



*На правах рукописи*

Сеплярский Владимир Борисович

**Паттерны и молекулярные механизмы  
мутагенеза у эукариот**

03.01.09 - математическая биология, биоинформатика

Автореферат  
диссертации на соискание степени  
кандидата биологических наук

Москва, 2015

Работа выполнена на факультете биоинженерии и биоинформатики Федерального государственного бюджетного образовательного учреждения высшего профессионального образования Московского государственного университета имени М.В. Ломоносова и в секторе молекулярной эволюции Федерального государственного бюджетного учреждения науки Институт проблем передачи информации им. А.А. Харкевича Российской академии наук.

**Научный руководитель:** Кандидат биологических наук  
**Базыкин Георгий Александрович**  
Федеральное государственное бюджетное учреждение науки  
Институт проблем передачи информации им. А.А. Харкевича Российской академии наук

**Официальные оппоненты:** Доктор биологических наук  
**Орлов Юрий Львович**  
Федеральное государственное бюджетное учреждение науки  
Институт цитологии и генетики Сибирского отделения  
Российской академии наук

Кандидат физико-математических наук  
**Кулаковский Иван Владимирович**  
Федеральное государственное бюджетное учреждение науки  
Институт молекулярной биологии имени В.А. Энгельгардта  
Российской академии наук

**Ведущая организация:** **Федеральное государственное бюджетное учреждение науки  
Институт математических проблем биологии Российской академии наук**

Защита диссертации состоится 14 декабря 2015 года в \_\_\_\_ на заседании диссертационного совета Д 002.077.04 при Федеральном государственном бюджетном учреждении науки Институте проблем передачи информации им. А.А. Харкевича Российской академии наук по адресу: 127994, г. Москва, ГСП-4, Большой Каретный переулок, д. 19, стр. 1.

С текстом автореферата и диссертации можно ознакомиться в библиотеке Федерального государственного бюджетного учреждения науки Института проблем передачи информации им. А.А. Харкевича Российской академии наук, а также на сайте ИППИ РАН по адресу <http://www.iitp.ru/ru/dissertation/>\*

Автореферат разослан "\_\_" \_\_\_\_\_ 2015 г.

Ученый секретарь диссертационного совета

доктор биологических наук, профессор



Г.И. Рожкова

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность проблемы.** Обилие данных, как по наборам полиморфных мутаций, так и по полным геномам разных видов в совокупности с тем, что стали известны эпигенетические свойства ДНК, поставило ряд совершенно новых задач. Была открыта зависимость между временем репликации и скоростью мутирования участков ДНК, однако механизмы, определяющие эту зависимость, остались неизвестными. Появился новый пласт работ, фокусирующих на предсказании особенностей мутационных процессов в клетке по данным о наблюдаемых мутациях. Но, не смотря на прогресс в понимании процессов мутирования, очень малоизвестно об особенностях мутационных процессов в участках генома сильно подверженных мутированию (горячих точках мутагенеза). Так же недостаточно изучены сложные мутации – замены, затрагивающие одновременно несколько сайтов. С использованием данных, полученные из межвидовых сравнений и внутривидового полиморфизма, стало возможным изучать сложные мутации и редкие события, такие как мутации, многократно попавшие в один и тот же сайт ДНК. Более того, в результате экспериментов, стали известны механизмы возникновения некоторых сложных мутаций, что позволило изучать какие именно свойства ДНК, ассоциированы с этими механизмами.

Нами были исследованы горячие точки мутагенеза в хорошо аннотированных и качественно собранных геномах *Homo sapiens* и *Drosophila melanogaster*. Кроме того, используя межвидовые сравнения, мы обнаружили множество мутаций, затрагивающих сразу два нуклеотида – динуклеотидных мутаций (ДНМ). Особое внимание в наших исследованиях мы уделили GC→AA/TT ДНМ, являющимся результатом работы неточной полимеразы зета (пол ζ).

### **Цели и задачи исследования.**

Целью настоящего исследования была изучить особенности точечного и множественного мутирования у эукариот. В рамках поставленной цели решались следующие общие и частные задачи:

Собрать данные по многократно произошедшим в одном сайте (множественным) мутациям и данных по скорости мутирования в окрестности уже случившейся «условной» мутации.

Проанализировать соотношение разных типов мутаций в сайтах множественных мутаций и сравнить их со среднегеномными параметрами.

Рассмотреть сложные мутации по данным о межвидовой дивергенции и вычислить соотношение сложных и однонуклеотидных мутаций в *H. sapiens* и *D. melanogaster*.

Изучить распределение GC→AA/TT ДНМ, мутационной подписи полимераза  $\zeta$ , вдоль генома. Выяснить какие именно свойства ДНК влияют на частоту GC→AA/TT ДНМ. Исследовать роль пол  $\zeta$  в мутациях, приводящих к генетическим заболеваниям.

#### **Научная новизна и практическая значимость.**

Горячие точки мутагенеза и ДНМ плохо изучены, в том числе и из-за того, что ошибки выравнивания или сборки геномов могут выглядеть как кластеры мутаций. Тем не менее, наличие полных геномов для модельных организмов позволяет разрешить эту проблему. Наша работа сфокусирована на геномах с высококачественной сборкой: геномах *H. sapiens* и *D. melanogaster*, что и позволяет исследовать гипермутабельные участки и ДНМ.

В данной работе впервые изучено соотношение различных типов мутаций в горячих точках мутагенеза и расширены предыдущие работы по ДНМ. Исследованы особенности работы пол  $\zeta$  и показано, что её активность зависит от времени репликации и, вероятно, приводит к тяжёлым генетическим заболеваниям: синдрому Костелло и амиотрофическому латеральному склерозу.

Знание о процессе мутирования крайне полезно для изучения взаимодействия нескольких мутаций под действием отбора и помогает искать гены, связанные с генетическими заболеваниями или развитием рака.

**Публикации и апробация работы.** По материалам диссертации опубликовано 5 статей в международных рецензируемых научных журналах. Результаты работы были представлены на международных конференциях MCCMB'09,11 SMBE'14, THMRM'15 и российской конференциях ИТИС'09.

Апробация работы проведена 21 сентября 2015 г. на совместном семинаре лаборатории эволюционной геномики Факультета биоинженерии и биоинформатики МГУ им. М.В.Ломоносова и сектора молекулярной эволюции Института проблем передачи информации РАН им. А.А. Харкевича.

**Структура и объем работы.** Диссертация изложена на 93 страницах машинописного текста и содержит следующие разделы: введение, материалы и методы, результаты в четырех главах и выводы. Диссертация включает 26 рисунков, 6 таблиц и список литературы, содержащий 120 ссылок.

## СОДЕРЖАНИЕ РАБОТЫ

### 1) Изменчивость локальной скорости мутирования.

Изучена изменчивость локальной скорости мутирования на различных филогенетических расстояниях в двух группах metazoa: в роде *Drosophila* и семействе гоминид (Hominidae). Для этого мы анализировали, как меняется скорость мутирования у человека или *D. melanogaster* при наличии другой «условной» мутации. «Условная» мутация могла произойти как в виде, в котором мы измеряем скорость мутирования, так и на другой филогенетической ветке (Рисунок 1). Мы проделали анализы для четырех различных филогенетических тестов (Анализы I, II, III и IV) (Рисунок 1). Кроме того, мы исследовали, как меняется скорость мутирования в зависимости от нуклеотидного расстояния до «условной» мутации (Рисунки 2 и 3) или при наличии дополнительной «условной» мутации в том же сайте (Таблицы 1 и 2).

В анализе I изучалась частота однонуклеотидного полиморфизма (SNP), при условии другого SNP в том же виде поблизости от исследуемых сайтов (Рисунок 2А и 3А). В человеке мы обнаружили увеличение скорости мутирования лишь для сайта, непосредственно соседнего с «условным» SNP. В сайтах, в которых уже содержится SNP, вероятность другого SNP растёт в 2.6 раза в сравнении со средним по геному (Таблица 1).

У *D. melanogaster* наблюдалось повышение плотности SNP на расстоянии до 15 нуклеотидов от «условного» SNP (Рисунок 3А). Для ситуации, когда «условный» SNP наблюдается в изучаемом сайте, скорость мутирования повышена в 3.5 раза.

