

Поиск адаптивной эволюции после событий реассортации в генах вируса гриппа А типа H3N2

Ксения Лежнина
Лаборатория
эволюционной геномики
ФББ МГУ
oxia.com@gmail.com

Алексей Неверов
Лаборатория
эволюционной геномики
ФББ МГУ
neva_2000@mail.ru

Аннотация

В работе было исследовано, как события реассортации генов влияют на значение отношения d_N/d_S в вирусе гриппа. Для последовательностей генов были построены филогенетические деревья (Mr.Bayes), с помощью Giraf были определены реассортанты. Посредством HyPhy, была проверена гипотеза о том, что события реассортации увеличивают отношение d_N/d_S . Для гена нейраминидазы было показано, что наблюдается сильное увеличение значения d_N/d_S на ветвях сразу после событий реассортации.

1. Введение

Вирус гриппа А, как правило, вызывает заболевание средней или сильной степени тяжести. Поражает как человека, так и некоторых животных (лошадь, свинья, хорек) и птиц. Именно вирусы гриппа А ответственны за появление пандемий и тяжелых эпидемий.

Вирус гриппа А способен быстро мутировать в процессе репликации. Кроме того, вирус гриппа А может эволюционировать посредством реассортации и рекомбинации. Глобальные пандемии были вызваны вирусами гриппа, которые прежде подверглись процессам реассортации (H1N1, H2N2, H3N2). Однако в настоящее время влияние реассортаций на эволюцию вируса гриппа изучено недостаточно [10].

Целью настоящей работы является исследование реассортаций вируса гриппа H3N2 и их влияния на эволюцию вирусных белков.

Для определения событий реассортации была использована программа Giraf [2], которая определяет множества последовательностей, являющиеся потомками одного события реассортации для каждой пары сегментов. Затем на близких расстояниях от событий реассортации был измерен селективный отбор (d_N/d_S) с помощью пакета HyPhy[6].

2. Постановка задачи

Пандемии могут возникнуть вслед за неожиданным антигенным сдвигом, появившимся в результате реассортации. Реассортантный вирус гриппа обладает новой комбинацией антигенов, к которой в популяции еще не выработан иммунитет – сегментированный геном вируса гриппа только способствует обмену РНК-сегментами между разными штаммами.

В настоящее время влияние реассортаций на эволюцию вируса гриппа мало исследовано [10].

Ранее было показано, что реассортация, приводящая к смене хозяина, вызывает кратковременный положительный отбор в NS гене гриппа [1]. Смена антигенных вариантов, например WU95 на SY97 по времени согласуется с реассортацией [Holmes 2005]. Целью настоящей работы является определить частоту реассортаций вируса гриппа А H3N2 и их влияние на эволюцию вирусных белков.

3. Материалы и методы

Для решения поставленной задачи необходимо:

1. Построить выравнивание по каждому гену вируса гриппа

2. Построить филогенетические деревья для каждого из генов вируса гриппа
3. Определить события реассортаций на каждом филогенетическом дереве
4. Сравнить селективный отбор на ветках дерева сразу после событий реассортации и отбор на остальных ветках дерева.

3.1. Последовательности

Все последовательности геномов вируса гриппа А типа H3N2 (всего 2205) были получены из GenBank [7]. Было исследовано 7 генов из 8 для всех имеющихся последовательностей вируса гриппа типа H3N2.

3.2. Построение выравнивания генов и кластеризация последовательностей

Выравнивание для каждого из генов было построено с помощью программы MUSCLE. Неоднозначные символы в последовательностях были заменены на консенсусный нуклеотид в колонке выравнивания.

Последовательности гена NA были собраны в кластеры с уровнем сходства более 99% с помощью CD-HIT-EST [11]. Всего было получено 126 кластеров, каждый кластер был представлен единственной последовательностью, в результате было сформировано множество из 126 геномов, сегменты которых образуют выравнивания, используемые в работе. Из выравниваний последовательностей сегментов M, PB1 были удалены области, кодирующие перекрывающиеся рамки считывания.

3.3. Построение филогенетических деревьев

Консенсусные филогенетические деревья были построены с помощью MrBayes [3][4] (outgroup=A/Hong_Kong/1-11/1968 mcmc ngen=1000000 samplefreq=1000).

3.4. Определение реассортантов

Реассортанты были определены с помощью Giraf [2]. На филогенетическом дереве реассортанты можно определить по топологической несовместимости двух деревьев: если на дереве первого гена клада группируется с одной группой последовательностей, а на дереве второго гена – с другой группой последовательностей, тогда можно считать, что клада образована реассортантами.

Giraf, используя деревья, построенные MrBayes, представляет каждое дерево в виде набора разбиений вершин. Два разбиения A|B и X|Y несовместимы, если все четыре пересечения A∩X, A∩Y, B∩X и B∩Y не пусты. Тогда, два дерева филогенетически несовместимы тогда и только тогда, когда они содержат несовместимые разбиения вершин. Далее Giraf строит граф несовместимостей для каждой пары последовательностей так, что узлами в этом графе являются разбиения каждого из деревьев, а ребро соединяет 2 узла, если 2 разбиения несовместимы. Таким образом, в этом графе формируются биклики, которые могут служить свидетельством реассортации [2].

