

*На правах рукописи*



Вахрушева Ольга Александровна

**Эволюционно-генетический сигнал отрицательного отбора и  
рекомбинации в полногеномных данных**

1.5.8. – математическая биология, биоинформатика

**АВТОРЕФЕРАТ**

диссертации на соискание ученой степени  
кандидата биологических наук

Москва – 2022

Работа выполнена в Федеральном государственном бюджетном учреждении науки Институте проблем передачи информации им. А.А. Харкевича Российской академии наук (ИППИ РАН).

**Научный руководитель:**

**Базыкин Георгий Александрович,**

доктор биологических наук,

Федеральное государственное бюджетное учреждение науки Институт проблем передачи информации им. А.А. Харкевича Российской академии наук.

**Официальные оппоненты:**

**Трифонов Владимир Александрович,**

доктор биологических наук, заведующий лабораторией,

Федеральное государственное бюджетное учреждение науки Институт молекулярной и клеточной биологии Сибирского отделения Российской академии наук.

**Медведева Юлия Анатольевна,**

кандидат биологических наук, руководитель группы,

Федеральное государственное учреждение «Федеральный исследовательский центр «Фундаментальные основы биотехнологии» Российской академии наук».

**Ведущая организация:**

Федеральное государственное бюджетное учреждение науки Институт общей генетики им. Н.И. Вавилова Российской академии наук.

Защита диссертации состоится 06 июня 2022 года в 17-00 на заседании диссертационного совета 24.1.101.01 при Федеральном государственном бюджетном учреждении науки Институте проблем передачи информации им. А.А. Харкевича Российской академии наук по адресу: 127051, г. Москва, Большой Каретный переулок, д. 19, стр. 1.

С диссертацией можно ознакомиться в библиотеке Федерального государственного бюджетного учреждения науки Института проблем передачи информации им. А.А. Харкевича Российской академии наук (ИППИ РАН), а также на сайте ИППИ РАН.

Автореферат разослан «\_\_\_» \_\_\_\_\_ 2022 г.

Ученый секретарь диссертационного совета  
доктор биологических наук

Казенников О.В.

## Общая характеристика работы

### Актуальность темы исследования

Отрицательный отбор и рекомбинация – важные факторы эволюции геномных последовательностей. Накопленные за последние годы данные секвенирования геномов для большого количества видов позволили понять многое про то, как последовательности изменяются в результате мутаций, отбора и рекомбинации. Развитие технологий секвенирования дает возможность определять и сравнивать не только последовательности геномов особей из разных видов, но и анализировать множественные последовательности геномов особей, принадлежащих к одному и тому же виду. В результате появилась возможность изучать действие отбора и других эволюционных процессов на различных эволюционных масштабах.

Методы сравнительной геномики позволяют проводить поиск участков генома, находящихся под действием отбора, а также изучать зависимость эффективности отбора от разных факторов, в том числе от частоты рекомбинации. Одним из наиболее интересных направлений исследований, которые могут быть проведены с использованием данных по внутривидовой изменчивости, накопленных за последние годы, является изучение эпистаза – явления зависимости эффектов мутации от геномного контекста, в котором она произошла.

Другое направление исследований, ставшее доступным с распространением технологий секвенирования, заключается в поиске «подписей» рекомбинации в геномах видов, считающихся бесполовыми. В случае микроскопических организмов убедительно доказать отсутствие полового размножения или выявить свидетельства криптического обмена генетическим материалом с применением классических подходов может быть чрезвычайно сложно. В то же время с помощью сравнения последовательностей геномов можно провести анализ совместимости структуры внутривидовой изменчивости с тем, что ожидается в случае отсутствия полового размножения и рекомбинации. Так, с использованием такого подхода признаки «криптической» рекомбинации были выявлены в геномах паразита *Giardia lamblia*<sup>1</sup>, считавшегося бесполом, и геномах вида из группы Placozoa<sup>2</sup>, группы, вопрос о существовании полового размножения в которой длительное время оставался без ответа. Таким образом, выявление и изучение сигналов отбора и рекомбинации в полногеномных данных позволяет отвечать на различные важные биологические вопросы, многие из которых долгое время оставались открытыми.

---

<sup>1</sup> Cooper M.A. et al. Population genetics provides evidence for recombination in *Giardia* // Curr. Biol. 2007. Vol. 17, № 22. P. 1984–1988.

<sup>2</sup> Signorovitch A.Y., Dellaporta S.L., Buss L.W. Molecular signatures for sex in the Placozoa // PNAS. 2005. Vol. 102, № 43. P. 15518–15522.

## Степень разработанности темы

К настоящему моменту опубликовано множество работ, посвященных поиску сигнала отрицательного отбора и рекомбинации в геномных данных разных типов. Выявление сигнала отрицательного отбора по данным дивергенции обычно проводится на основе степени консервативности рассматриваемого участка в геномах далеких видов. Несмотря на большое число методов, направленных на поиск сигнала отрицательного отбора, в некоторых случаях выявление сигнала отрицательного отбора или определение типа отбора является сложной задачей. В данной работе рассмотрено два таких случая. Первый случай соответствует ситуации, в которой рассматриваются ортологичные последовательности из далеких видов, которые могли эволюционировать под действием продолжающегося отрицательного отбора, направленного на сохранение функции, но не обязательно направленного на сохранение последовательности. Такая ситуация может возникнуть, например, для регуляторных элементов, если происходит относительно быстрая эволюция набора сайтов связывания или их взаимного расположения. В этом случае методы, основанные на поиске консервативных участков генома, не выявят сигнал отрицательного отбора. Другой случай, в котором детекция отрицательного отбора может быть осложнена, соответствует ситуации поиска сигнала отбора (в первую очередь в этом контексте интересен эпистатический отбор) на основании данных по внутривидовому полиморфизму, сигнал отбора или эпистатических взаимодействий в которых может быть очень слабым.

В области исследований сигнала отрицательного отбора отдельное направление посвящено изучению эволюции консервативных некодирующих последовательностей. Такие последовательности были описаны как для геномов позвоночных, так и для геномов двукрылых<sup>3, 4</sup>. Согласно оценкам, приведенным в одной из ранних работ в этой области, от 0.3 до 1% генома человека соответствует консервативным некодирующим областям, находящимся под давлением сильного отбора у большинства млекопитающих<sup>3</sup>. Несмотря на то, что функциональное значение большинства консервативных некодирующих элементов неизвестно, результаты значительного числа исследований указывают на то, что такие элементы, по всей видимости, часто выполняют регуляторную функцию<sup>5, 6</sup>, в частности играют роль энхансеров или инсуляторов. Однако оценки доли генома, находящейся под действием отрицательного отбора, полученные на основе оценки консервативности,

---

<sup>3</sup> Dermitzakis E.T. et al. Evolutionary discrimination of mammalian conserved non-genic sequences (CNGs) // Science. 2003. Vol. 302, № 5647. P. 1033–1035.