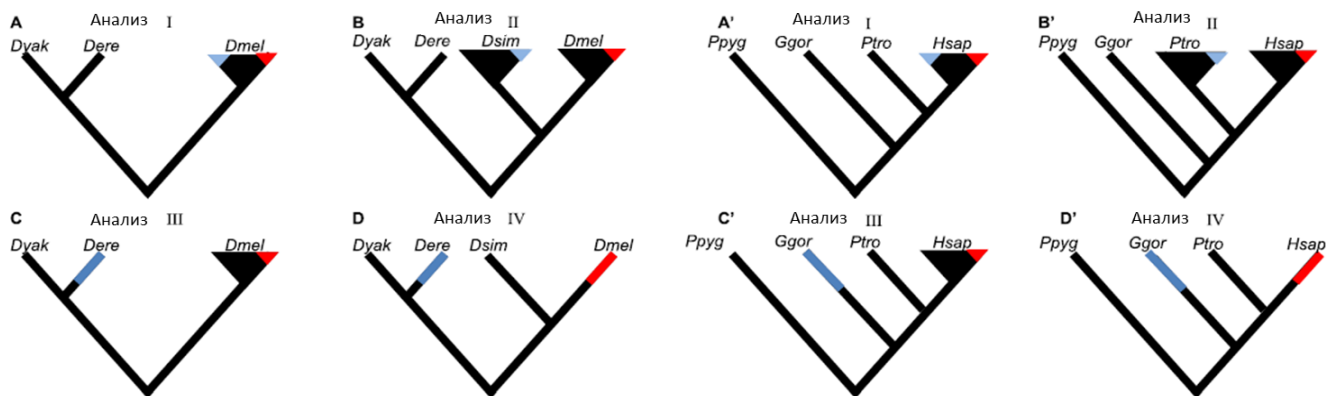
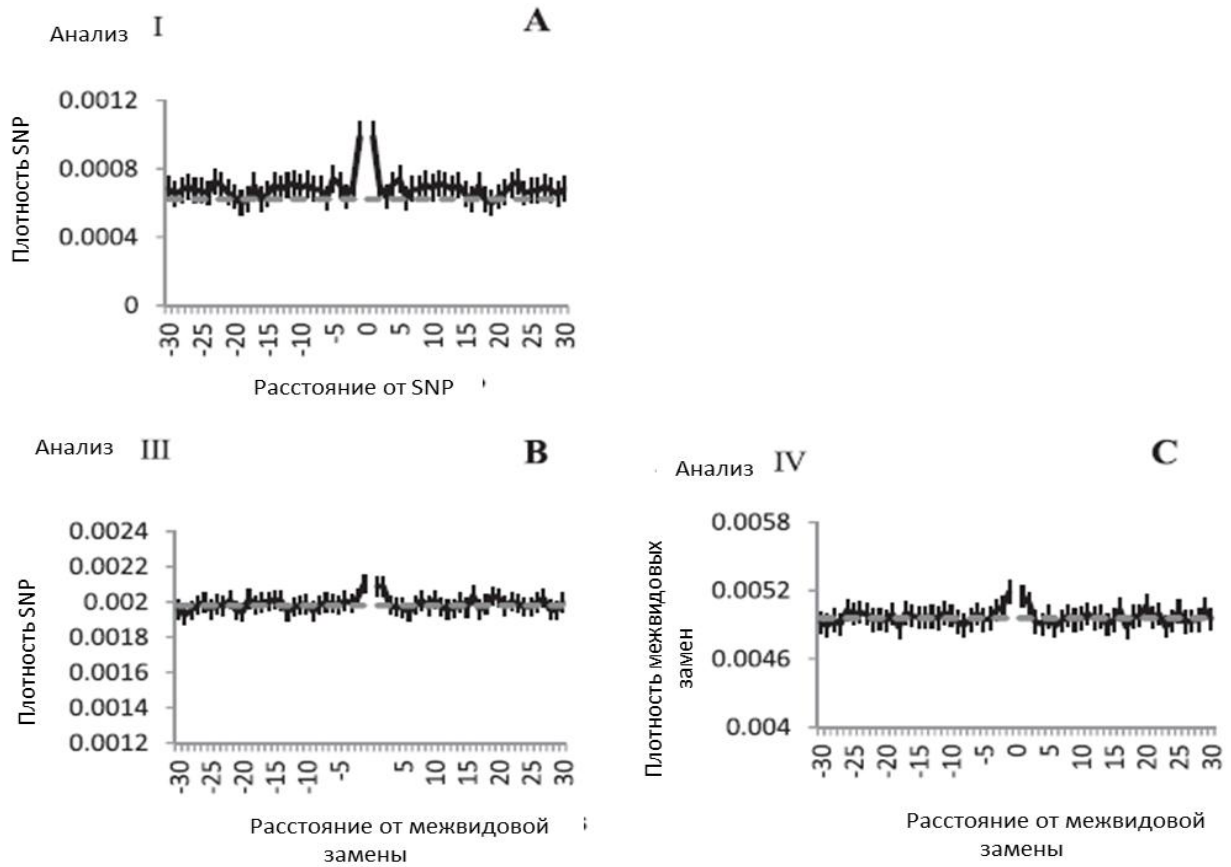
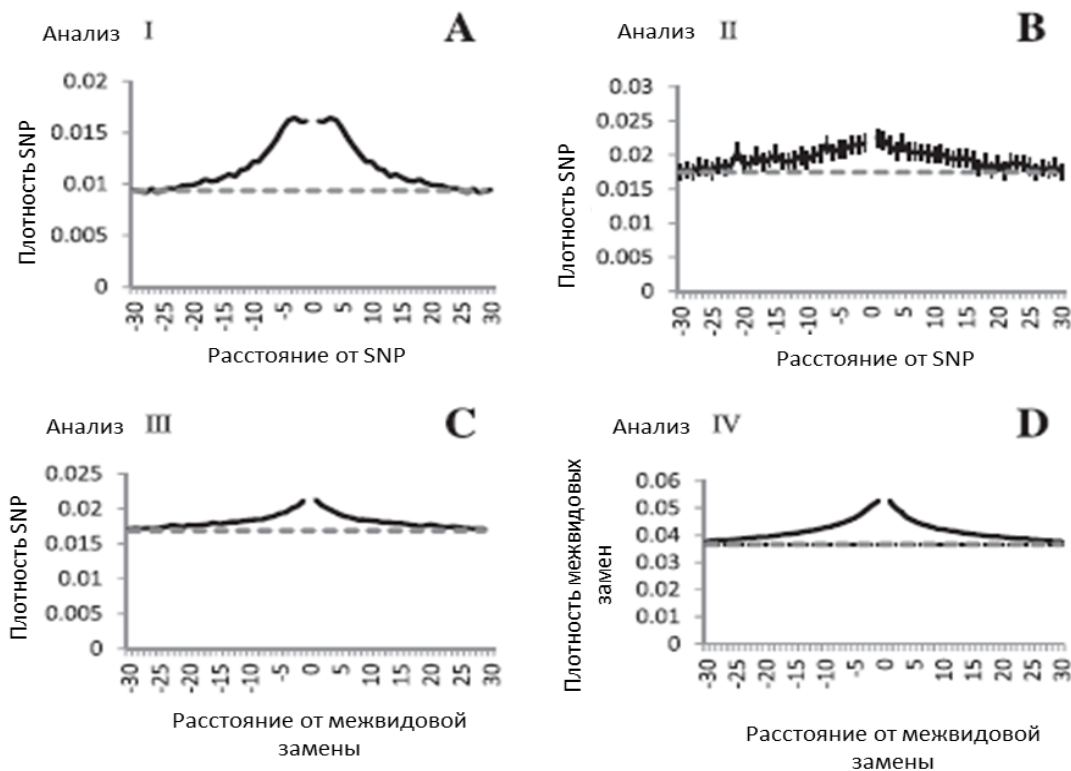


Рисунок 1. Четыре типа филогенетических анализов, сделанных для изучения неравномерности скорости мутирования вдоль генома. Линии обозначают филогенетические ветви, закрашенные треугольники – полиморфизм. Черным цветом обозначен предковый аллель, а синим и красным – различные производные аллели. В каждом анализе мы измеряли вероятность SNP или замены, такие события обозначены красным цветом, при условии события, обозначенного синим цветом. *Dyak*, *Dere*, *Dsim* и *Dmel* обозначают *D. yakuba*, *D. erecta*, *D. simulans* и *D. melanogaster*; *Ppyg*, *Ggor*, *Ptro*, и *Hsap* обозначают *Pongo pygmaeus*, *Gorilla gorilla*, *Pan troglodytes* и *Homo sapiens*.



