

ВАРИАЦИОННОЕ РАЗНООБРАЗИЕ И МУТАЦИОННОЕ ДАВЛЕНИЕ ГЕНА ГЕМАГГЛЮТИНИНА ПАНДЕМИЧЕСКОГО ВИРУСА ГРИППА A/CALIFORNIA/04/2009(H1N1)

Павлов К. И., Титов Л. П., Синюк К. В., Грибкова Н. В., Шмелёва Н. П.

*Государственное учреждение «Республиканский научно-практический центр
эпидемиологии и микробиологии», город Минск, Республика Беларусь*

Резюме

После пандемии H1N1 2009 было собрано большое количество полных нуклеотидных последовательностей гена гемагглютинина – ключевой антигенной детерминанты, что позволило сформировать объективную выборку и изучить нуклеотидный полиморфизм и филогенетическое разнообразие. Было выявлено, что за период с 01.04.2009 по 31.12.2009 отмечается накопление как синонимичных, так и несинонимичных нуклеотидных замен с формированием филогенетического разнообразия и множества квазивидов. Наиболее варибельным оказался участок, кодирующий Са эпитоп, при относительной интактности Sa и Sb эпитопов. Большинство произошедших аминокислотных замен полипептида были консервативны по физико-химическим свойствам и показателю разности гидрофобностей. Выраженное колебание GC содержания нуклеотидных последовательностей не наблюдалось.

Ключевые слова: вирус гриппа, пандемия, гемагглютинин, эволюция.

Введение

Пандемическое распространение нового генетического варианта вируса гриппа A H1N1/California 2009 создало серьезную угрозу для популяции человека. Предварительные итоги и уроки пандемии свидетельствуют о достаточно выраженном эпидемическом потенциале данного вируса, быстрым распространением, большим числе тяжелых случаев инфекции, повсеместном возникновении кластеров инфекции (групповых вспышек) и летальности. Характер динамики распространения данного инфекционного агента, вероятно, обусловлен как высокой прослойкой восприимчивого населения на разных континентах и странах, интенсивными миграционными процессами людей, так и молекулярно-биологическими особенностями и фитнесом возникшего вируса. Геномические и иммунобиологические характеристики вируса и его потенциал изучены недостаточно.

Генетическая информация вируса гриппа А зашифрована в 8 сегментах одноцепочечной РНК. Они кодируют 7 белков - PB1, PB2, PA, NA, NA, NP и M1. Ключевая роль в изменчивости вирусов гриппа принадлежит поверхностному гликопротеину гемагглютинуину [1,2,3,4], который зафиксирован в перикапсиде в виде гомотримера и отвечает за взаимодействие с сиалорецепторами эпителия дыхательного тракта хозяина. Гемагглютинин состоит из двух фрагментов HA1 и HA2 (Рисунок 1). HA1 содержит сигнальный пептид, стебель и глобулу, HA2 фрагмент характеризуется конформационными изменениями молекулы, влияя на слияние перикапсида с мембраной клетки. В глобуле локализуется 4 антигенных сайта (эпитопа). Sa и Sb эпитопы расположены возле рецепторного кармана, а Ca (состоящий из 5 фрагментов) и Cb расположены в областях, где аминокислотная цепь образует петли, выступающие из поверхности глобулы. Возникающие в этих сайтах мутации по типу антигенного дрейфа определяют сезонную варибельность популяций вирусов гриппа,

интенсивность эпидемического процесса, стимуляцию определенных клонов В- и Т-лимфоцитов. Антитела, вырабатываемые к этим фрагментам в ходе иммунного ответа на инфекцию, соответственно, и определяют выраженность протективного иммунитета [4].

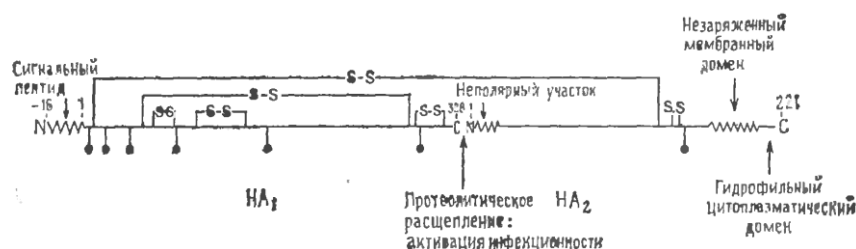


Рисунок 1 - Схематическое изображение первичной структуры НА-белка вируса гриппа. Отмечены дисульфидные мостики (-S-S-), сайты присоединения углеводов (вертикальные черточки с чёрным кружком) и другие значимые области.

