mutation pattern of HPV16 and predicts cervical cancer. *Microbiol. Spectr.* **10**(4), e0140122. doi: 10.1128/spectrum.01401-22
9. Escobar-Escamilla N., González-Martínez B.E., Araiza-Rodríguez A., Fragoso-Fonseca D.E., Pedroza-Torres A., Landa-Flores M.G., Garcés-Ayala F., Mendieta-Condado E., Díaz-Quinonez J.A., Castro-Escarpulli G., Ramírez-González J.E. (2019) Mutational landscape and intra-host diversity of human papillomavirus type 16 long control region and E6 variants in cervical samples. *Arch. Virol.* **164**(12), 2953–2961. doi: 10.1007/s00705-019-04407-6
10. Mesplède T., Gagnon D., Bergeron-Labrecque F., Azar I., Sénéchal H., Coutlée F., Archambault J. (2012) p53 degradation activity, expression, and subcellular localization of E6 proteins from 29 human papillomavirus genotypes. *J. Virol.* **86**(1), 94–107. doi: 10.1128/JVI.00751-11
11. Бестаева Н.В., Назарова Н.М., Прилепская В.Н., Трофимов Д.Ю., Бурменская О.В., Суламанидзе Л.А. (2013) Папилломавирусная инфекция: новые взгляды на диагностику и лечение (обзор литературы). *Гинекология*. **3**, 4–7. <https://gynecology.orscience.ru/2079-5831/article/view/28190>
12. Burk R.D., Harari A., Chen Z. (2013) Human papillomavirus genome variants. *Virology*. **445**(1–2), 232–243. doi: 10.1016/j.virol.2013.07.018
13. Pimenoff V.N., de Oliveira C.M., Bravo I.G. (2017) Transmission between archaic and modern human ancestors during the evolution of the oncogenic human papillomavirus 16. *Mol. Biol. Evol.* **34**(1), 4–19. doi: 10.1093/molbev/msw214
14. Kirnbauer R., Hubbert N.L., Wheeler C.M., Becker T.M., Lowy D.R., Schiller J.T. (1994) A virus-like particle enzyme-linked immunosorbent assay detects serum antibodies in a majority of women infected with human papillomavirus type 16. *J. Natl. Cancer Inst.* **86**(7), 494–499. doi: 10.1093/jnci/86.7.494
15. Shally M., Alloul N., Jackman A., Muller M., Gissmann L., Sherman L. (1996) The E6 variant proteins E6I–E6IV of human papillomavirus 16: expression in cell free systems and bacteria and study of their interaction with p53. *Virus Res.* **42**(1–2), 81–96. doi: 10.1016/0168-1702(96)01301-9
16. Zhao J., Zhu J., Guo J., Zhu T., Zhong J., Liu M., Ruan Y., Liao S., Li F. (2020) Genetic variability and functional implication of HPV16 from cervical intraepithelial neoplasia in Shanghai women. *J. Med. Virol.* **92**(3), 372–381. doi: 10.1002/jmv.25618
17. Ortiz-Ortiz J., Alarcón-Romero L.D.C., Jiménez-López M.A., Garzón-Barrientos V.H., Calleja-Macias I., Barrera-Saldaña H.A., Leyva-Vázquez M.A., Illades-Aguar B. (2015) Association of human papillomavirus 16 E6 variants with cervical carcinoma and precursor lesions in women from Southern Mexico. *Virol. J.* **12**, 29. doi: 10.1186/s12985-015-0242-3
18. Ramas V., Mirazo S., Bonilla S., Ruchansky D., Arbiza J. (2018) Analysis of human papillomavirus 16 E6, E7 genes and long control region in cervical samples from Uruguayan women. *Gene*. **654**, 103–109. doi: 10.1016/j.gene.2018.02.023
19. Szostek S., Zawilinska B., Klimek M., Kosz-Vnenchak M. (2017) HPV16 E6 polymorphism and physical state of viral genome in relation to the risk of cervical cancer in women from the south of Poland. *Acta Biochim. Pol.* **64**(1), 143–149. doi: 10.18388/abp.2016_1364
20. Tsakogiannis D., Papadopoulou A., Kontostathi G., Ruether I.G.A., Kyriakopoulou Z., Dimitriou T.G., Orfanoudakis G., Markoulatos P. (2013) Molecular and evolutionary analysis of HPV16 E6 and E7 genes in Greek women. *J. Med. Microbiol.* **62**, 1688–1696. doi: 10.1099/jmm.0.055491-0
21. He H., Li H., Fan P., Zhu J., Pan Z., Pan H., Wu D., Ren X., Guo X., Li D., Pan Z., Shao R. (2016) Variants of human papillomaviruses 16 (HPV16) in Uigur women in Xinjiang, China. *Infect. Agent. Cancer*. **11**, 44. doi: 10.1186/s13027-016-0089-2
22. Pan Z., Xu S. (2020) Population genomics of East Asian ethnic groups. *Hereditas*. **157**(1), 49. doi: 10.1186/s41065-020-00162-w