**Рисунок 2. Плотность SNP или нуклеотидных замен как функция расстояния от другого SNP или нуклеотидной замены для межгенных интервалов у гоминид; анализы соответствуют анализам на рисунке 1. Положительные расстояния соответствуют 3` позициям, а отрицательные соответствуют 5` позициям. Чёрная линия обозначает плотность SNP (A, B) или межвидовых замен (C) поблизости от сайтов, содержащих «условную» мутацию; серая пунктирная линия показывает соответствующее значение, посчитанное по 100 нуклеотидным окнам с центром на изучаемом сайте. Ошибки посчитаны как 95% биномиальные доверительные интервалы. Анализ II для человека отсутствует, поскольку не было данных по полиморфизму шимпанзе.**



**Рисунок 3. Плотность SNP, или нуклеотидных замен как функция расстояния от «условного» SNP или нуклеотидной замены для межгенных интервалов у *Drosophila*; анализы соответствуют анализам на рисунке 1. Положительные расстояния соответствуют 3` позициям, а отрицательные соответствуют 5` позициям. Чёрная линия обозначает плотность SNP (A-C) или межвидовых замен (D) поблизости от сайтов, содержащих «условную» мутацию; серая пунктирная линия показывает соответствующее значение, посчитанное по 100 нуклеотидным окнам, с центром на изучаемом сайте. Ошибки посчитаны как 95% биномиальные доверительные интервалы; мы не рисовали доверительные интервалы, когда они были малы и плохо различимы.**



Для анализов II-IV, соответствующим различным филогенетическим расстояниям между «условной» заменой и изучаемым сайтом (Рисунок 1), мы наблюдали те же паттерны, что и для анализа I: в человеке скорость мутирования повышена на расстоянии в 1-2 нуклеотида от «условной» мутации (Рисунок 2); в *D. melanogaster* – на расстоянии 10-15 нуклеотидов. Скорость мутирования значительно повышается в сайте, в котором произошла «условная» мутация, как в человеке, так и в плодовой мушке (Таблицы 1 и 2).

**Таблица 1. Число и частота (в скобках) случаев одной и двух мутаций в сайте для трёх из четырёх типов анализов для Homenidae. Ожидание во всех случаях достоверно отличается от наблюдения ( $P < 10^{-100}$ ,  $\chi^2$ ).**

	Тип участка	Одна мутация в сайте	Две мутации в сайте, наблюдение	Две мутации в сайте, ожидание	Наблюдение /ожидание
<b>Триаллельный SNP (анализ I)</b>	Интроны	981033 ( $5.82 \cdot 10^{-4}$ )	819 ( $4.92 \cdot 10^{-7}$ )	442 ( $2.73 \cdot 10^{-7}$ )	2.55
	Межгенные интервалы	1321296 ( $6.93 \cdot 10^{-4}$ )	1113 ( $5.88 \cdot 10^{-7}$ )	703 ( $3.97 \cdot 10^{-7}$ )	2.21
<b>SNP в сайтах с заменой (анализ III)</b>	Интроны	575893 ( $3.84 \cdot 10^{-4}$ )	2403 ( $8.13 \cdot 10^{-7}$ )	1442 ( $5.03 \cdot 10^{-7}$ )	1.83
	Межгенные интервалы	720188 ( $4.28 \cdot 10^{-4}$ )	3139 ( $9.66 \cdot 10^{-7}$ )	1,854 ( $5.86 \cdot 10^{-7}$ )	1.85
<b>Замена в сайтах с другой заменой (анализ IV)</b>	Интроны	1582163 ( $1.52 \cdot 10^{-3}$ )	8159 ( $3.98 \cdot 10^{-6}$ )	5,591 ( $2.83 \cdot 10^{-6}$ )	1.76
	Межгенные интервалы	1873583 ( $1.67 \cdot 10^{-3}$ )	1880 ( $4.90 \cdot 10^{-6}$ )	7,239 ( $3.41 \cdot 10^{-6}$ )	1.83

**Таблица 2. Число и частота (в скобках) случаев одной и двух мутаций в сайте для четырёх типов анализов для *Drosophila*. Ожидание во всех случаях достоверно отличается от наблюдения ( $P < 10^{-100}$ ,  $\chi^2$ ).**

	Тип участка	Одна мутация в сайте	Две мутации в сайте, наблюдение	Две мутации в сайте, ожидание	Наблюдение /ожидание
<b>Триаллельный SNP (анализ I)</b>	Интроны	392688 ( $6.91 \cdot 10^{-3}$ )	7315 ( $1.24 \cdot 10^{-4}$ )	2525 ( $4.12 \cdot 10^{-5}$ )	3.48
	Межгенные интервалы	523,130 ( $5.72 \cdot 10^{-3}$ )	9725 ( $1.00 \cdot 10^{-4}$ )	3268 ( $3.23 \cdot 10^{-5}$ )	3.59
<b>Совпадающий SNP (анализ II)</b>	Интроны	392688 ( $6.91 \cdot 10^{-3}$ )	2743 ( $4.59 \cdot 10^{-5}$ )	1,223 ( $2.07 \cdot 10^{-5}$ )	2.28
	Межгенные интервалы	523130 ( $5.27 \cdot 10^{-3}$ )	3486 ( $3.59 \cdot 10^{-5}$ )	1502 ( $1.65 \cdot 10^{-6}$ )	2.35
<b>SNP в сайтах с заменой (анализ III)</b>	Интроны	392688 ( $6.91 \cdot 10^{-3}$ )	39398 ( $3.14 \cdot 10^{-4}$ )	20291 ( $1.76 \cdot 10^{-4}$ )	1.95
	Межгенные интервалы	523130 ( $5.72 \cdot 10^{-3}$ )	53533 ( $3.34 \cdot 10^{-4}$ )	25953 ( $1.63 \cdot 10^{-4}$ )	2.07
<b>Замена в сайтах с другой заменой (анализ IV)</b>	Интроны	766220 ( $9.26 \cdot 10^{-3}$ )	80597 ( $5.02 \cdot 10^{-4}$ )	37319 ( $2.32 \cdot 10^{-4}$ )	2.23
	Межгенные интервалы	914971 ( $9.73 \cdot 10^{-3}$ )	95564 ( $5.02 \cdot 10^{-4}$ )	47947 ( $2.51 \cdot 10^{-4}$ )	2.06

## 2) Локальная изменчивость отношения транзиций к трансверсиям.

Однонуклеотидные мутации, меняющие тип азотистого основания, называются трансверсиями. Мутации, не меняющие тип азотистого основания – транзициями. Отношение транзиций к трансверсиям, нормированное на количество классов таких мутаций, обозначается символом **к**. Если рассматривать только различные производные варианты, то для определённого предкового нуклеотида, возможны две транзиции и одна трансверсия. Для изучения **к** в сайтах множественных мутаций, мы исследовали отношение транзиций к трансверсиям при «условии» трансверсии ( $k_{tv}$ ), а отношение этих же типов мутаций для всего генома мы обозначили  $k_{tv}$  (см. Таблицы 3 и 4).