После получения первых просеквенированных последовательностей (Porter, Barber, Carey, 1979) вирус гриппа стал одним из основных объектов в исследованиях молекулярной эволюции и эпидемиологии (Bush, 1999; Holmes, 2005; Zhou, 2005; Kryazhimskiy, 2008). Однако, скорость накопления фактических первичных данных (изоляты, нуклеотидные последовательности, данные классической эпидемиологии) по-прежнему преобладают над их глубоким анализом и интерпретацией. В ряде работ последних лет анализировалась как вся совокупность доступных последовательностей (Plotkin, Dushoff, 2002; Wang, Smith, 2010), так и выборка определённых штаммов, зачастую именно пандемических (Both, 1983; Jhang-Wei, 2011; Huang, 2009). Наиболее оптимальны для таких исследований полные и равные по длине последовательности, которых, к тому же, имеется ещё и достаточное количество. При изучении неполных последовательностей или фрагментов возможно, конечно, выравнивание и анализ только совпадающих участков, как, в частности, было сделано для H3 молекулы H3N2 вируса (Kryazhimskiy, 2008). Но такой фрагмент не большой, не даёт представления о функциональной изменчивости молекулы, и не возможен для проекции 3D структуры. Полные по длине последовательности гена гемагглютинина длиной 1701 нуклеотидов (как и в H1N1 2009) проанализированы Igarashi с реконструированного A/South Carolina/1/1918 до A/Brisbane/59/2007 с построением 3D структуры (сравнивался и пандемический штамм 2009 года) [5], но подобная выборка содержит иногда по 1 штамму на 10 лет, что не даёт полного представления о картине мутагенеза.

В динамике развития пандемии гриппа А 2009 г. в разных странах мира вирусологическими лабораториями в режиме реального времени от пациентов с пандемическим гриппом выделялось множество изолятов, осуществлялось как полногеномное, так и сегментарное секвенирование, накоплено огромное количество полных сиквенсов, что является первичным и удобным материалом для проведения филогенетического анализа, оценки популяционного разнообразия и трансформации пандемического штамма в сезонный, оценки тенденций молекулярной эволюции патогена.

Цель настоящей работы – изучение темпов дивергенции, характера нуклеотидных замен в последовательностях гена гемагглютинина и оценка мутационного давления за период интрадукции вируса в популяцию человека.

Материалы и методы

Материалы. Источник нуклеотидных последовательностей - Influenza Research Database and Influenza virus resource, NCBI. В данной работе были использованы только *полные* нуклеотидные последовательности гена НА пандемических штаммов длиной 1701 нуклеотидов, изолятов из 35 стран мира. Сроки выделения штаммов: 01.04.2009 - 31.12.2009 - течение пандемии. При изучении *мутационного давления* [6,7] в выборку были включены также вирусы, эквивалентные пандемическим, выделенные от животных в данный период, содержащие информацию о месяце выделения (расширенная выборка составила 1370 сиквенсов). Для исследования были отобраны *все* сиквенсы, подходящие по обозначенным критериям, содержащиеся в базе данных NIH (NIAID) на 01.05. 2011. Соответственно, при составлении выборки последовательностей гена НА в название была добавлена *дата выделения штамма*. За первые 6 недель вирус распространился в 213 стран мира по глобальным коммуникациям. Пик заболеваемости гриппом для большинства регионов мира приходится на ноябрь 2009 года. В 2009 году имеют место два пика сбора образцов и секвенирования – апрель и ноябрь. При изучении полиморфизма нуклеотидных последовательностей и филогении использованы только штаммы, выделенные от людей с указанием дня и месяца выделения вируса (всего 1170).