<sup>4</sup> Bergman C.M., Kreitman M. Analysis of conserved noncoding DNA in *Drosophila* reveals similar constraints in intergenic and intronic sequences // Genome Res. 2001. Vol. 11, № 8. P. 1335–1345.

<sup>5</sup> Visel A. et al. Ultraconservation identifies a small subset of extremely constrained developmental enhancers // Nature Genetics. 2008. Vol. 40, № 2. P. 158–160.

<sup>6</sup> Xie X. et al. Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites // PNAS. 2007. Vol. 104, № 17. P. 7145–7150.

вероятно, являются заниженными, поскольку не все функциональные элементы сохраняют сходство последовательностей на больших эволюционных расстояниях. Вопрос о возможном сохранении функции участка генома без сохранения сходства последовательностей особенно интересен в контексте эволюции регуляторных последовательностей. Так, для функциональных некодирующих последовательностей ДНК описан ряд случаев, в которых некодирующие последовательности из разных организмов могут выполнять похожие функции и, скорее всего, имеют общее происхождение, несмотря на отсутствие между ними осмысленного выравнивания<sup>7</sup>. Например, энхансер гена человека может обеспечивать нормальную экспрессию ортологичного гена в трансгенных *Danio rerio*, притом что сходство последовательностей между энхансерами человека и *D. rerio* отсутствует<sup>8</sup>. Однако насколько нам известно, до нашей работы на уровне полного генома не проводилось изучения явления, при котором отрицательный отбор может продолжать действовать на ортологичные участки генома, утратившие в далеких видах сходство последовательностей.

Помимо вопроса о действии отбора на определенные участки генома, большое значение имеет вопрос о типе отбора, в частности, действует ли отбор на каждую мутацию независимо от геномного контекста или является эпистатическим. Изучение распространенности и типа эпистатического отбора на вредные мутации в естественных популяциях важно для понимания того, как популяциям человека и других живых существ удается противостоять постоянному притоку вредных мутаций. Большой объем теоретических работ посвящен влиянию эпистатического отбора на мутационный груз в популяциях с половым и бесполом размножением<sup>9, 10, 11</sup>. Основным результатом этих работ заключается в том, что в случае существования синергического (усиливающего) эпистаза между вредными мутациями мутационный груз в популяции с половым размножением ниже, чем в том случае, если отбор действует на каждую мутацию по отдельности. При этом при бесполом размножении мутационный груз не зависит от типа отбора<sup>9</sup>. Таким образом, с одной стороны, синергические эпистатические взаимодействия между вредными мутациями являются возможным объяснением парадокса мутационного груза<sup>12</sup>, а, с другой стороны, могут быть одним из ключевых факторов, определяющих преимущество полового размножения над бесполом

---

<sup>7</sup> Taher L. et al. Genome-wide identification of conserved regulatory function in diverged sequences // *Genome Res.* 2011. Vol. 21, № 7. P. 1139–1149.

<sup>8</sup> Fisher S. et al. Conservation of *RET* regulatory function from human to zebrafish without sequence similarity // *Science.* 2006. Vol. 312, № 5771. P. 276–279.

<sup>9</sup> Kimura M., Maruyama T. The mutational load with epistatic gene interactions in fitness // *Genetics.* 1966. Vol. 54, № 6. P. 1337–1351.

<sup>10</sup> Kondrashov A.S. Selection against harmful mutations in large sexual and asexual populations // *Genet. Res.* 1982. Vol. 40, № 3. P. 325–332.

<sup>11</sup> Charlesworth B. Mutation-selection balance and the evolutionary advantage of sex and recombination // *Genet. Res.* 1990. Vol. 55, № 3. P. 199–221.

<sup>12</sup> Kondrashov A.S. Deleterious mutations and the evolution of sexual reproduction // *Nature.* 1988. Vol. 336, № 6198. P. 435–440.

<sup>10, 11, 12</sup>. Несмотря на то, что эпистатические взаимодействия разных типов были описаны для большого числа мутаций, до недавнего времени вопрос о том, насколько распространен синергический эпистаз между вредными мутациями на уровне всего генома, оставался открытым.

Помимо гипотезы, объясняющей преимущество полового размножения и рекомбинации более низким мутационным грузом (в случае присутствия синергических взаимодействий), существует целый ряд теорий, предлагающих другие объяснения тому факту, что половое размножение преобладает среди эукариот. Так, существуют гипотезы, объясняющие преимущество полового размножения над бесполом эффектами, связанными с дрейфом генов<sup>13, 14</sup>, более высокой эффективностью положительного отбора<sup>15, 16</sup>, и более высокой скоростью приспособления к изменяющимся условиям среды<sup>17</sup>.

Какие именно из этих факторов в действительности создают преимущество для полового размножения, неизвестно, но, по всей видимости, это преимущество является значительным, т.к. переход к бесполому размножению обычно заканчивается относительно быстрым вымиранием. На этом основании бесполое размножение часто рассматривается как «эволюционный тупик». В связи с этим внимание исследователей привлекли немногочисленные исключения из этого правила – предположительно древние группы бесполой организмов. В качестве наиболее яркого примера такой группы обычно приводили класс бделлоидных коловраток, группу микроскопических беспозвоночных, как считалось, отказавшихся от полового размножения десятки миллионов лет назад. Основным аргументом в пользу строго бесполого размножения у видов этой группы служило отсутствие самцов среди сотен тысяч особей бделлоидных коловраток, проанализированных разными исследователями<sup>18</sup>. Однако молекулярно-генетические и геномные данные, полученные в последние годы, не позволяли сделать убедительный вывод о существовании или отсутствии полового размножения и рекомбинации у видов этой группы. Так, анализ первого опубликованного генома бделлоидной коловратки *Adineta vaga*, вышедший в 2013, дал основания предполагать, что структура этого генома несовместима с классическим мейозом<sup>19</sup>. Однако эти данные не нашли подтверждения при

---

<sup>13</sup> Muller H.J. The relation of recombination to mutational advance // *Mutat. Res.* 1964. Vol. 106. P. 2–9.

<sup>14</sup> Haigh J. The accumulation of deleterious genes in a population—Muller’s Ratchet // *Theor. Popul. Biol.* 1978. Vol. 14, № 2. P. 251–267.

<sup>15</sup> Fisher R.A. *The genetical theory of natural selection.* Oxford: Oxford University Press, 1930.

<sup>16</sup> Rice W.R., Chippindale A.K. Sexual recombination and the power of natural selection // *Science.* 2001. Vol. 294, № 5542. P. 555–559.

<sup>17</sup> Kondrashov A.S. Classification of hypotheses on the advantage of amphimixis // *J. Hered.* 1993. Vol. 84, № 5. P. 372–387.