**Таблица 3. Отношение частот транзиций к трансверсиям в биаллельных и триаллельных SNP (анализ I) у Hominidae. P < 0.001 для всех сравнений  $\kappa_{tv}$  и  $\kappa_{tvc}$ .**

Транзиция и трансверсия, использованные для расчета $\kappa_{tv}$	«Условная» трансверсия, использованная для расчёта $\kappa_{tvc}$	Интроны			Межгенные интервалы		
		$\kappa_{tv}$	$\kappa_{tvc}$	$\kappa_{tvc}/\kappa_{tv}$	$\kappa_{tv}$	$\kappa_{tvc}$	$\kappa_{tvc}/\kappa_{tv}$
A→G/A→T	A→C	4.76	1.74	0.37	4.35	1.51	0.35
A→G/A→C	A→T	3.91	1.56	0.40	3.84	1.30	0.34
G→A/G→T	G→C	3.48	1.18	0.34	3.20	1.34	0.42
G→A/G→C	G→C	3.01	0.57	0.19	3.31	1.17	0.35
C→T/C→G	C→A	3.01	0.93	0.31	3.36	1.04	0.31
C→T/C→A	C→G	3.58	1.34	0.37	3.22	1.09	0.34
T→C/T→G	T→A	3.82	1.08	0.28	3.89	1.12	0.29
T→C/T→A	T→G	4.87	1.00	0.21	4.34	1.38	0.32
Среднее значение		3.82	1.18	0.31	3.69	1.24	0.34

**Таблица 4. Отношение частот транзиций к трансверсиям в биаллельных и триаллельных SNP (анализ I) у *Drosophila*. \*,\*\*,\*\*\* обозначают  $P<0.05$ ,  $P<0.01$  и  $P<0.001$  соответственно для отличия  $K_{tvc}$  и  $K_{tv}$ .**

Транзиция и трансверсия, использованные для расчёта $K_{tv}$	«Условная» трансверсия, использованная для расчёта $K_{tvc}$	Интроны			Межгенные интервалы		
		$K_{tv}$	$K_{tvc}$	$K_{tvc}/K_{tv}$	$K_{tv}$	$K_{tvc}$	$K_{tvc}/K_{tv}$
A→G/A→T	A→C	1.02	0.78	0.76***	1.03	0.82	0.80**
A→G/A→C	A→T	1.88	1.76	0.94	1.88	1.70	0.90*
G→A/G→T	G→C	1.84	1.45	0.79***	1.83	1.26	0.69***
G→A/G→C	G→C	3.52	2.16	0.61***	3.30	2.02	0.61***
C→T/C→G	C→A	3.50	2.17	0.62***	3.33	2.14	0.64***
C→T/C→A	C→G	1.85	1.26	0.68***	1.81	1.30	0.72***
T→C/T→G	T→A	1.85	1.50	0.81***	1.85	1.61	0.87***
T→C/T→A	T→G	1.01	0.81	0.81**	1.02	0.84	0.82**
Среднее значение		2.06	1.49	0.75	2.01	1.46	0.76

Для всех филогенетических анализов, в которых «условная» мутация случилась в изучаемом сайте, обнаружено, что значение  $K$  снижается относительно среднегеномных значений. Как у *Hominidae*, так и у *Drosophila* наиболее выражено падение  $K$  в анализе I, то есть для триаллельных полиморфизмов (Таблицы 3 и 4). В *Hominidae* среднее отношение  $K_{tvc}$  к  $K_{tv}$  составляет 0.31, 0.91 и 0.58 для анализов I, III и IV, соответственно. У *Drosophila* эти значения для анализов I–IV – 0.76, 0.91, 0.95 и 0.86, соответственно.

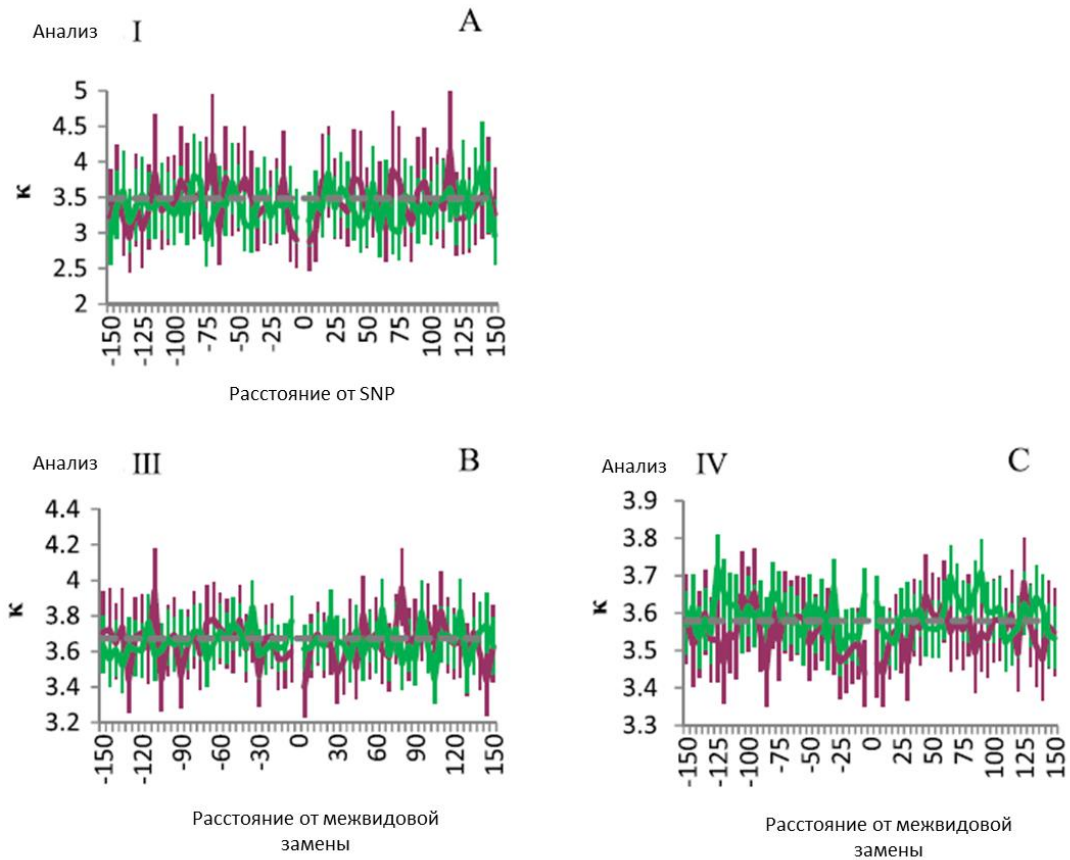
У *S. commune* в анализе II обнаружено, что вероятность SNP в одной популяции повышается при «условии» SNP в другой популяции в 1.42 раза. Этот эффект сильнее, когда оба биаллельных SNP – трансверсии (1.58), и слабее, когда один биаллельный SNP

транзигия, а другой – трансверсия (1.32). Таким образом, и у *S. commune*, трансверсии сильнее скоррелированы вдоль генома.

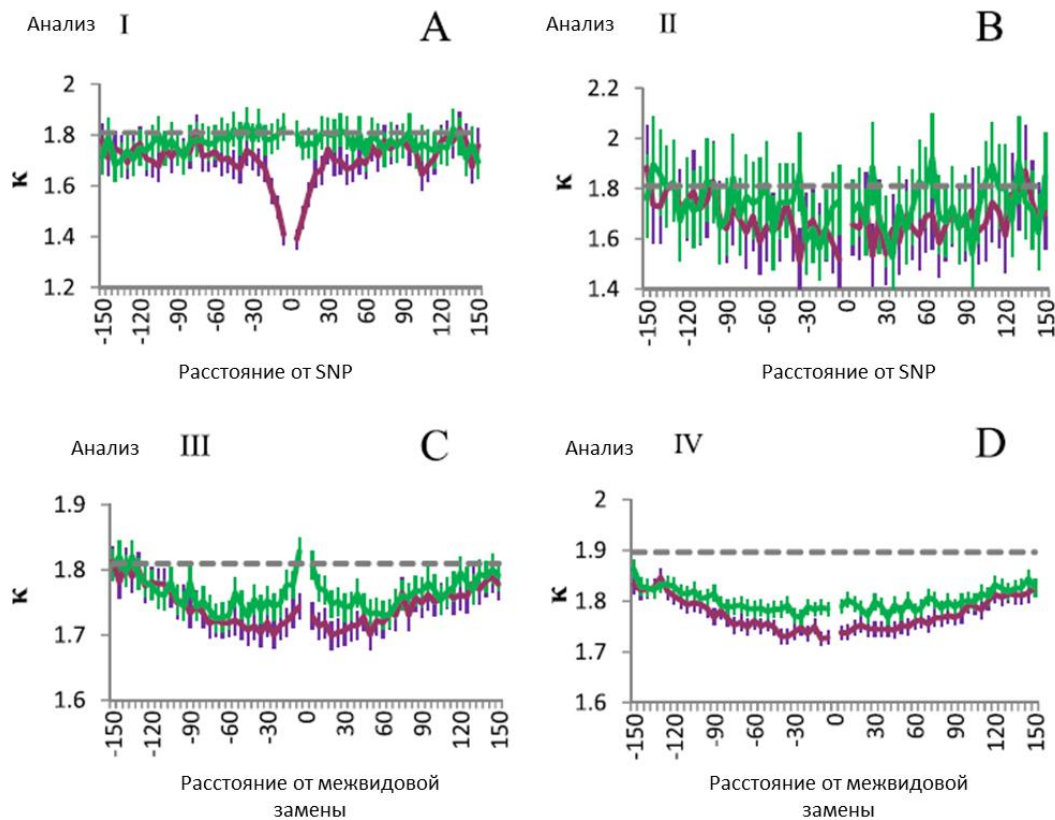
Также исследовалось, как меняется значение  $k$  для различных нуклеотидных расстояний от «условной» мутации, которая может быть транзигией или трансверсией.