Методы. Анализ нуклеотидных последовательностей осуществлялся с помощью пакета программ MEGA 5. Полиморфизм нуклеотидных последовательностей оценивался при помощи программы DnaSP 5.10.01. Подсчёт GC показателей произведён в пакете программ MEGA 5, а также алгоритмами VVK34 и VVK Consensus. Исследовался коэффициент корреляции между общим GC содержанием и GC содержанием в каждой позиции кодонов. Для подтверждения полученных результатов было рассчитано среднее 1GC, 2GC, 3GC, GC, отдельно G и C показатели содержания для каждого месяца (исходя из количества собранных за этот месяц последовательностей), для объективизации результатов расчет велся в абсолютных числах нуклеотидов, тенденция оценивалась по направлению линейного тренда и значениям стандартного отклонения и доверительного интервала. Оценка характера аминокислотных замен проводилась по критериям Снита, Волькенштейна и Бачинского [8]. Метод Снита появился в 1966 году. Он базируется на оценке всех изученных к тому моменту физико-химических свойств аминокислот (наличие/отсутствие –ОН групп, растворимость L-изомеров, наличие бензольного кольца, свободные электроны и т. д.) на положение в белковой макромолекуле. При этом все эти признаки, априори, приняты как равнозначные. Если коэффициент Снита больше 0,416 то замена одной аминокислоты на другую считается консервативной, если меньше, то радикальной [8]. Волькенштейн предложил использовать разность гидрофобностей в качестве показателя взаимозаменяемости аминокислот. Для этого пользуются шкалой гидрофобностей аминокислот, установленной Ч. Танфордом. Замена считается консервативной при разности гидрофобностей менее 1,28 ккал/моль для замен в целом и 1,22 ккал/моль для одношаговых замен. Бачинский в 1976 году предложил «показатель функциональной близости аминокислот» при их заменах. Показатель этот получен при анализе аминокислотных замен в изофункциональных белках. Матрицы изофункциональных замен построены, используя 28 семейств белков. Замена одной АК на другую считается консервативной при ФБА больше 12,4 для одношаговых замен. При этом за анцестральный штамм был взят Influenza A virus (A/Mexico/3955/2009(H1N1)). Филогенетический анализ осуществлялся с использованием метода связывания ближайших соседей (Neighbour-joining). Эволюционные дистанции измерялись при помощи метода Maximum Composite

Likelihood. Визуализация филогенетических деревьев осуществлялась с помощью программы FigTree v.1.3.1.

Статистика. Полученные данные были обработаны методами описательной статистики, для оценки корреляции использовался коэффициент регрессии.

Результаты

При изучении полиморфизма нуклеотидных последовательностей заметно возрастание количества переменных сайтов, причём как синглетонных, так и парсимоничных с течением пандемии (Рисунок 2). Установлено, что при попадании вируса в новую для себя популяцию хозяина, происходило вовлечение всё большего числа сайтов в мутационный процесс по мере распространения среди восприимчивой части населения. Среднее число нуклеотидных замен на последовательность (k) также поступательно повышается (от 3,442 в апреле до 8,756 в декабре), что говорит о накоплении в популяции происходящих по переменным сайтам мутаций.

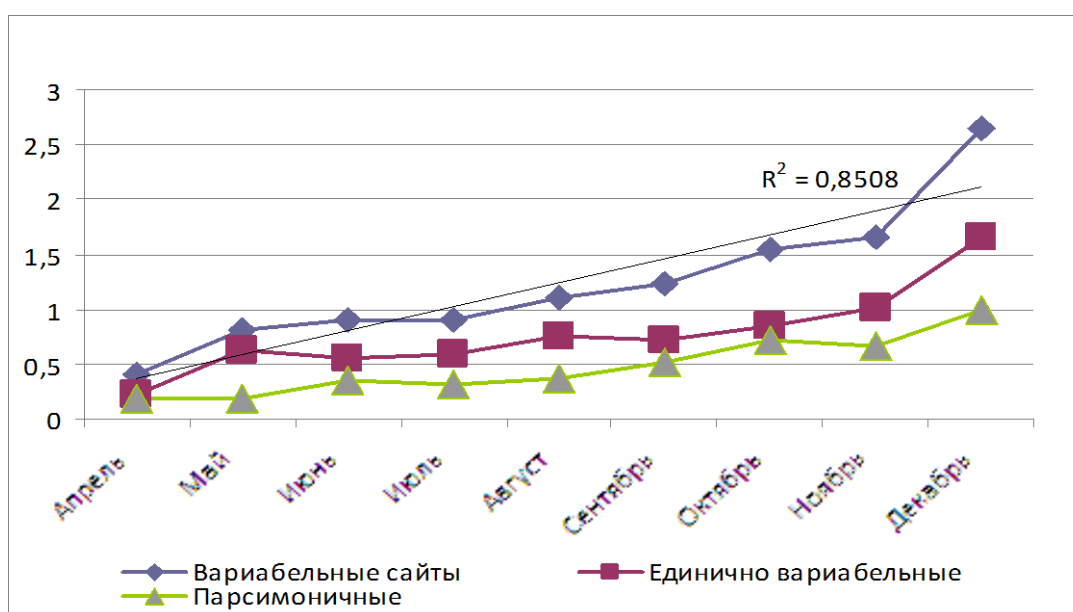


Рисунок 2 – Динамика нарастания количества переменных сайтов

Показатели изменчивости, существенно изменяются в зависимости от количества секвенсов, приходящихся на определенный месяц. Поэтому показатели полиморфизма дополнительно делились на число имеющихся за данный месяц последовательностей (Рисунок 2).