<sup>18</sup> Birky C.W. Positively negative evidence for asexuality // *J. Hered.* 2010. Vol. 101 Suppl 1. P. S42–45.

<sup>19</sup> Flot J.-F. et al. Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga* // *Nature.* 2013. Vol. 500, № 7463. P. 453–457.

анализе генома другой коловратки из рода *Adineta*, *A. ricciae*<sup>20</sup>, а также при анализе новой сборки генома хромосомного уровня, недавно полученной для *A. vaga* с использованием комбинации различных технологий секвенирования<sup>21</sup>. Первая работа в области популяционной геномики бделлоидных коловраток была опубликована в 2015 году: в этой работе был проведен анализ нескольких участков ядерного генома у 6 особей вида *Macrotrachela quadricornifera*<sup>22</sup>. Результаты этого анализа были интерпретированы как вероятное свидетельство чрезвычайно редкого типа мейоза, описанного ранее у растений из рода *Oenothera*, происходящего таким образом, что рекомбинация затрагивает только теломерные области хромосом, без выравнивания гомологичных хромосом по длине относительно друг друга<sup>22</sup>. Работ, в которых на полногеномных данных проводили бы поиск рекомбинации и «подписей» генетического обмена у бделлоидных коловраток, до последнего времени опубликовано не было.

### Цели и задачи исследования

Целью данной работы являлся поиск сигнала отрицательного отбора и рекомбинации в данных разных типов, в том числе поиск возможных свидетельств рекомбинации в геномах бделлоидных коловраток, которых ранее рассматривали как группу предположительно древних бесполок видов.

Для достижения этой цели были поставлены следующие задачи:

- 1) изучение возможного продолжения действия отрицательного отбора на ортологичные некодирующие участки генома, потерявшие сходство последовательностей в далеких видах;
- 2) поиск сигнала эпистатического отбора, действующего на вредные аллели в белок-кодирующих генах, по данным внутривидовой изменчивости *Drosophila melanogaster*;
- 3) поиск сигнала рекомбинации и обмена генетическим материалом в данных по внутривидовой изменчивости бделлоидной коловратки вида *Adineta vaga* и исследование совместимости данных по внутривидовой изменчивости *A. vaga* с различными эволюционными сценариями.

### Научная новизна

В данной работе был получен ответ на ряд новых и сформулированных ранее, но оставшихся открытыми, вопросов в области эволюционной геномики. Так, нами на полногеномных данных было получено свидетельство о продолжении действия отрицательного отбора на ортологичные некодирующие

---

<sup>20</sup> Nowell R.W. et al. Comparative genomics of bdelloid rotifers: Insights from desiccating and nondesiccating species // PLOS Biology. 2018. Vol. 16, № 4. P. e2004830.

<sup>21</sup> Simion P. et al. Chromosome-level genome assembly reveals homologous chromosomes and recombination in asexual rotifer *Adineta vaga* // Sci. Adv. 2021. Vol. 7, № 41. P. eabg4216.

<sup>22</sup> Signorovitch A. et al. Allele sharing and evidence for sexuality in a mitochondrial clade of bdelloid rotifers // Genetics. 2015. Vol. 200, № 2. P. 581–590.

участки генома, потерявшие в далеких видах сходство последовательностей. Примеры этого явления были описаны ранее на основании экспериментальных данных, но, насколько нам известно, до нашей работы на уровне всего генома это явление исследовано не было.

Кроме того, в данной работе впервые на уровне всего генома был выявлен сигнал синергического эпистатического отбора, действующего на вредные аллели в популяции *D. melanogaster*. Возможное существование синергических эпистатических взаимодействий между вредными мутациями ранее широко обсуждалось в литературе, в первую очередь, в теоретических работах, в которых исследовалась зависимость мутационного груза от присутствия и типа эпистатических взаимодействий. Тем не менее вопрос о существовании и преобладающем типе эпистатических взаимодействий на уровне всего генома у эукариотических видов оставался малоизученным.

В работе впервые на полногеномных данных выявлен сигнал рекомбинации и обмена генетическим материалом для вида, относящегося к группе бделлоидных коловраток. Вопрос о возможном существовании обмена генетическим материалом у бделлоидных коловраток исследовался ранее, но в предыдущих работах, посвященных изучению этого вопроса, использовались или полногеномные последовательности одного индивидуума, или последовательности небольшого числа геномных локусов. Поиск сигнала рекомбинации у бделлоидных коловраток на полногеномных данных ранее также не проводился. Таким образом, в данной работе впервые получены полногеномные свидетельства рекомбинации у бделлоидных коловраток. В предыдущих исследованиях выдвигались различные гипотезы относительно возможного механизма обмена генетическим материалом у бделлоидных коловраток. В частности, обсуждалась возможность существования горизонтального переноса генов внутри популяций видов из этой группы. В рамках выполнения исследования получены указания на то, что половое размножение является более вероятным объяснением обмена генетическим материалом и рекомбинации в популяциях бделлоидных коловраток.

### **Теоретическая и практическая значимость**

Результаты, полученные в данной работе, позволяют лучше понять некоторые фундаментальные эволюционные закономерности и имеют в первую очередь теоретическую значимость. Исследование сигнала отрицательного отбора, продолжающего действовать на ортологичные некодирующие участки генома после утраты очевидного сходства последовательностей, представляет интерес с точки зрения изучения принципов эволюции функциональных некодирующих последовательностей. Результаты, указывающие на вероятное существование синергического эпистаза, сигнал которого был выявлен в данной работе у *D. melanogaster* для аллелей, вызывающих потерю функции гена, являются важными с точки зрения исследований парадокса мутационного груза и возможных причин преобладания полового размножения. Свидетельства рекомбинации и обмена генетическим материалом у бделлоидной коловратки *A. vaga*, выявленные в данной работе, имеют



фундаментальное значение, поскольку потенциально позволяют объяснить эволюционный парадокс, которым считалось существование видов этой группы. Кроме того, обнаружение сигнала рекомбинации в геномах бделлоидных коловраток, которых в течение длительного времени относили к группам древних бесполок видов, по всей видимости, является важным аргументом, говорящим о важности рекомбинации для долгосрочного эволюционного успеха.

Некоторые полученные результаты и подходы, примененные в рамках данной работы, могут иметь практическую значимость. Так, результаты, указывающие на то, что отрицательный отбор может продолжать действовать на имеющие общее происхождение участки генома даже после того, как они разошлись в далеких видах до неузнаваемости, могут иметь значение с точки зрения разработки методов выявления функциональных элементов, методов поиска последовательностей, находящихся под действием отрицательного отбора, и поиска гомологичных последовательностей на далеких эволюционных расстояниях. Подход к поиску сигнала обмена генетическим материалом, основанный на анализе трехаллельных сайтов и использовавшийся в данной работе при анализе данных *A. vaga*, может быть применен и в других исследованиях с похожей проблематикой.