У человека во всех анализах  $k$  снижена для 5 нуклеотидов вблизи от мутации, случившейся как трансверсия, но не транзигия; этого эффекта не наблюдается на расстоянии более 5 нуклеотидов (Рисунок 4).

У *D. melanogaster*, значение  $k$  понижено в окрестности трансверсии сильнее, чем транзигии (Рисунок 5), что свидетельствует о корреляции трансверсий вдоль генома. В анализе IV  $k$  так же понижена вблизи транзигий. В анализе III мы видим немонотонность  $k$  по отношению к положению «условной» мутации, что, возможно, объясняется присутствием нескольких мутационных эффектов с различными характерными расстояниями.



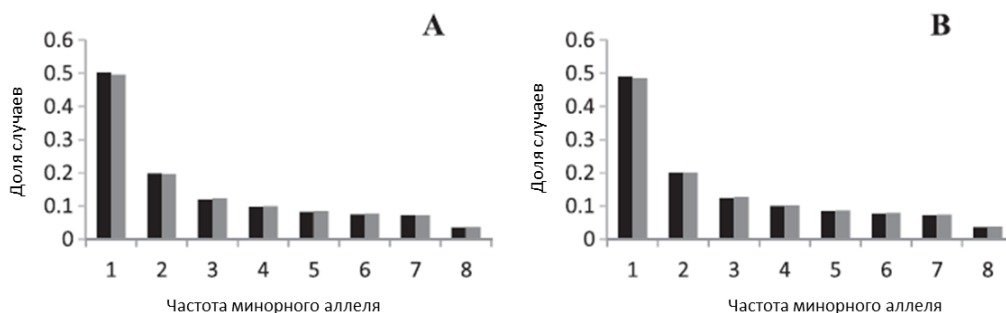
**Рисунок 4. Отношение транзиций к трансверсиям ( $\kappa$ ) как функция расстояния до транзиции (зеленый) или трансверсии (фиолетовый) для межгенных интервалов у *Hominidae*; анализы соответствуют анализам на рисунке 1. Положительные расстояния соответствуют 3` позициям рассматриваемых нуклеотидов, а отрицательные соответствуют 5` позициям рассматриваемых нуклеотидов. Серая пунктирная линия соответствует среднегеномному значению  $\kappa$ . Значения  $\kappa$  даны по 5 нуклеотидным окнам. Ошибки посчитаны как 95% доверительные интервалы по 1000 испытаний.**



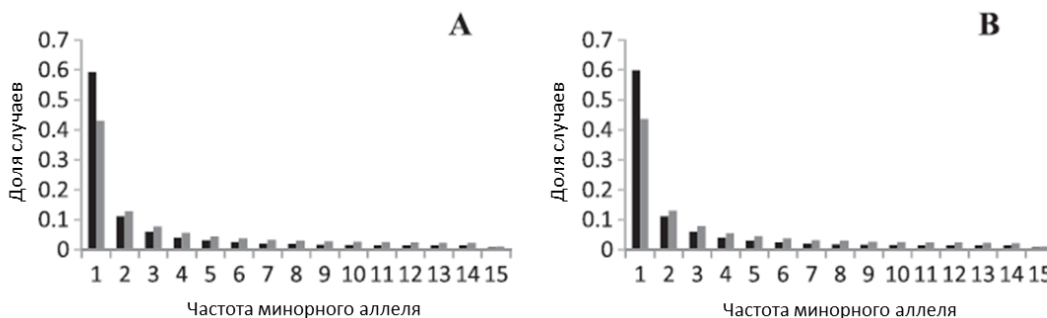
**Рисунок 5. Отношение транзиций к трансверсиям ( $\kappa$ ) как функция расстояния до транзиции (зеленый) или трансверсии (фиолетовый) для межгенных интервалов у *Drosophila*; анализы соответствуют анализам на рисунке 1. Положительные расстояния соответствуют 3' позициям рассматриваемых нуклеотидов, а отрицательные соответствуют 5' позициям рассматриваемых нуклеотидов. Серая пунктирная линия соответствует среднегеномному значению  $\kappa$ . Значения  $\kappa$  даны по 5 нуклеотидным окнам. Ошибки посчитаны как 95% доверительные интервалы по 1000 испытаний.**

Сильную скоррелированность трансверсий вдоль генома можно объяснять не только особенностями мутационных механизмов, но и влиянием отбора. Так, если трансверсии более вредны, чем транзиции, то распределение транзиций вдоль генома будет в меньшей степени зависеть от отрицательного отбора. Мы построили спектр аллельных частот для трансверсий и транзиций в человеке и *D. melanogaster* и

установили, что спектры частот сходны для транзиций и трансверсий в человеке (Рисунок 6). Одинаковое распределение частот для транзиций и трансверсий, свидетельствует о сходном давлении отбора на эти типы мутаций. Напротив, в *D. melanogaster* спектр аллельных частот для трансверсий смещён в сторону редких вариантов (Рисунок 7), что говорит о более сильном отборе против трансверсий, и, вероятно, частично объясняет более тесную ассоциацию трансверсий по геному. Тем не менее, самое сильное снижение  $K$  в сайтах, содержащих трансверсию, мы наблюдали именно для человека, что согласуется с тем, что мутационные процессы играют главенствующую роль в повышении доли трансверсий в горячих точках мутагенеза.



**Рисунок 6.** Спектр частот минорного аллеля для полиморфных трансверсий (черный) и транзиций (серый) у человека в интронах (А) и межгенных интервалах (В). Горизонтальная ось соответствует количеству генотипов (из 16), несущих минорный аллель.



**Рисунок 7.** Спектр частот минорного аллеля для полиморфных трансверсий (черный) и транзиций (серый) у *D. melanogaster* в интронах (А) и межгенных



интервалах (В). Горизонтальная ось соответствует количеству генотипов (из 30), несущих минорный аллель.

### 3) Мультинуклеотидные замены в эволюции приматов и *Drosophila*.

Замены могут кластеризоваться не только из-за того, что скорость мутирования отличается вдоль генома, но и из-за того, что многие мутаций происходят, как события, затрагивающие сразу несколько нуклеотидов.  $\alpha_d(k)$  – отношение частоты динуклеотидных мутаций (ДНМ), затрагивающих два нуклеотида на расстоянии  $k$  друг от друга, к числу однонуклеотидных мутаций. Рассчитать  $\alpha_d(k)$  можно как разность частоты пар мутаций, случившихся в разных видах ( $d_d(k)$ ), и частоты пар мутаций, случившихся только в одном виде ( $s_d(k)$ ), нормированную на частоту однонуклеотидных мутаций. Этот подход работает благодаря тому, что пары мутаций, случившиеся на одной филогенетической ветви, могут происходить как ДНМ, или как пара одиночных мутаций, а пара мутаций на разных ветвях случается всегда как два независимых события.

В человеке сумма значений  $\alpha_d(k)$  для  $1 \leq k \leq 10$  составляет 0.023 и наибольшее значение  $\alpha_d(k)$  достигает при  $k=1$  (Рисунок 8). Мы обнаружили, что величина  $\alpha_d(1)$  очень сходна для всех изучаемых троек видов из семейства *Hominidae*, несмотря на то, что длина веток отличается до 3.8 раз, и отношение числа ДНМ к числу всех пар мутаций произошедших на одной ветке падает с филогенетическим расстоянием (Рисунок 9).