Значительный интерес представляет внутримолекулярная изменчивость гемагглютинаина [6,7], особенно его иммунодоминантных эпитопов. При определении наиболее переменных участков нуклеотидной последовательности резко превалирует один из эпитопов – Sa, показатель нуклеотидных замен на сайт в котором почти в три раза выше, чем во всей остальной последовательности (Рисунок 3). Такая закономерность сохраняется как при использовании всей выборки, так и при составлении равного количества нуклеотидных последовательностей за каждый месяц (1370 и 90 секвенсов соответственно). Относительно всей молекулы высок также уровень нуклеотидных замен на сайт в сигнальном участке. Интактными выглядят Sa и Sb эпитопы, расположенные в области рецепторсвязывающего кармана. Любопытна также меньшая активность в участках, кодирующих глобулу сравнительно с

фрагментами, кодирующими стебель, трансмембранный домен и конформационно перестраивающийся HA2 участок молекулы.

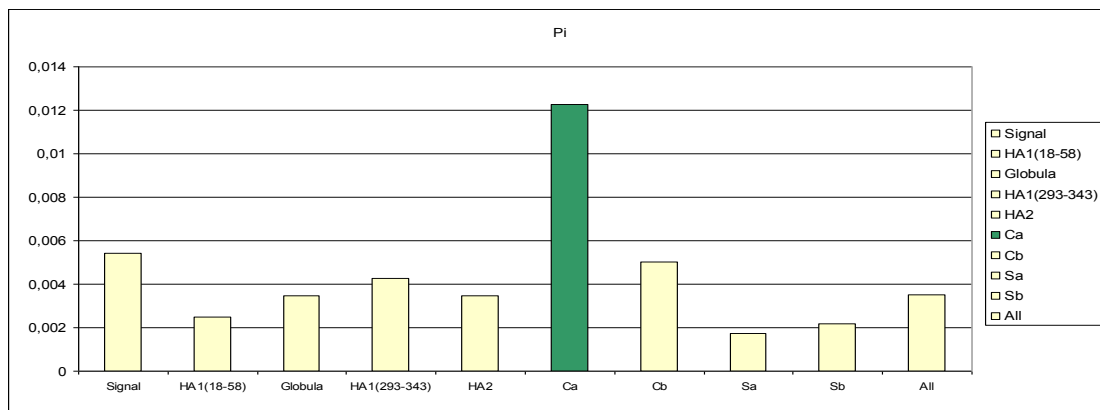


Рисунок 3 – Число нуклеотидных замен на сайт внутри молекулы гена гемагглютинаина

При оценке синонимичных и несинонимичных нуклеотидных замен данный эпитоп также резко выделяется на фоне всей молекулы (Рисунок 4). Разность средних значений показателей синонимичных и несинонимичных аминокислотных замен (dS-dN) является отрицательным числом, т.е. несинонимичные нуклеотидные замены значительно преобладают. Следовательно, преобладают аминокислотные замены. Число нуклеотидных замен на сайт для Sa и Sb эпитопов ниже, чем в целом по молекуле. Глобула, рассматриваемая как основной накопитель мутаций, не преобладает над фрагментами стебля.