### **Методология и методы исследования**

Для решения поставленных задач использовалось множество методов биоинформатики, сравнительной и популяционной геномики. Использованные методы включают выравнивание ортологичных белковых и нуклеотидных последовательностей, определение ортологичных последовательностей в разных геномах, построение филогенетических деревьев (с помощью методов максимального правдоподобия и метода ближайших соседей). Часть задач осуществлялась с использованием открытых геномных данных. Кроме того, в рамках реализации работы были получены собственные данные секвенирования 11 геномов бделлоидных коловраток вида *A. vaga*. Для последующего анализа полученных данных секвенирования использовались стандартные методы первичной обработки и картирования прочтений, определения однонуклеотидных полиморфизмов. Определение однонуклеотидных полиморфизмов для особей *A. vaga* подразумевало использование референтного генома. В связи с этим для одной из анализируемых особей была получена геномная сборка, далее использовавшаяся в качестве референтной последовательности. Кроме того, для локальной реконструкции гаплотипов в работе был использован метод вычислительного фазирования полиморфизмов. Поиск сигнала рекомбинации проводили с помощью широко применяемых методов для выявления рекомбинации и с применением модифицированного нами варианта четырехгаметного теста. Исследование внутрипопуляционной изменчивости *A. vaga* включало определение коэффициента инбридинга для биаллельных полиморфных сайтов и анализ трехаллельных сайтов.

## **Основные положения, выносимые на защиту**

1. Действие отрицательного отбора на ортологичные интроны может продолжаться даже тогда, когда между последовательностями имеющих общее происхождение интронов в далеких видах уже не существует осмысленного выравнивания. Сигнал, свидетельствующий о том, что давление отрицательного отбора на ортологичные, но разошедшиеся до неузнаваемости последовательности, может сохраняться, присутствует в геномных данных как для позвоночных, так и для двукрылых. Это явление может быть связано с сохранением предковой функции некодирующих участков генома, утративших в далеких видах сходство последовательностей.

2. Понижение дисперсии мутационной нагрузки по сравнению с аддитивной дисперсией для аллелей, вызывающих потерю функции гена, в двух популяциях *D. melanogaster* указывает на существование синергических эпистатических взаимодействий между мутациями данного типа. Сигнал синергического эпистаза также присутствует в подмножестве несинонимических аллелей, попадающих в гены *D. melanogaster*, находящиеся под сильным давлением отрицательного отбора.

3. В геномах бделлоидных коловраток вида *A. vaga* выявлен сигнал рекомбинации, который не может быть объяснен исключительно действием генной конверсии и, вероятно, связан с реципрокной рекомбинацией. Данные по внутривидовой изменчивости *A. vaga* свидетельствуют об обмене генетическим материалом, происходящем в популяциях этого вида. Некоторые закономерности, выявленные при анализе филогений гаплотипов разных особей, указывают на то, что половое размножение является более вероятным механизмом обмена генетическим материалом внутри популяций бделлоидных коловраток, чем горизонтальный перенос генов.

## **Степень достоверности и апробация результатов**

По материалам диссертации опубликовано три статьи в рецензируемых научных журналах. Результаты работы были представлены на встречах Общества молекулярной биологии и эволюции (Society for Molecular Biology and Evolution) в 2012, 2017 и 2019 годах (SMBE 2012 – Дублин, Ирландия; SMBE 2017 – Остин, Техас, США; SMBE 2019 – Манчестер, Англия) и Московской международной конференции по вычислительной молекулярной биологии в 2019 году (Moscow Conference on Computational Molecular Biology, MCCMB'19 – Москва, Россия), а также на конференциях «Информационные технологии и системы» в 2012, 2016 и 2018 годах (ИТиС 2012 – Петрозаводск, Россия; ИТиС 2016 – Репино, Санкт-Петербург, Россия; ИТиС 2018 – Казань, Россия).

## **Структура и объем диссертации**

Диссертация изложена на 218 страницах машинописного текста и содержит следующие разделы: введение, обзор литературы, результаты и обсуждение в трех главах, заключение и выводы. В конце приведён список

литературы. Материал включает 42 рисунка, 31 таблицу, 4 таблицы в приложении, а также список литературы, содержащий 216 ссылок.

## Содержание работы

**Глава 1** содержит обзор литературы по теме диссертации, включающий четыре раздела.

В **разделе 1.1** приводится краткий обзор работ, посвященных поиску консервативных некодирующих элементов. Обсуждается возможное функциональное значение консервативных некодирующих элементов и рассматриваются подходы к их поиску.

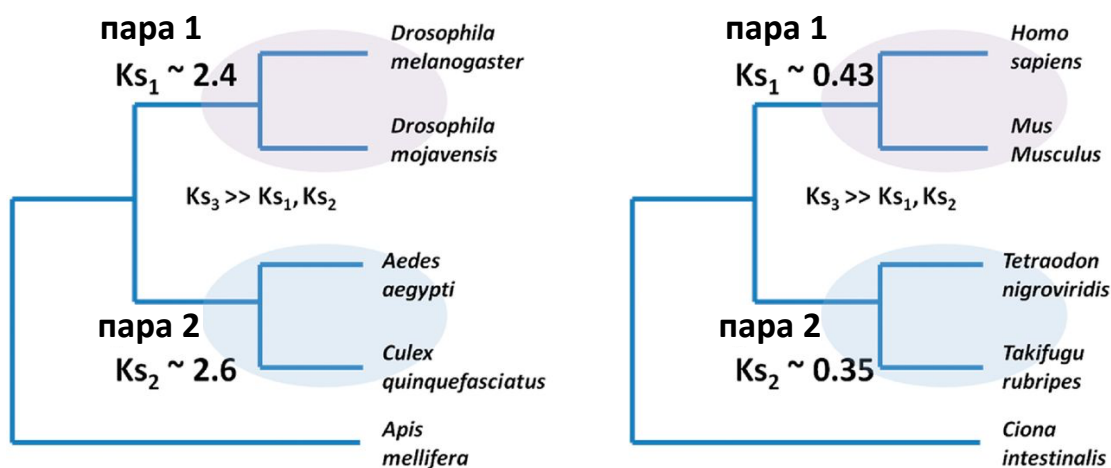
В **разделе 1.2** рассматривается явление синергического эпистаза и обсуждается парадокс мутационного груза. Синергический эпистаз рассматривается как вероятный фактор, благодаря которому половое размножение может получать преимущество перед бесполом. Приводится обзор теоретических работ, в которых исследовали зависимость мутационного груза от существования и типа эпистатических взаимодействий.

**Раздел 1.3** содержит обзор различных гипотез, объясняющих преимущество полового размножения перед бесполом и преобладание полового размножения среди эукариот.