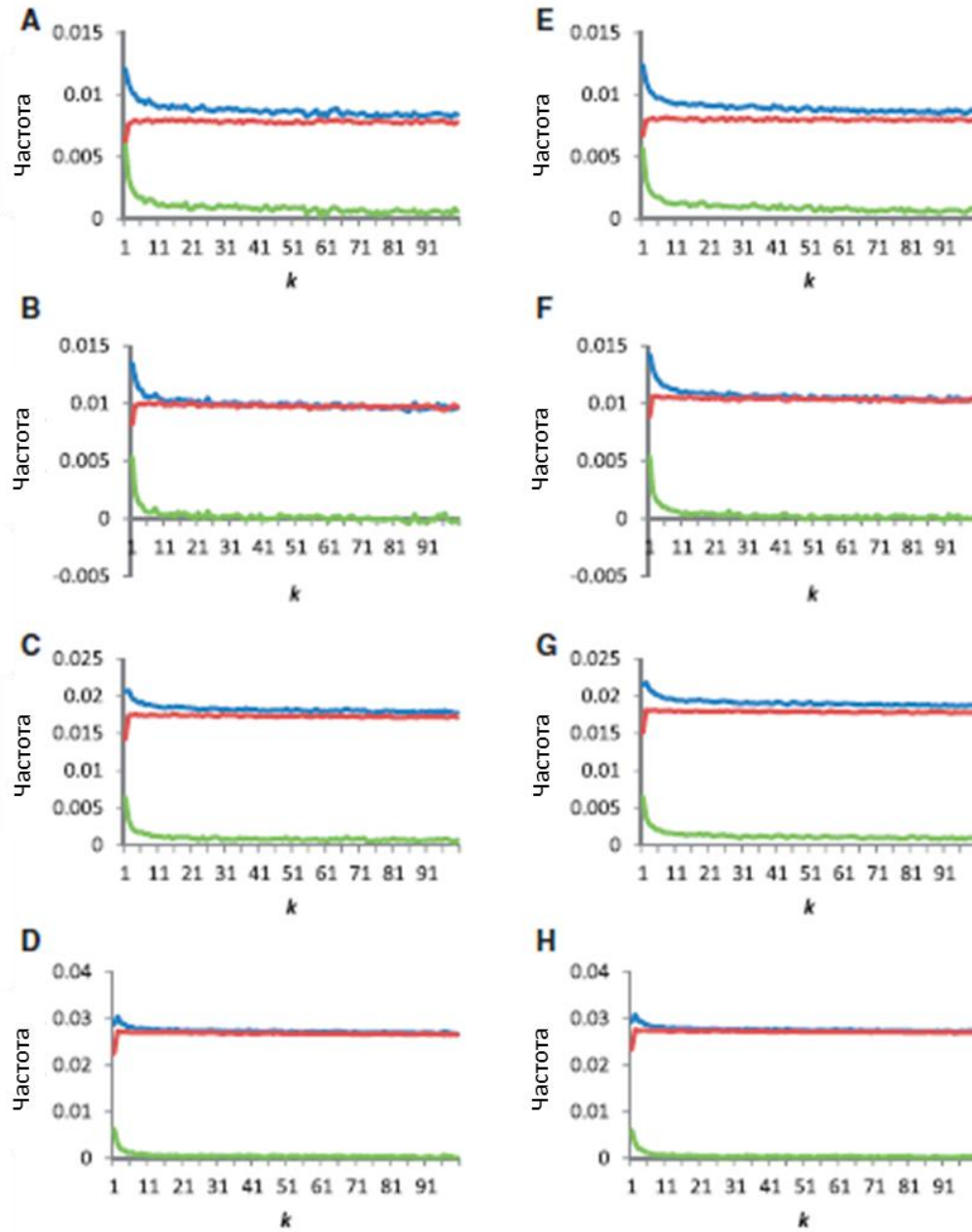


Рисунок 8. Частота ДНМ в приматах для сайтов на разных расстояниях.  $d_d(k)$  (красный),  $s_d(k)$  (синий) и  $\alpha_d(k)$  (зеленый) для расстояний между сайтами  $1 \leq k \leq 100$  (горизонтальная ось). Левая колонка (A-D), интроны; правая колонка (E-H), межгенные интервалы. *Homo sapiens* и *Pan troglodytes* (*Gorilla gorilla* в качестве внешнего вида), (B,F) *H. sapiens* и *G. gorilla* (*Pongo pygmaeus* в качестве внешнего вида), (C,G) *H. sapiens* и *P. pygmaeus* (*Macaca mulatta* в качестве внешнего вида), and (D,H) *H. sapiens* и *M. mulatta* (*Callithrix jacchus* в качестве внешнего вида). Исключение CpG динуклеотидов приводит к симметричной недооценке  $d_d(1)$  и  $s_d(1)$ .

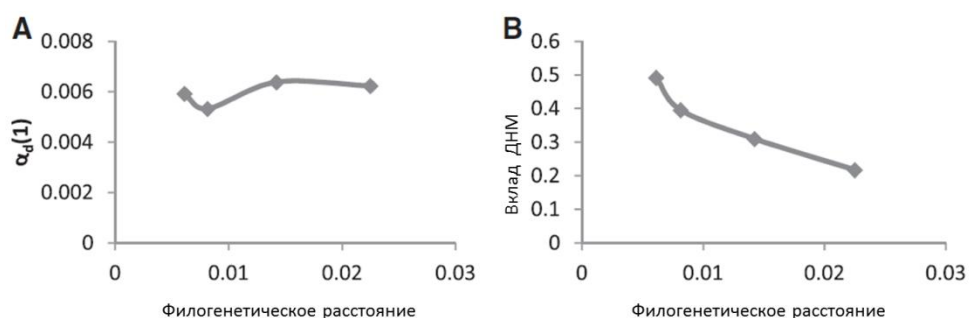
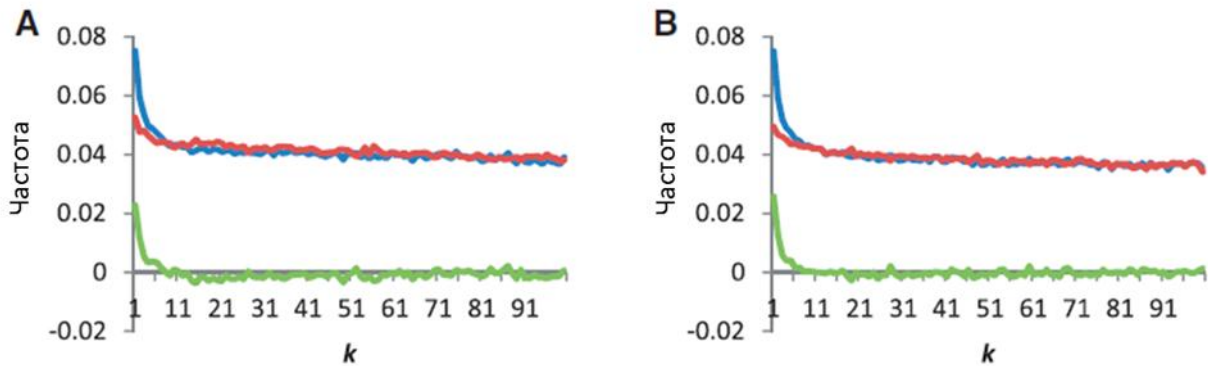


Рисунок 9. Зависимость вклада ДНМ от длины филогенетической линии в интронах. (A)  $\alpha_d(1)$  как функция  $d_d(k)$  и (B)  $\alpha_d(1)/s_d(k)$  как функция  $d_d(k)$ . На обоих графиках четыре точки соответствуют заменам, случившимся на линии *H. sapiens* после отхождения от общего предка с *P. troglodytes*, *G. gorilla*, *P. pygmaeus* и *M. mulatta* соответственно.