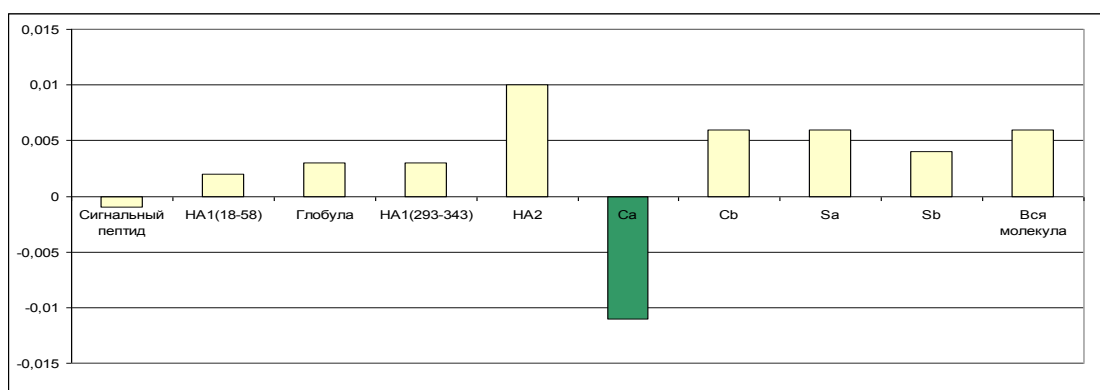


Рисунок 4 – Разность между уровнями синонимичных и несинонимичных нуклеотидных замен в молекуле гена гемагглютинаина

Корреляция между динамикой Pi и аминокислотных замен ($R^2=0,910$) говорит о высоком уровне несинонимичности замен и накоплении аминокислотных изменений в молекуле HA за данный период. Оценка характера аминокислотных замен проводилась после филогенетического анализа с каждым верхним и нижним сиквенсом выявленных трёх больших клайдов со штаммом A/Mexico/3955/2009(H1N1). В 97,43% случаев аминокислотные замены были консервативны по своим физико-химическим свойствам (критерии Снита и Бачинского) и показателю разности гидрофобностей (критерий Волькенштейна), что говорит о значительной функциональной консервативности структуры молекулы гемагглютинаина и стабилизирующем отборе.

Мутационное давление - важный фактор молекулярной эволюции – характеризуется увеличением или снижением GC содержания всего генома и отдельных позиций кодонов. Для оценки этого фактора исследовалась корреляция колебания GC содержания генома в целом и каждой позиции кодонов (Рисунок 5.)

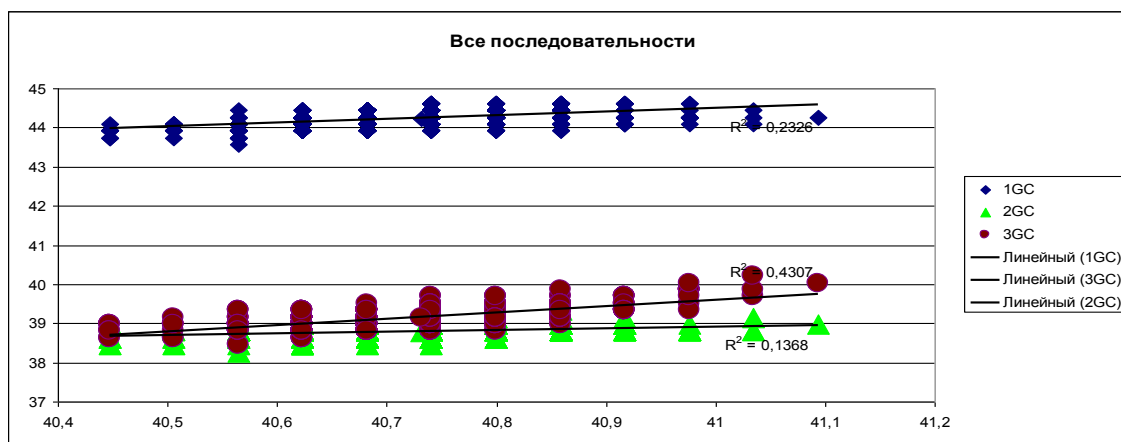


Рисунок 5 – Корреляция GC трёх позиций кодонов и общего GC содержания

Значительной корреляции с общим GC содержанием не наблюдается ни для одной позиции кодонов: для первой позиции кодонов $R^2=0,231$, для второй $R^2=0,1362$, для третьей $R^2=0,4307$. Таким образом, только для третьей позиции кодонов прослеживается некоторая корреляция. При изучении средних GC показателей рассчитанных в абсолютных числах нуклеотидов отмечается рост 3GC уровня (с 221,4 GC нуклеотидов в апреле до 222,3 GC нуклеотидов в декабре) на 0,9 нуклеотида. Общее GC и 2GC показатели остаются практически неизменными. 1GC уровень снижается (с 251,4 GC нуклеотидов в апреле до 250,8 GC нуклеотидов в декабре) на 0,6 нуклеотида. Однако, на данные расчеты значительное влияние оказывает имеющееся за месяц число последовательностей (что выражается в больших значениях доверительного интервала). Тем не менее, варибельность 1GC показателя говорит о несинонимичности нуклеотидных замен [7].