В **разделе 1.4** обсуждаются предполагаемые геномные «подписи» бесполого размножения, которые ожидаются в популяционно-генетических данных, полученных для бесполок видов. Приводится обзор предположительно древних групп бесполок организмов. Рассматриваются данные, поддерживающие бесполой статус таких групп или свидетельствующие против него.

В **главе 2** представлены результаты анализа действия отрицательного отбора на ортологичные интроны в далеких видах. С одной стороны, отбор на сохранение функции гомологичных участков генома может сопровождаться сохранением очевидного сходства на уровне последовательностей, которое может поддерживаться в течение миллиардов лет. С другой стороны, известны случаи, когда имеющие общее происхождение, но утратившие сходство последовательностей участки генома продолжают выполнять предковую функцию в далеких видах. Так, например, энхансер гена человека может обеспечивать нормальную экспрессию ортологичного гена в трансгенных *D. rerio*, притом что сходство последовательностей между энхансерами человека и *D. rerio* отсутствует<sup>8</sup>. Однако изучение таких ситуаций ранее в основном заключалось в описании отдельных случаев. В связи с этим большой интерес представляет систематическое изучение явления, при котором отбор может продолжать действовать на ортологичные участки генома, которые уже

утратили сходство на уровне последовательностей. Для того, чтобы изучить возможное существование этого явления на уровне всего генома, мы проанализировали гомологичные интроны в двух четверках видов – в четверке видов насекомых и в четверке видов позвоночных. Каждая четверка состоит из двух пар геномов, подобранных так, что геномы внутри пары находятся на умеренно большом филогенетическом расстоянии друг от друга, а расстояние, разделяющее разные пары, значительно превышает расстояния внутри пар (Рисунок 1). Первая четверка – четверка двукрылых – включает в себя два вида из рода *Drosophila*, *D. melanogaster* и *D. mojavensis* (пара 1), и два вида комаров, *Culex quinquefasciatus* и *Aedes aegypti* (пара 2). Вторая четверка – четверка позвоночных – состоит из двух видов млекопитающих, *Homo sapiens* и *Mus musculus* (пара 1), и двух видов рыб, *Tetraodon nigroviridis* и *Takifugu rubripes* (пара 2). Для анализа были выбраны интроны, поскольку ортологические соответствия между интронами могут быть установлены на основании выравнивания фланкирующих экзонов даже на таких филогенетических расстояниях, на которых осмысленное выравнивание между последовательностями интронов отсутствует.

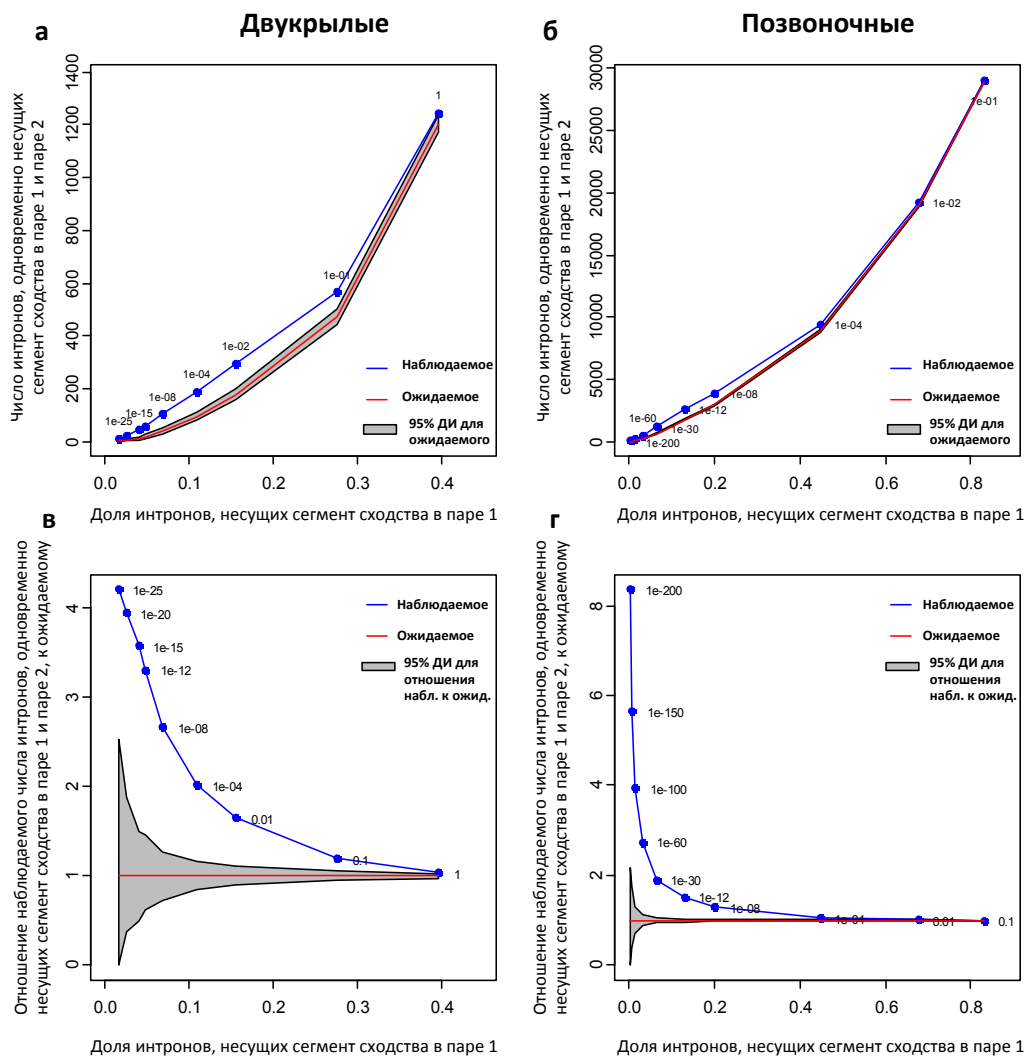


**Рисунок 1.** Две четверки видов, для которых проводился анализ. Четверка двукрылых (слева) и четверка позвоночных (справа). Для каждой четверки отображена соответствующая аутгруппа. Показаны эволюционные расстояния для видов внутри каждой пары, выраженные через ожидаемое число синонимических замен на сайт,  $Ks$ .