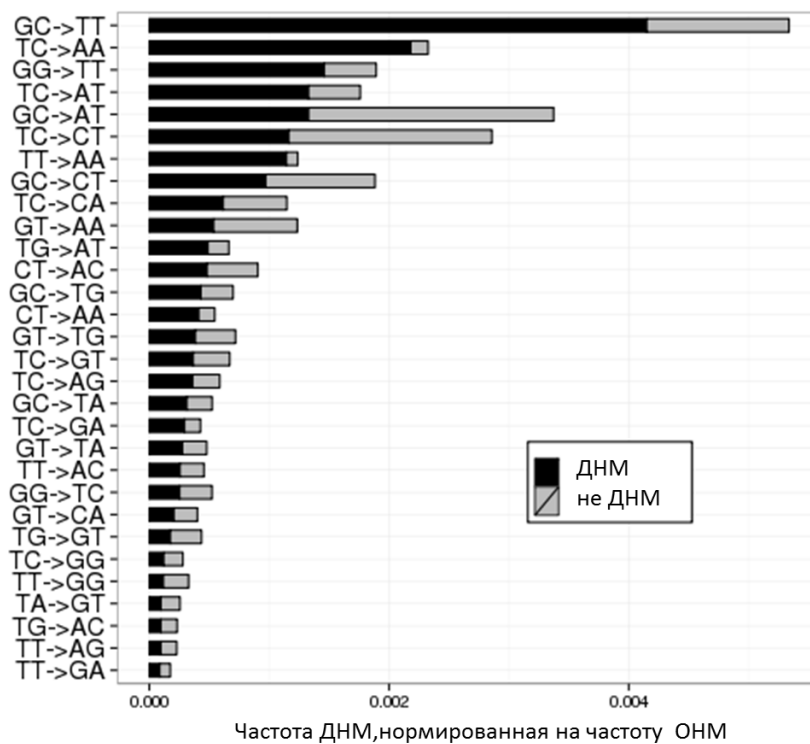
В *D. melanogaster* сумма значений  $\alpha_d(k)$  для  $1 \leq k \leq 10$  составляет 0.056 и наибольшее значение  $\alpha_d(k)$  достигает при  $k=1$  (Рисунок 10). Подытоживая, мы показали, что ДНМ, затрагивающие пары сайтов на расстоянии до 10 нуклеотидов, составляют 2.3% и 5.6% от однонуклеотидных замен в Hominidae и *Drosophila* соответственно. Эти оценки учитывают локальную неравномерность скорости мутирования, если эта неравномерность сохраняется между близкими видами, использованными в анализах.



**Рисунок 10. Частота ДНМ у *Drosophila* для сайтов на разных расстояниях.  $d_a(k)$ (красным),  $s_a(k)$  (синим) и  $a_a(k)$  (зеленым) для расстояний между сайтами  $1 \leq k \leq 100$  (горизонтальная ось). (А) интроны; (В) межгенные интервалы.**

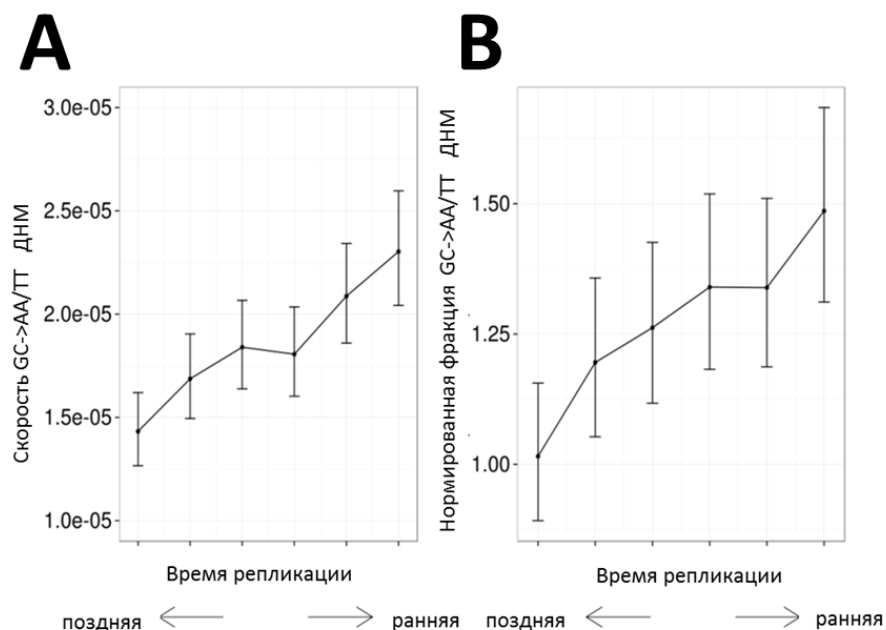
#### **4) Использование динуклеотидной мутационной подписи для изучения свойств полимеразы зета.**

В экспериментах на дрожжах и на мыши было показано, что неточная пол  $\zeta$ , вызывает GC→AA/TT ДНМ. Используя межвидовые сравнения, мы исследовали частоты различных ДНМ и обнаружили, что GC→AA/TT ДНМ является самым часто встречающимся типом ДНМ (Рисунок 11), что свидетельствует о важной роли пол  $\zeta$  в мутагенезе человека.



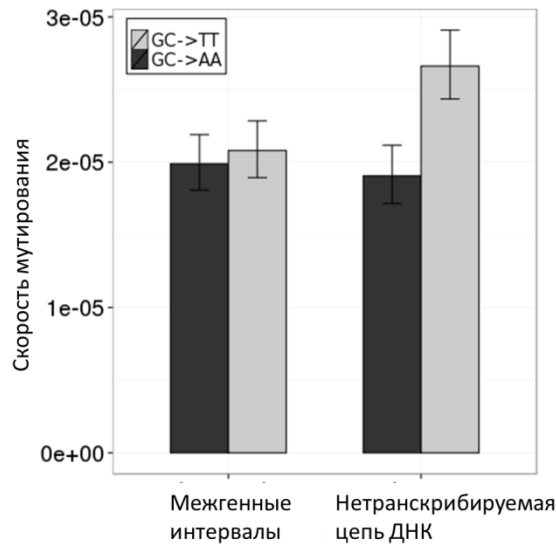
**Рисунок 11. 30 наиболее частых ДНМ. Мы объединяли пары комплементарных мутаций. Частота ДНМ (черным) и тандемных мутаций, произошедших как два независимых события (серым). GC→AA/TT ДНМ является наиболее частой среди всех классов ДНМ.**

Используя GC→AA/TT ДНМ в качестве мутационной подписи пол ζ, мы решили исследовать гипотезу о том, что неточные полимеразы особенно активно принимают участие в удвоении ДНК с поздним временем репликации. Частота GC→AA/TT ДНМ растёт от рано реплицирующихся регионов ДНК к поздно реплицирующимся. Мы показали, что зависимость между временем репликации и частотой GC→AA/TT ДНМ, сильнее чем для ДНМ в целом, и доля GC→AA/TT ДНМ среди всех ДНМ растёт в поздно реплицирующихся регионах (Рисунок 12).