Филогенетический анализ всей выборки выявил широкую дивергенцию образованием большого количества жизнеспособных субклайдов (Рисунок 6, А). При изучении филогении для отдельных эпитопов, также наблюдается дивергенция с появлением закрепившихся в популяции мутантов, причём для Са эпитопа эта дивергенция гораздо шире, чем в других эпитопах (Рисунок 6, Б,В). Отмечается значительное расширение многообразия популяции пандемического вируса, образование многочисленных жизнеспособных субклайдов, содержащих свободно циркулирующие в популяции квазивиды.

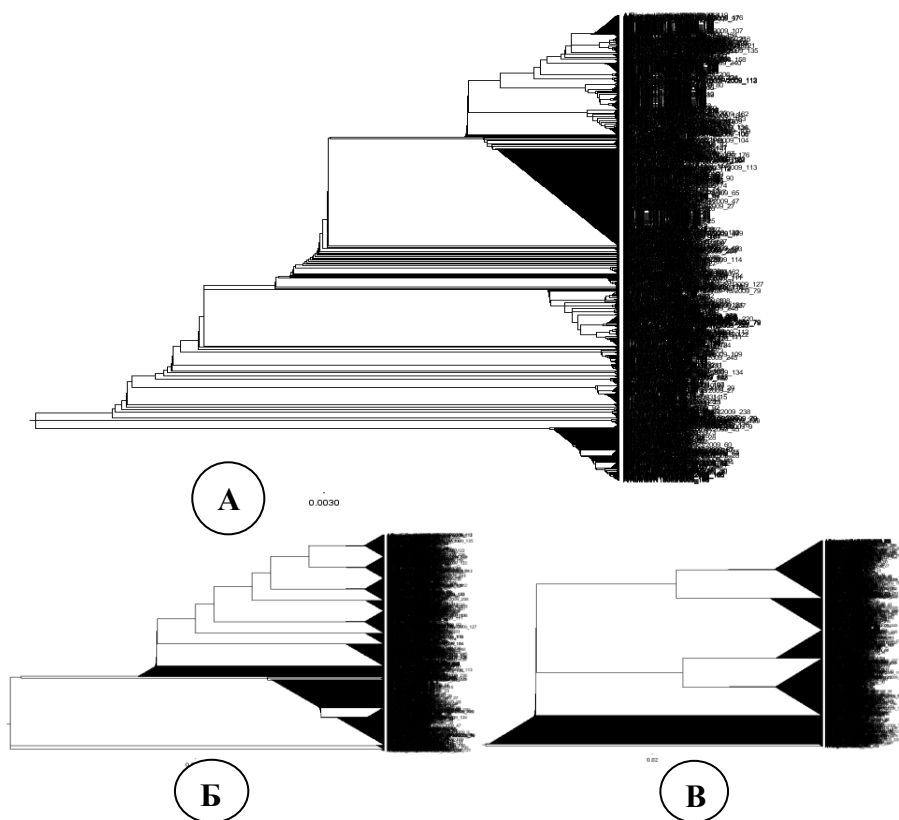


Рисунок 6 – Сравнение филогенетического разнообразия в Sa (А) и Sb (Б) эпитопах и всей молекуле (В)

Обсуждение. Полученные данные о высоких темпах как синонимичных, так и несинонимичных нуклеотидных замен последовательностей гена гемагглютинина для пандемических вирусов H1N1 2009 совпадают с имеющимися данными анализа всей совокупности предыдущих сиквенсов (Kryazhimskiy, Bazykin, 2008). Отбор для H3N2 пандемии с 1968 по 1990 и, особенно, с 1990 по 2007, шёл в направлении сохранения одной доминирующей линии, без явления совместной циркуляции разнообразных изолятов.[1,4] Поэтому можно предположить дальнейшее урезание полученной ветвистой дендрограммы (Рисунок 6, А) и элиминацию большинства возникших квазивидов. Отмеченная взаимозаменяемость аминокислотных замен (критерии Снита, Волькенштейна, Бачинского) рассчитана только в крайних последовательностях кластеров дендрограмм и не учитывает все мутации, тем не менее указывает на структурно-функциональную консервативность молекулы гемагглютинина.