23. Hudson M.J., Nakagome S., Whitman J.B. (2020) The evolving Japanese: the dual structure hypothesis at 30. *Evol. Hum. Sci.* **2**, e6.
doi: 10.1017/ehs.2020.6
24. van der Weele P., Meijer C.J.L.M., King A.J. (2017) Whole-genome sequencing and variant analysis of human papillomavirus 16 infections. *J. Virol.* **91**(19), e00844-17.
doi: 10.1128/JVI.00844-17
25. He J., Li Q., Ma S., Li T., Chen Y., Liu Y., Cui Y., Peng J., Shi Y., Wei X., Ding X. (2022) The polymorphism analysis and epitope predicted of Alphapapillomavirus 9 E6 in Sichuan, China. *Virol. J.* **19**, 14.
<https://doi.org/10.1186/s12985-021-01728-4>
26. Tomaic V. (2016) Functional roles of E6 and E7 oncoproteins in HPV-induced malignancies at diverse anatomical sites. *Cancers* (Basel). **8**, 95.
doi.org/10.3390/cancers8100095
27. Brauburger K, Hume AJ, Mühlberger E, Olejnik J. (2012) Forty-five years of Marburg virus research. *Viruses*. **4**(10), 1878–927.
doi: 10.3390/v4101878
28. Kyuregyan K.K., Kichatova V.S., Karlsen A.A., Isaeva O.V., Solonin S.A., Petkov S., Nielsen M., Isagulians M.G., Mikhailov M.I. (2020) Factors influencing the prevalence of resistance-associated substitutions in NS5A protein in treatment-naïve patients with chronic hepatitis C. *Biomedicines*. **8**(4), 80.
doi: 10.3390/biomedicines8040080
29. Mehta A.M., Mooij M., Branković I., Ouburg S., Morré S.A., Jordanova E.S. (2017) Cervical carcinogenesis and immune response gene polymorphisms: a review. *J. Immun. Res.* **2017**, 8913860.
doi: 10.1155/2017/8913860
30. Pandey N.O., Chauhan A.V., Raithatha N.S., Patel P.K., Khandelwal R., Desai A.N., Choksi Y., Kapadia R.S., Jain N.D. (2019) Association of *TLR4* and *TLR9* polymorphisms and haplotypes with cervical cancer susceptibility. *Sci. Rep.* **9**(1), 9729.
doi: 10.1038/s41598-019-46077-z
31. Tan S.C., Ismail M.P., Duski D.R., Othman N.H., Ankathil R. (2017) FAS c.-671A>G polymorphism and cervical cancer risk: a case-control study and meta-analysis. *Cancer Genet.* **211**, 18–25.
doi: 10.1016/j.cancergen.2017.01.004
32. Rader J.S., Tsaih S.W., Fullin D., Murray M.W., Iden M., Zimmermann M.T., Flister M.J. (2019) Genetic variations in human papillomavirus and cervical cancer outcomes. *Int. J. Cancer.* **144**(9), 2206–2214.
doi: 10.1002/ijc.32038
33. Ellis J.R., Keating P.J., Baird J., Hounsell E.F., Renouf D.V., Rowe M., Hopkins D., Duggan-Keen M.F., Bartholomew J.S., Young L.S., Stern P.L. (1995) The association of an HPV16 oncogene variant with HLA-B7 has implications for vaccine design in cervical cancer. *Nat. Med.* **1**(5), 464–470.
doi: 10.1038/nm0595-464
34. Duvvuri V.R., Duvvuri B., Cuff W.R., Wu G.E., Wu J. (2009) Role of positive selection pressure on the evolution of H5N1 hemagglutinin. *Genomics Proteomics Bioinformatics*. **7**(1–2), 47–56.
doi: 10.1016/S1672-0229(08)60032-7
35. Stern A., Andino R. (2016) Viral evolution: it is all about mutations. In: *Viral Pathogenesis* (3rd edition). Eds Katz M.G., Lynn Law G., Korth M.J., Nathanson N. Elsevier, pp. 233–240.
doi: 10.1016/B978-0-12-800964-2.00017-3
36. Kumar A., Hussain S., Yadav I.S., Gissmann L., Nataraajan K., Das B.C., Bharadwaj M. (2015) Identification of human papillomavirus-16 E6 variation in cervical cancer and their impact on T and B cell epitopes. *J. Virol. Methods*. **218**, 51–58.
doi: 10.1016/j.jviromet.2015.03.008
37. Kichatova V.S., Kyuregyan K.K., Soboleva N.V., Karlsen A.A., Isaeva O.V., Isagulians M.G., Mikhailov M.I. (2018) Frequency of interferon-resistance conferring substitutions in amino acid positions 70 and 91 of core protein of the Russian HCV 1b isolates analyzed in the T-cell epitopic context. *J. Immunol. Res.* **7**, 7685371.
doi: 10.1155/2018/7685371
38. Avila-Rios S., Carlson J.M., John M., Mallal S., Brumme Z.L. (2019) Clinical and evolutionary consequences of HIV adaptation to HLA: implications for vaccine and cure. *Curr. Opin. HIV AIDS*. **14**(3), 194–204.
doi: 10.1097/COH.0000000000000541
39. Lumley S.F., McNaughton A.L., Klenerman P., Lythgoe K.A., Matthews P.C. (2018) Hepatitis B virus adaptation to the CD8⁺ T cell response: consequences for host and pathogen. *Front. Immunol.* **9**, 1561.
doi: 10.3389/fimmu.2018.01561