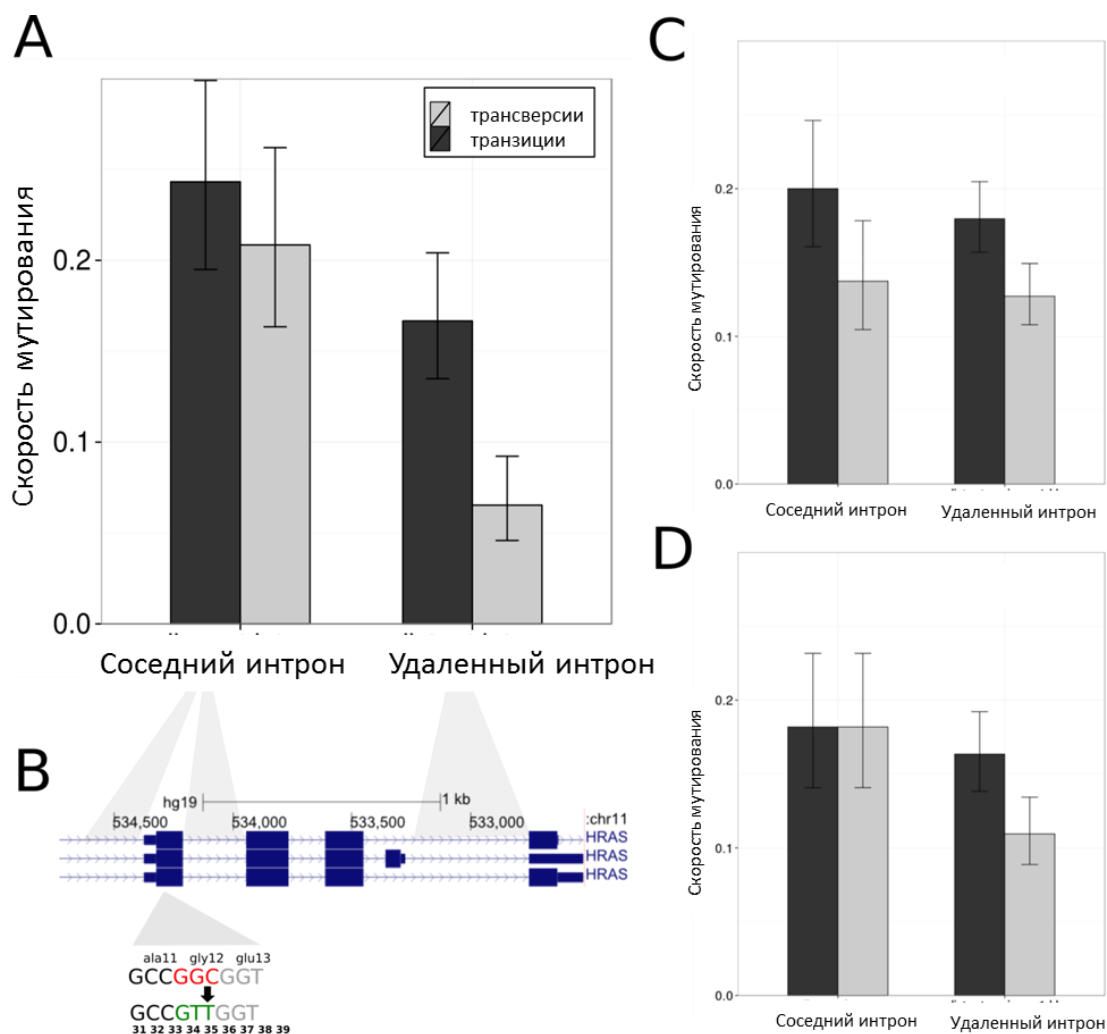


**Рисунок 12. Связь между скоростью GC→AA/TT ДНМ и временем репликации. (А) Скорость GC→AA/TT ДНМ растет от раннего к позднему времени репликации, как и доля GC→AA/TT ДНМ среди всех ДНМ (В).**

Частота GC→AA/TT ДНМ выше в интронах, чем в межгенных интервалах, и, кроме того, мы обнаружили, что на нетранскрибируемой цепи частота GC→ТТ ДНМ превосходит частоту комплиментарной к ней GC→AA ДНМ на 40% (Рисунок 13). Вероятно, при удвоении ДНК повреждения, накопившиеся из-за транскрипции, вызывают остановку основных репликативных полимераз, и пол  $\zeta$  участвует в прохождении этих повреждений.



**Рисунок 13. Связанная с транскрипцией асимметрия между GC→AA и GC→TT ДНМ. Для межгенных интервалов частоты GC→AA и GC→TT ДНМ посчитаны для референтной цепи.** Известны случаи, многократно повторяющихся GC→AA/TT мутаций. Эти мутации в генах HRAS и SOD1 вызывают синдром Костелло и боковой амиотрофический склероз, соответственно. Экспериментально показано, что пол ζ вставляет гораздо больше одиночных мутаций, чем основные репликационные полимеразы, её спектр мутаций обогащён трансверсиями, и, кроме того, она реплицирует короткие участки ДНК (меньше 1000 нуклеотидов). Мы обнаружили, что несколько сотен нуклеотидов по соседству с сайтом, в котором многократно наблюдали GC→AA/TT ДНМ, характеризуются повышенной скоростью мутирования, повышенной долей трансверсии, и этот эффект пропадает для участков удалённых больше, чем на 1000 нуклеотидов (Рисунок 14). Это косвенное свидетельство того, что пол ζ – причина наблюдаемых мутаций и её активность в генах HRAS и SOD1, вызывает тяжёлые заболевания.



**Рисунок 14. GC→TT/AA ДНМ в генах HRAS и SOD1, ассоциированных с синдромом Костелло и амиотрофическим латеральным склерозом, соответственно. (А) ген HRAS. Интроны, соседствующие с горячей точкой для GC→TT ДНМ, мутируют значительно быстрее и имеют пониженную к, по сравнению с чуть более удаленными интронами. (В) Структура гена HRAS. GC→TT ДНМ происходит в двенадцатом глицине первого экзона. (С) Ген KRAS – контроль для гена HRAS. (D) ген SOD1.**



## **Выводы**

- 1) SNP попадает в сайт, содержащий другой SNP, в 3.5, 2.5 и 1.4 раза чаще у *D. melanogaster*, *H. sapiens* и *S. commune*, соответственно, что говорит о сильной гетерогенности скорости мутирования на уровне однонуклеотидных позиций.
- 2) Соотношение транзиций и трансверсий смещается в сторону трансверсий в сайтах поблизости от трансверсий и в сайтах, содержащих трансверсию; так, в человеке доля трансверсий растёт втрое для сайтов, содержащих другую трансверсию.
- 3) Многонуклеотидные мутации – распространённый феномен у *Metazoa*, и доля мутаций, затрагивающих 2 сайта на расстоянии до 10 нуклеотидов составляет 5.6% и 2.3% от однонуклеотидных мутаций у *D. melanogaster* и *H. sapiens*.
- 4) Половина тандемных замен в линии человека после отделения от шимпанзе происходят как двойные мутации.
- 5) Активность полимеразы зета связана со временем репликации.
- 6) Полимераза зета в 1.4 раза интенсивнее работает на нетранскрибируемой цепи.
- 7) Вероятно, полимеразы зета делают мутации, приводящие к синдрому Костелло и амиотрофическому латеральному склерозу.

## **Список публикаций по теме диссертации**

### **Статьи в изданиях, рекомендованных перечнем ВАК**

1. **Seplyarskiy VB**, Kharchenko P, Kondrashov AS, Bazykin GA. Heterogeneity of the transition/transversion ratio in *Drosophila* and Hominidae genomes. *Mol Biol Evol.* 2012 – V 29(8) – PP. 1943-1955.

2. Terekhanova NV, Bazykin GA, Neverov A, Kondrashov AS, **Seplyarskiy VB**.  
Prevalence of multinucleotide replacements in evolution of primates and *Drosophila*. *Mol Biol Evol*. 2013 – V 30(6) – PP.1315-1325.
3. **Seplyarskiy VB**, Logacheva MD, Penin AA, Baranova MA, Leushkin EV, Demidenko NV, Klepikova AV, Kondrashov FA, Kondrashov AS, James TY. Crossing-over in a hypervariable species preferentially occurs in regions of high local similarity. *Mol Biol Evol*. 2014 - V 31(11) - PP.3016-3025.
4. Baranova MA, Logacheva MD, Penin AA, **Seplyarskiy VB**, Safonova YY, Naumenko SA, Klepikova AV, Gerasimov ES, Bazykin GA, James TY, Kondrashov AS.  
Extraordinary Genetic Diversity in a Wood Decay Mushroom. *Mol Biol Evol*. 2015 – V 32(10) – PP. 2775-2783.
5. **Seplyarskiy VB**, Bazykin GA, Soldatov RA. Polymerase  $\zeta$  activity is linked to replication timing in humans: evidence from mutational signatures. *Mol Biol Evol*. 2015 Sep 15.

#### **Труды конференций**

1. **Seplyarskiy VB**, Bazykin GA. Polyallelic SNPs in population of *Drosophila melanogaster* // MCCMB 2009: Proceedings of the International Moscow Conference on Computational Molecular Biology. - Moscow, Russia, 2009. - P. 321
2. **Seplyarskiy VB**, Bazykin GA. Higher cryptic variation in rate of transversions than transitions. // ITaS 2009: Information Technology and Systems – 2009. - Bekasovo, Russia, 2009. - P. 382
3. **Seplyarskiy VB**, Kondrashov AS, Bazykin GA. Small scale heterogeneity in mutation rate and mutation biases in *Drosophila*. // MCCMB 2011: Proceedings of the International Moscow Conference on Computational Molecular Biology. - Moscow, Russia, 2011. - P. 340
4. **Seplyarskiy VB**, Logacheva MD, Penin AA, Baranova MA, Leushkin EV, Demidenko NV, Klepikova AV, Kondrashov FA, Kondrashov AS, James TY. // Negative selection

facilitates crossing-over in a hypervariable species. Annual Meeting of the Society for Molecular Biology and Evolution. – Puerto Rico, USA, 2014

5. **Septyarskiy VB**, Bazykin GA, Soldatov RA. Polymerase  $\zeta$  activity is linked to replication timing in humans: evidence from mutational signatures. // The Human Mutation Rate Meeting 2015 – Leipzig, Germany 2015 – P. 35.