Участки, кодирующие иммунодоминантные эпитопы, подвергаются более сильному селекционному давлению, чем остальная часть молекулы (Рисунок 7). Известно, что для птичьих вирусов антигенный дрейф не характерен – за год-два популяция полностью меняется и иммунологический пресс фактически отсутствует. В популяциях свиней дрейф идёт гораздо медленнее, чем среди людей. В свете этого любопытна сравнительная интактность Sb и Sa эпитопов (Рисунок 3) выявленная в ходе данной работы. То же самое наблюдалось в ранних известных потомках H1N1 1918 - A/Puerto Rico/8/1934 и A/Bellamy/1942.[5] Причём уже в штаммах 50-х годов содержатся значительные аминокислотные замены, как если изменения в Sa и Sb произошли в конце сороковых годов, более чем через 20 лет после пандемии 1918 года.(Igarashi, 2008).

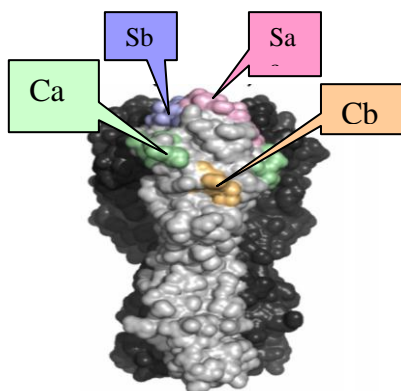


Рисунок 7 - пространственная топография иммунодоминантных эпитопов гемагглютинина

В Са эпитопе, состоящем из 5 фрагментов (6+5+3+2+3 аминокислот) уровень нуклеотидных замен на сайт был довольно высоким (около 0,012). Возникает вопрос, чем можно объяснить такой высокий показатель мутаций, учитывая иммунологическую наивность популяции к штамму A/California/04/2009, и предположительно полное отсутствием иммунной прослойки в популяции? Как для выборки пандемических штаммов гена гемагглютинина 01.04.2009 - 31.12.2009, так и для аналогичной выборки последовательностей гена нейраминидазы, не было отмечено значительной корреляции между общим GC содержанием и GC в каждой позиции кодонов. Известно, что для вирусов гриппа (в отличии ,скажем, от вирусов герпеса) колебания AT и GC содержания геномов не носят выраженный, тенденциозный характер, варьируя скорее в пределах неких усреднённых показателей с незначительными подъёмами и спадами.

Значительный интерес представляет дальнейшая филодинамика пандемического штамма при его трансформации в сезонный возбудитель эпидемий – как долго будет сохраняться возникшее генетическое разнообразие, какие субклаиды окажутся наиболее жизнеспособными? В следующие несколько лет станет понятно, будет ли «мексиканский» вирус интенсивно мутировать или сохранит подобие анцестральному. Образовавшийся теперь иммунный пресс окажет определённое влияние на антигенный дрейф, в свете чего любопытно, сохранится ли и в дальнейшем мутационная интактность Sa и Sb эпитопов.

Заключение

Биоинформатический анализ выборки полных нуклеотидных последовательностей гена гемагглютинина пандемических штаммов H1N1 2009 собранных за период с 01.04.2009 по 31.12.2009 выявил следующие закономерности:

1. Возникновение и накопление как синонимичных, так и несинонимичных нуклеотидных замен ассоциируется с повышением филогенетического разнообразия и формированием множества квазивидов вируса;

2. Большинство произошедших аминокислотных замен полипептида были консервативны по физико-химическим свойствам и показателю разности гидрофобностей.

3. Показатель нуклеотидных замен на сайт в участке, кодирующем Са эпитоп почти в три раза выше в сравнении таковым для других фрагментов. Большинство этих замен несинонимичны, а наиболее мутационно-инертными оказались Sa и Sb эпитопы;

4. За период пандемии не отмечалось значительного колебания ни общего GC содержания, ни GC содержания трёх позиций кодонов.