На первом этапе мы проанализировали распределение консервативных сегментов по интронам в двух парах видов, принадлежащих к четверке. Мы предположили, что в том случае, если действие отбора продолжается и после того, как сходство последовательностей между видами из разных пар потеряно, то наличие консервативного (и, следовательно, вероятно, функционального) сегмента в одной паре видов, входящей в четверку, будет значимым предиктором наличия консервативного сегмента в ортологичном интроне между видами во второй паре. Рассмотрим ортологичные некодирующие участки генома, сохранившие одну и ту же функцию во всех видах,

принадлежащих к одной четверке. Сохранение функции должно приводить к более высокому уровню сохранения сходства последовательностей между видами внутри каждой пары по сравнению с ожидаемым в предположении нейтральной скорости эволюции. Кроме того, сильный отбор на сохранение функции может приводить к сохранению существенного сходства последовательностей и на значительно больших эволюционных расстояниях, разделяющих виды из разных пар. В этом случае консервативность последовательности может присутствовать между геномами из разных пар. Поскольку мы были заинтересованы в изучении возможных случаев сохранения функции без сохранения сходства последовательностей, мы исключили из рассмотрения интроны, несущие участки значимого сходства между видами из разных пар. Тем не менее мы показали, что интрон, имеющий участок локального сходства между видами в одной из пар внутри четверки, несет участок локального сходства в другой паре чаще, чем ожидалось бы в том случае, если бы участки локального сходства в каждой паре были распределены независимо друг от друга. Избыток интронов, одновременно несущих консервативный сегмент внутри каждой пары, присутствует как в том случае, если анализ проводился без учета длин интронов, так и в том случае, если при анализе учитывалась неоднородность длин интронов (Рисунок 2). Кроме того, мы показали, что данный результат не может быть объяснен свойствами гена или тем, что первые по порядку интроны чаще несут консервативные сегменты.

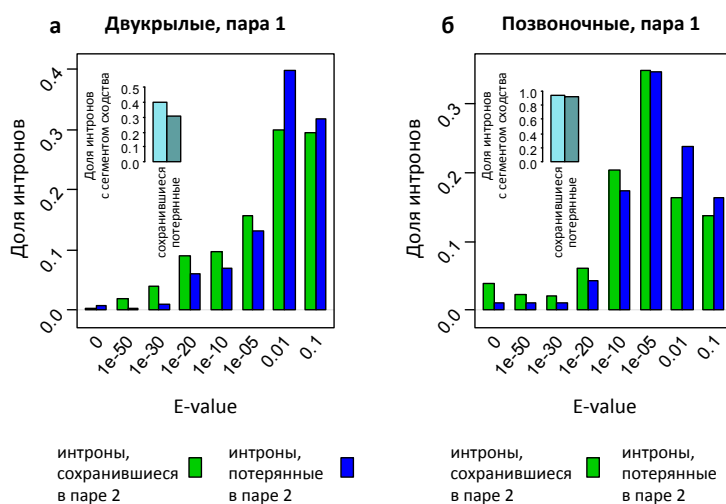
На следующем этапе мы сформулировали и проверили еще одно предсказание гипотезы о сохранении давления отрицательного отбора на ортологичные, но разошедшиеся до неузнаваемости последовательности. Мы предположили, что если некоторые регуляторные элементы существуют дольше, чем сходство последовательностей соответствующих участков генома, мы ожидали бы, что интроны, несущие регуляторный сегмент в паре 1, чаще, чем ожидается по случайным причинам, несут консервативный участок в паре 2. Для того, чтобы проверить эту гипотезу, мы использовали полногеномную разметку генома по регуляторным элементам, предсказанным на основе данных по модификациям хроматина. Действительно, мы обнаружили, что интроны *D. melanogaster*, пересекающиеся с областями генома, обогащенными активными модификациями хроматина (и, следовательно, вероятно вовлеченными в регуляторные процессы), до ~3-х раз чаще несут участок значимого сходства последовательностей во второй паре из четверки двукрылых (комары *C. quinquefasciatus* – *A. aegypti*). Аналогичным образом, интроны, пересекающиеся с инсуляторами или энхансерами у человека, до 2.6 и 1.4 раз соответственно чаще по сравнению с ожиданием несут участок значимого сходства последовательностей во второй паре из четверки позвоночных (рыбы *T. nigroviridis* – *T. rubripes*).



**Рисунок 2.** Интроны, несущие сегмент локального сходства последовательностей между видами из одной пары в четверке, с повышенной частотой также несут сегмент локального сходства в другой паре видов в четверке (с поправкой на неоднородность длин интронов). Синие точки соответствуют различным значениям E-value bl2seq (указаны рядом с точками), использованным как пороговые значения для поиска участков локального сходства между ортологичными интронами в паре 1 и паре 2. Более низкие значения E-value соответствуют более строгим порогам для выявления сходства последовательностей. На горизонтальной оси для данного значения E-value отмечена доля интронов, несущих сегмент локального сходства между ортологичными интронами в паре 1, среди интронов, присутствующих во всех видах в четверке. На вертикальной оси отмечено число (**а, б**) или отношение наблюдаемого к ожидаемому (**в, г**) для числа интронов, одновременно несущих сегмент сходства в паре 1 и паре 2. (**а, в**) – двукрылые; (**б, г**) – позвоночные. Отношение наблюдаемого числа к ожидаемому рассчитывали как отношение наблюдаемого числа интронов с участком сходства в обеих парах в четверке к ожидаемому в том случае, если бы участки сходства были случайно распределены по интронам в каждой паре видов, контролируя на длину интронов. Красная линия и серая область соответствуют среднему значению и 95% доверительному интервалу (ДИ) для ожидаемых значений, рассчитанным на основе 10,000 случайных выборок интронов.

Затем мы предположили, что присутствие функционального сегмента в интронах может выражаться не только в сохранении последовательности интронов в далеких видах, но и в низкой скорости потерь таких интронов в эволюции. Чтобы проверить эту гипотезу, мы проанализировали распределение участков локального сходства для тех интронов из пары 1, которые сохранились в паре 2, по сравнению с интронами, которые были потеряны в паре 2. Поскольку при проведении соответствующего анализа необходимо отличать потери интронов от приобретений интронов, мы использовали информацию из вида-аутгруппы для определения предкового состояния (присутствие или отсутствие интрона в общем предке четверки видов).

В соответствии с гипотезой о продолжении действия отрицательного отбора после потери сходства на уровне последовательностей как в четверке двукрылых, так и в четверке позвоночных участки локального сходства в паре 1 оказались более консервативными среди интронов, сохранившихся в паре 2, по сравнению с интронами, потерянными в паре 2 (Рисунок 3). Консервативность оценивали как E-value локального выравнивания, построенного для пары интронов с применением программы bl2seq. Распределение значений E-value участков локального сходства в паре 1 для интронов, сохранившихся в паре 2, смещено в сторону более низких значений по сравнению с распределением E-value для участков локального сходства в паре 1 для интронов, потерянных в паре 2 (критерий Манна-Уитни, P-значение = 0.00103 для четверки двукрылых, P-значение = 0.0203 для четверки позвоночных; Рисунок 3). Похожие результаты были получены при аналогичном сравнении интронов из пары 2, сохранившихся в паре 1, с интронами из пары 2, потерянными в паре 1.