Amino Acid Substitution Patterns in the E6 and E7 Proteins of HPV type 16: Phylogeography and Evolution

E. E. Zelenova^{1, 2, *}, A. A. Karlsen^{3, 4}, D. V. Avdoshina⁵, K. K. Kyuregyan^{3, 4}, M. G. Belikova^{3, 5, 6}, I. D. Trotsenko^{3, 7}

¹Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, 119991 Russia

²Blokhin National Medical Research Center of Oncology, Moscow, 115478 Russia

³Peoples' Friendship University of Russia, Moscow, 117198 Russia

⁴Mechnikov Research Institute of Vaccines and Sera, Moscow, 105064 Russia

⁵Chumakov Federal Scientific Center for Research and Development of Immunobiological Drugs, Russian Academy of Sciences, Moscow, 117218 Russia

⁶National Research Center for Epidemiology and Microbiology named after Honorary Academician N. F. Gamaleya, Ministry of Health of the Russian Federation, Moscow, 123098 Russia

⁷Moscow Oncological Society, Moscow, 111123 Russia

*e-mail: zelenovayeye@gmail.com

The E6 and E7 proteins of human papillomavirus (HPV) play a key role in the oncogenesis of papillomavirus infection. Data on the variability of these proteins are limited, and the factors affecting their variability are poorly understood. We analyzed the variability of the currently known sequences of HPV type 16 (HPV16) E6 and E7 proteins, taking into account their geographic origin and year of sample collection, as well as the direction of their evolution in major geographic regions of the world. All sequences belonging to HPV16 genome fragments encoding E6 and E7 oncoproteins were downloaded from the NCBI GenBank database on October 6, 2022. Samples were filtered according to the following parameters: the sequence includes at least one of the two whole open reading frames, the collection date and the country of origin are known. A total of 3,651 full-genome nucleotide sequences encoding the E6 protein and 4,578 full-genome nucleotide sequences encoding the E7 protein were sampled. The nucleotide sequences obtained after sampling and alignment were converted to amino acid sequences and analyzed using MEGA11, R, RStudio, Jmodeltest 2.1.20, BEAST v1.10.4, Fastcov, and Biostrings software. The highest variability in E6 protein structure was recorded at positions 17, 21, 32, 85, and 90, while in E7, positions 28, 29, 51, and 77 were the most variable. The samples were divided geographically into 5 heterogeneous groups: African, European, American, Southwest and South Asia and Southeast Asia. Unique amino acid substitutions (AA-substitutions) in the E6/E7 proteins of HPV16, presumably characteristic of certain ethnic groups, were identified for a number of countries. They are mainly localized in sites of known B- and T-cell epitopes and relatively rarely in structural and functional domains. The revealed differences in AA-substitutions in different ethnic groups and their colocalization with clusters of B- and T-cell epitopes suggest their possible relationship with the geographical distribution of alleles and haplotypes of the major histocompatibility complex (HLA). This may lead to the recognition of a different set of B- and T-cell epitopes of the virus, resulting in regional differences in the direction of epitope drift. Phylogenetic analysis of the nucleotide sequences encoding the E6 protein of HPV16 revealed a common ancestor, confirmed regional clustering of the E6 protein gene sequences by the set of the most common AA-substitutions, and identified cases of reversion of individual AA-substitutions when the virus distribution region changed. For the E7 protein, a similar analysis was not possible due to high sequence homology. Covariance analysis of the pooled sample revealed that there was no relationship between amino acid residues in the E6 protein, in the E7 protein, and between E6 and E7. Data obtained are important for the development of therapeutic vaccines against HPV of high carcinogenic risk.

Keywords: human papillomavirus type 16, E6 protein, E7 protein, primary structure, evolution, phylogenetic analysis, covariance analysis, B-cell epitopes, T-cell epitopes