Использованная литература:

1. Both GW, Sleight MJ, Cox NJ, Kendal AP. Antigenic Drift in Influenza Virus-H3 Hemagglutinin from 1968 to 1980 - Multiple Evolutionary Pathways and Sequential Amino-Acid Changes at Key Antigenic Sites. *Journal of Virology*. 1983;48(1):52–60
2. Bush RM, Bender CA, Subbarao K, Cox NJ, Fitch WM. Predicting the evolution of human influenza A. *Science*. 1999;286(5446):1921–1925. doi: 10.1126/science.286.5446.1921
3. Kryazhimskiy, S., Bazykin, G., Plotkin, J., Dushoff, J. Directionality in the evolution of Influenza A haemagglutinin. *Proc. R. Soc. B.* (2008) 275, 2455-2464.
4. Jhang-Wei Huang and Jinn-Moon Yang. Changed epitopes drive the antigenic drift for influenza A (H3N2) viruses. *BMC Bioinformatics*. 2011; 12(Suppl 1): S31. PMID: PMC3044287
5. Igarashi M, Ito K, Yoshida R, Tomabechi D, Kida H, et al. (2010) Predicting the Antigenic Structure of the Pandemic (H1N1) 2009 Influenza Virus Hemagglutinin. *PLoS ONE* 5(1): e8553. doi:10.1371/journal.pone.0008553
6. Титов Л.П., Вотяков В.И. Геномическо-протеомические основы молекулярной эволюции вирусов. *Весті НАН Беларусі. Сер. Мед. Навук.*-2011. №1. С.109-116.
7. Титов Л.П., Вотяков В.И.. Вирусы, вириды, прионы: эволюция, таксономия и номенклатура. – *Весті НАН Беларусі. Сер. Мед. Навук.*-2008. №4. С.23-34.
8. Sneath, P. H. A. “Relations between Chemical Structure and Biological. Activity in Peptides”. *J. Theor. Bioi.* 1966, 12, 157-195

Вариационное разнообразие и мутационное давление в популяции пандемических вирусов гриппа A/California/04/2009(H1N1).

Павлов Кирилл Игоревич, Титов Леонид Петрович, Синюк Константин Викторович, Грибкова Наталья Васильевна, Шмельёва Наталья Петровна

Государственное учреждение «Республиканский научно-практический центр эпидемиологии и микробиологии», город Минск, Республика Беларусь. Улица Филимонова, д. 23.

Abstract.

After 2009 pandemic, there are a great number of complete Influenza virus sequences stored at NCBI, including sequences with the same nucleotide length. So we formed HA gene complete sequences group from 01.04.2009 to 31.12.2009 for polymorphism estimation and phylogenetic analysis. We relived significant synonymous and nonsynonymous rate for nucleotide substitutions especially in Ca antigenic epitop, controversy to Sa and Sb epitops. Amino acid substitutions were mostly conserved in their physical and chemical features and hydrophoby. We relived small role for GC mutation pressure in HA gene evolution for this period.

Key words: Influenza virus, pandemic, hemagglutinin, evolution.

Павлов Кирилл Игоревич, аспирант, ассистент кафедры микробиологии, вирусологии, иммунологии, БГМУ, Минск. (017) 245-33-19, +375-57 10-911, zabrrr2008@rambler.ru

Титов Леонид Петрович, доктор медицинских наук, профессор, член-корреспондент НАН РБ, заслуженный деятель науки РБ, заведующий лабораторией экспериментальной и клинической микробиологии, Государственное учреждение «Республиканский научно-практический центр эпидемиологии и микробиологии», город Минск, Республика Беларусь. Тел.: +375-127 67 69 98. E-mail : leonidtitov@tut.by

Синюк Константин Викторович, студент пятого курса БГМУ, Минск Тел.: +375-29 235 6989 E-mail: КО-17@yandex.ru

Грибкова Наталья Васильевна, кандидат медицинских наук, заведующая лабораторией гриппа и гриппоподобных заболеваний, Государственное учреждение «Республиканский научно-практический центр эпидемиологии и микробиологии», город Минск, Республика Беларусь. Тел.: +375-172 37 62 95 E-mail: gribkovanatalia0@gmail.com

Шмелёва Наталья Петровна, кандидат медицинских наук, ведущий научный сотрудник, Государственное учреждение «Республиканский научно-практический центр эпидемиологии и микробиологии», город Минск, Республика Беларусь. Тел.: +375-172 3762 95, E-mail: shmelevanataliya@mail.ru

Павлов Кирилл Игоревич _____

Титов Леонид Петрович, доктор медицинских наук,
профессор, член-корреспондент НАН РБ,
заслуженный деятель науки РБ _____

Синюк Константин Викторович _____

Грибкова Наталья Васильевна
кандидат медицинских наук _____

Шмелёва Наталья Петровна
кандидат медицинских наук _____