**Рисунок 3.** Последовательность интронов, сохранившихся в видах из пары 2, более консервативна в видах из пары 1. Для двух четверок видов (а – двукрылые; б – позвоночные) показаны распределения значений E-value для сегментов локального сходства между ортологичными интронами в паре 1. Распределения показаны для двух групп интронов – для интронов, сохранившихся в двух видах в паре 2 (зелёный) и потерянных хотя бы в одном виде из пары 2 (синий). Значения E-value, отмеченные на горизонтальной оси, соответствуют нижнему пороговому значению. На врезках для тех же двух групп показаны доли интронов,

для которых в паре 1 находится хотя бы какое-то локальное выравнивание (bl2seq E-value  $\leq 1$ ).

Таким образом, различные типы анализов свидетельствуют о том, что действие отрицательного отбора на ортологичные участки генома может продолжаться даже тогда, когда между последовательностями этих участков в далеких видах уже не существует осмысленного выравнивания. Возможное объяснение этого явления заключается в том, что предковая функция некодирующих участков генома в далеких видах может сохраняться и после потери сходства последовательностей.

В главе 3 представлены результаты поиска сигнала синергического эпистаза между вредными аллелями в данных по популяционной изменчивости *D. melanogaster*. Вопрос о существовании и распространенности синергических эпистатических взаимодействий между вредными мутациями представляет большой интерес, поскольку синергические эпистатические взаимодействия могут быть возможным объяснением важных проблем в эволюционной биологии: парадокса мутационного груза и причин преобладания полового размножения среди эукариот. Ранее в теоретических работах было показано, что присутствие синергического эпистаза должно изменять форму распределения мутационной нагрузки, снижая дисперсию ( $\sigma^2$ ) этого распределения относительно аддитивной дисперсии ( $V_A$ )<sup>11, 23, 24</sup>. Опираясь на эти теоретические результаты, в данной главе мы провели поиск подписей эпистатического отрицательного отбора на открытых данных по полногеномному полиморфизму *D. melanogaster*. Для этого мы проанализировали распределение мутационной нагрузки в двух популяциях *D. melanogaster* для разных типов аллелей. В то время как распределение мутационной нагрузки для синонимических аллелей характеризуется сверхдисперсией ( $\sigma^2/V_A > 1$ ), в замбийской популяции *D. melanogaster* (DPGP3) для вредных аллелей (нонсенс-аллелей и общей категории аллелей, вызывающих потерю функции гена, включающей нонсенс-аллели и аллели, вызывающие поломку сайтов сплайсинга) было выявлено понижение дисперсии относительно аддитивной дисперсии ( $\sigma^2/V_A < 1$ , Таблица 1). Так, для распределения числа редких минорных аллелей, вызывающих потерю функции гена,  $\sigma^2/V_A = 0.929$ , если же рассматривать все минорные аллели, вызывающие потерю функции гена,  $\sigma^2/V_A = 0.851$  (Таблица 1). Понижение дисперсии относительно аддитивной дисперсии для распределения мутационной нагрузки вредных аллелей указывает на вероятное существование у *D. melanogaster* синергических эпистатических взаимодействий между вредными аллелями. Значимость понижения дисперсии для вредных аллелей была показана с помощью сравнения наблюдаемых значений  $\sigma^2/V_A$  для вредных аллелей со

<sup>23</sup> Bulmer M.G. The mathematical theory of quantitative genetics. Oxford: Oxford University Press, 1980. 276 p.

<sup>24</sup> Kondrashov A.S. Dynamics of unconditionally deleterious mutations: Gaussian approximation and soft selection // Genet. Res. 1995. Vol. 65, № 2. P. 113–121.

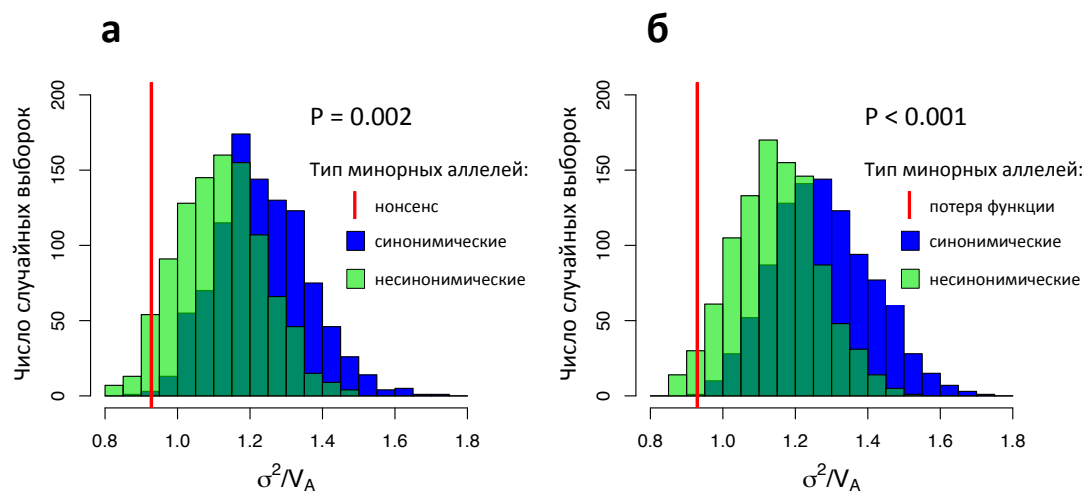


значениями  $\sigma^2/V_A$  в 1000 контрольных выборок синонимических аллелей с соответствующими популяционными частотами (Таблица 1, Рисунок 4).

Популяция DPGP3, весь геном					
Частота минорного аллеля $\leq 5$					
Тип минорных аллелей	Среднее число аллелей на особь	$V_A$	Дисперсия ( $\sigma^2$ )	$\sigma^2/V_A$	Р-значение
Синонимические	3,577.06	3,544.10	203,688.98	57.473	
Несинонимические	2,051.52	2,036.36	37,746.26	18.536	
Нонсенс	10.21	10.15	9.42	0.928	0.002
Поломка сайта сплайсинга	2.60	2.59	2.45	0.948	0.177
Потеря функции	12.81	12.74	11.84	0.929	<0.001
Частота минорного аллеля <50%					
Тип минорных аллелей	Среднее число аллелей на особь	$V_A$	Дисперсия ( $\sigma^2$ )	$\sigma^2/V_A$	Р-значение
Синонимические	50,666.23	37,671.02	1,418,266.56	37.649	
Несинонимические	11,980.92	9,520.01	25,950.18	2.726	
Нонсенс	25.84	22.89	20.44	0.893	0.010
Поломка сайта сплайсинга	9.62	7.96	6.94	0.872	0.070
Потеря функции	35.46	30.85	26.26	0.851	0.001

**Таблица 1.** Мутационная нагрузка минорных аллелей разных типов в африканской популяции *D. melanogaster* DPGP3. Анализ проводили для редких минорных аллелей (встречающихся не более чем в 5 особях из 191 особи в популяции DPGP3) и для всех минорных аллелей (встречающихся менее чем в 50% особей в данной популяции). Значимость понижения дисперсии по отношению к аддитивной дисперсии для вредных аллелей (нонсенс-аллелей, аллелей, вызывающих поломку сайтов сплайсинга, и аллелей, вызывающих потерю функции гена) оценивали путем сравнения значения  $\sigma^2/V_A$  для данной категории вредных аллелей со значениями  $\sigma^2/V_A$  в 1000 случайных выборок синонимических аллелей с таким же распределением популяционных частот. Односторонние Р-значения для вредных аллелей рассчитывали как долю 1000 случайных выборок синонимических аллелей, в которых значение  $\sigma^2/V_A$  было ниже или равно значению  $\sigma^2/V_A$  для соответствующей категории вредных аллелей.