Программа Giraf определяет множества последовательностей, которые являются потомками одного события реассортации для каждой пары сегментов. На филогенетическом дереве для каждого сегмента были определены положения всех реассортаций, который данный сегмент образует в паре с некоторым другим сегментом.

3.5. Поиск адаптивной эволюции

Селективный отбор измеряется отношением несинонимичных (d_N) замен к синонимичным (d_S) на сайт; $d_N > d_S$ указывает на возможность существования положительного отбора, $d_N < d_S$ – стабилизирующего отбора [5].

Для измерения адаптивной эволюции был использован пакет HyPhy [6] (*Hypothesis testing using Phylogenies*), анализ Positive Selection: TestBranchdNdS.bf. Этот анализ позволяет проверить гипотезу, значительно ли отличается сила отбора, действующая на выбранные ветви от отбора, действующего на остальную часть дерева (локальные значения $\omega = d_N/d_S$ для выбранных ветвей дерева отличаются ли от глобального ω по всему дереву).

В качестве результата выдаются значения p-value, логарифм отношения правдоподобия для глобальной модели и модели допускающей индивидуальные значения ω на выбранных ветвях дерева, значения ω для каждой выбранной ветви и значение ω для всего дерева в глобальной модели.

Параметры: Genetic code = Universal (универсальный генетический код), site-to-site rate variation model = no site-by-site rate variation (модель изменения параметров при переходе от сайта к сайту), Nucleotide substitution model = MG94xHKY85 ([8] [9]) (выбор нуклеотидной модели замен), Amino acid class model = Default

(выбор модели аминокислотных замен: 1 класс несинонимичных замен, все замены равнозначны).

Выбор ветвей: локальное значение ω было разрешено на ветвях дерева, находящихся на маленьком расстоянии от события реассортации. Расстояние измерялось в dS , для TestBranchdNdS.bf выбирались ветви с $dS < 0,001$ от события реассортации, определенного Giraf. Значение dS для каждой ветки было получено также HyPhy: анализ AnalyzeCodonData.bf (выбрана локальная модель подсчета ω).

Для каждой ветки дерева расстояние dS до ближайшей реассортации вычислено поиском в ширину. На ветви с реассортацией $dS = 0$. Если у предков не было событий реассортации, то dS не определено.

Для значений dS была построена зависимость значения d_N/d_S от dS . Также построена таблица распределения значений d_N/d_S такая, что в категорию $95\%CI_{dN/dS} > A$ ($dN/dS > A$) - попадают все ребра, когда нижняя граница $95\%CI_{dN/dS}$ (значение dN/dS) больше порога (границы доверительного интервала берутся из результатов TestBranchdNdS.bf), все остальное попадает в другую категорию: $95\%CI_{dN/dS} \leq A$ ($dN/dS \leq A$), где A – пороговое значение d_N/d_S равное среднему значению d_N/d_S по дереву.

4. Результаты

На рисунках 1-2 показаны значения d_N/d_S в зависимости от расстояний dS до ближайшей реассортации для генов NA, NP и PA. В таблицах 1-2 представлены распределения d_N/d_S ветвей дерева по двум категориям – сразу после реассортации (расстояние $dS \leq 0,001$) и остальные ветви.

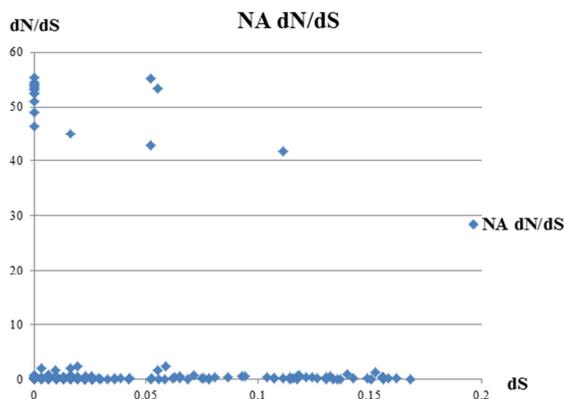


Рисунок 1. Распределение значений dN/dS в зависимости от расстояний dS до ближайшей реассортации для нейраминидазы

Таблица 1. Характеристики гистограммы NA dN/dS

	$dS \leq 0,001$	$dS > 0,001$
$d_N/d_S \leq 1$	18	133
$d_N/d_S > 1$	12	12
$95\%CI_{dN/dS} \leq 0,255$	17	132
$95\%CI_{dN/dS} > 0,255$	13	13

Согласно тесту Фишера, оба распределения значений d_N/d_S не случайны ($p_{value} < 0,0001$). Ветви с большим значением d_N/d_S расположены на очень маленьких расстояниях dS от событий реассортации.

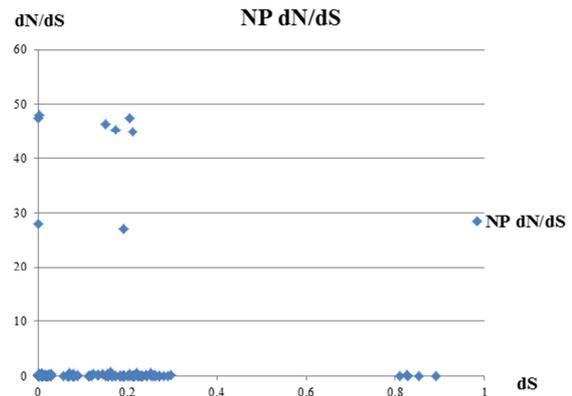


Рисунок 2. Распределение значений dN/dS в зависимости от расстояний dS до ближайшей реассортации для нуклеоптеидного белка

Таблица 2. Характеристики гистограммы NP dN/dS

	$dS \leq 0,001$	$dS > 0,001$
$d_N/d_S \leq 1$	26	165
$d_N/d_S > 1$	2	6
$95\%CI_{dN/dS} \leq 0,0758$	26	157
$95\%CI_{dN/dS} > 0,0758$	2	14

Как видно на рисунке 3, для гена NP только 2 значения с повышенным d_N/d_S лежат близко к событиям реассортации (тест Фишера: $p_{value} > 0,05$).

Данные для NA, MP, PB1, PB2, PA не представлены, но для них также были проведены расчеты.

Далее были получены результаты работы HyPhy (TestBranchdNdS.bf) для гена NA.

Таблица 3. Результат для гена NA

Ген	p-value	-2(Ln Likelihood Difference)	Количество реассортаций
NA	0,020767	7,7976	20

5. Обсуждение

Мы обнаружили, что реассортации могут вызывать увеличение d_N/d_S сразу после событий реассортации (на расстояниях $dS < 0,001$) в гене NA (см. Таб. 1).

Эволюционная модель, допускающая индивидуальные значения параметра d_N/d_S на ветвях дерева, находящихся на коротких временах в значениях dS от ближайшего события реассортации (локальная модель), значительно лучше глобальной модели, допускающей единственное значение этого параметра на всех ветвях, для гена NA, что, видимо, является следствием кратковременного увеличения d_N/d_S (табл. 1 и 3). Для остальных генов не обнаружено зависимости величины отбора от наличия реассортаций.

На рисунках 2-3 видно, что для всех генов есть значения с высоким d_N/d_S , одиночно лежащие на больших расстояниях от реассортаций. Вероятно, эти значения не связаны с событиями реассортации, а указывают на произошедшую полезную для вируса гриппа мутацию, которая быстро закрепилась в популяции вируса гриппа, однако, с другой стороны, возможно, эти значения с высоким d_N/d_S отвечают не найденным Giraf событиям реассортации.

Возможно, Giraf не смог определить все события реассортации по причине того, что не все гены были проанализированы (не были исследованы гены NS1, NS2, M2). Кроме того, важную роль в данной работе играет порог кластеризации последовательностей, который также влияет на количество реассортантов, которые могут быть определены Giraf. То есть, при увеличении количества последовательностей, что значит увеличении порога кластеризации, количество возможных реассортантов будет возрастать, что вероятно может увеличить значимость высоких d_N/d_S , связанных с событиями реассортации. Таким образом, для понимания эволюции гена гемагглютинина необходимо его дальнейшее изучение.

6. Выводы

1. Были определены реассортанты для генов NA, HA, NP, PA, PB1, PB2, MP.
2. Для гена нейраминидазы показано, что реассортации могут приводить к увеличению d_N/d_S на близких к событиям реассортации расстояниях ($dS < 0,001$)
3. Для остальных генов такого влияния на эволюцию генов не обнаружено.

7. Список литературы

- [1] Vijaykrishna D, Smith GJ, Pybus OG, Zhu H, Bhatt S, Poon LL, Riley S, Bahl J, Ma SK, Cheung CL, Perera RA, Chen H, Shortridge KF, Webby RJ, Webster RG, Guan Y, Peiris JS. *Long-term evolution and transmission dynamics of swine influenza A virus. Nature. 2011 May 26;473 (7348):519-22.*
- [2] Niranjan Nagarajan and Carl Kingsford *GiRaF: robust, computational identification of influenza reassortments via graph mining Nucleic Acids Research, 2010, 1–10*
- [3] Huelsenbeck, J.P. and Ronquist, F. (2001) *MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics, 17, 754–755.*
- [4] Drummond, A.J. and Rambaut, A. (2007) *BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol., 7, 214.*
- [5] Suzuki, Y. *Natural selection on the influenza virus genome. Mol. Biol. Evol. 23, 1902–1911 (2006).*
- [6] <http://www.datam0nk3y.org/hyphy/doku.php?id=start>
- [7] <http://www.ncbi.nlm.nih.gov/genbank/>
- [8] S. V. Muse and B. S. Gaut. *A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. Mol Biol Evol, 11:715–724, 1994.*
- [9] N. Goldman and Z. Yang. *A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol, 11(5):725–736, Sep 1994.*
- [10] Nelson MI and Holmes EC (2007). *The evolution of epidemic influenza. Nature Reviews Genetics 8: 196–205*
- [11] <http://nebc.nerc.ac.uk/bioinformatic/s/docs/cd-hit-est.html>