## DPGP3

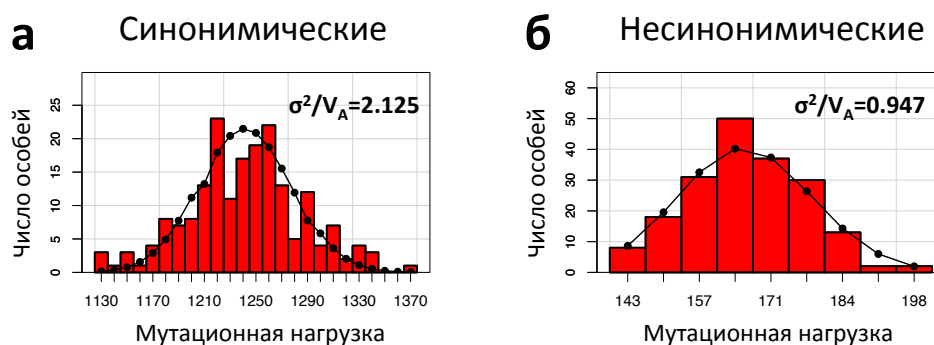


**Рисунок 4.** Распределение значений  $\sigma^2/V_A$  в 1000 случайных контрольных выборках синонимических (синий) и несинонимических (зеленый) минорных аллелей с популяционными частотами, соответствующими популяционным частотам нонсенс-аллелей (а) или аллелей, вызывающих потерю функции гена (б), полученное для популяции африканских мух DPGP3. Значение  $\sigma^2/V_A$  для нонсенс-аллелей (а) или аллелей, вызывающих потерю функции гена (б), отмечено красной вертикальной линией. Одностороннее Р-значение рассчитывали как долю 1000 случайных выборок синонимических аллелей, в которых значение  $\sigma^2/V_A$  было ниже или равно значению  $\sigma^2/V_A$  для нонсенс-аллелей (а) или аллелей, вызывающих потерю функции гена (б). В анализ включены редкие минорные аллели (встречающиеся не более чем в 5 особях из 191).

В североамериканской популяции *D. melanogaster* (DGRP) наблюдается сверхдисперсия ( $\sigma^2/V_A > 1$ ) для аллелей всех типов. Одним из вероятных факторов систематического повышения дисперсии может являться наличие популяционной структуры в данных. Поскольку ранее было показано, что в случае *D. melanogaster* значительная часть популяционной структуры в данных определяется присутствием полиморфных инверсий, мы повторили анализ, исключив из рассмотрения области генома, попадающие в участки известных инверсионных полиморфизмов *D. melanogaster*. После исключения областей инверсионных полиморфизмов сигнал синергического эпистаза ( $\sigma^2/V_A < 1$ ) для вредных аллелей был выявлен и в североамериканской популяции *D. melanogaster*.

Распределение мутационной нагрузки для несинонимических аллелей в обеих рассматриваемых популяциях характеризовалось сверхдисперсией, однако повышение дисперсии для несинонимических аллелей было выражено значительно слабее, чем для синонимических. Мы предположили, что более низкие значения  $\sigma^2/V_A$  для несинонимических аллелей связаны с действием отбора. Учитывая это, мы проанализировали распределение мутационной нагрузки для аллелей, попадающих в «необходимые» гены *D. melanogaster*, исходя из того, что несинонимические мутации в таких генах, вероятно, будут обогащены мутациями под сильным отрицательным отбором. Мы показали, что в популяции DPGP3 дисперсия распределения числа несинонимических аллелей, попадающих в «необходимые» гены, ниже аддитивной дисперсии

( $\sigma^2/V_A = 0.947$ , Рисунок 5). В то же время для синонимических аллелей в «необходимых» генах в популяции DPGP3 по-прежнему наблюдается сверхдисперсия ( $\sigma^2/V_A = 2.125$ , Рисунок 5). Понижение  $\sigma^2$  для распределения числа несинонимических аллелей, попадающих в «необходимые» гены, оказалось значимо как при сопоставлении с выборками синонимических аллелей из любого участка генома ( $P = 0.002$ ), так и при сопоставлении с выборками синонимических аллелей, попадающих в «необходимые» гены ( $P < 10^{-3}$ ).



**Рисунок 5.** Мутационная нагрузка в «необходимых» генах в популяции африканских мух DPGP3. Мутационную нагрузку рассчитывали на основе всех минорных аллелей (популяционная частота до 50%), попадающих в гены, проаннотированные как «необходимые» согласно базе данных DEG. Мутационная нагрузка синонимических (а) и несинонимических (б) аллелей в «необходимых» генах *D. melanogaster* в популяции африканских мух DPGP3. Наблюдаемое распределение числа аллелей на геном (красный) и ожидаемое в случае независимости аллелей в разных локусах (черная линия), соответствующая распределению Пуассона с таким же средним, как в данных).

Затем мы провели более общий анализ, изучив, как дисперсия распределения мутационной нагрузки для синонимических и несинонимических аллелей зависит от силы отрицательного отбора, действующего на ген. Для этого мы разделили гены в соответствии со значением отношения скоростей несинонимической и синонимической эволюции ( $dN/dS$ ), определенным для подгруппы *melanogaster* рода *Drosophila*, на 5 групп одинакового размера. В популяции DPGP3 дисперсия распределения мутационной нагрузки для несинонимических аллелей, попадающих в гены, белковая последовательность которых находится под наиболее сильным отрицательным отбором, ниже аддитивной дисперсии ( $\sigma^2/V_A = 0.98$ ). Этот результат находится в соответствии с пониженной дисперсией для несинонимических аллелей, попадающих в «необходимые» гены, в этой популяции ( $\sigma^2/V_A = 0.947$ ). Сравнение значений  $\sigma^2/V_A$  для сопоставленных по популяционным частотам выборок аллелей из генов с разным отношением скоростей несинонимической и синонимической эволюции показало, что в случае несинонимических аллелей  $\sigma^2/V_A$  падает с уменьшением  $dN/dS$ , то есть с увеличением силы отрицательного отбора на ген в обеих популяциях. В то же время в случае синонимических аллелей подобная зависимость отсутствует.

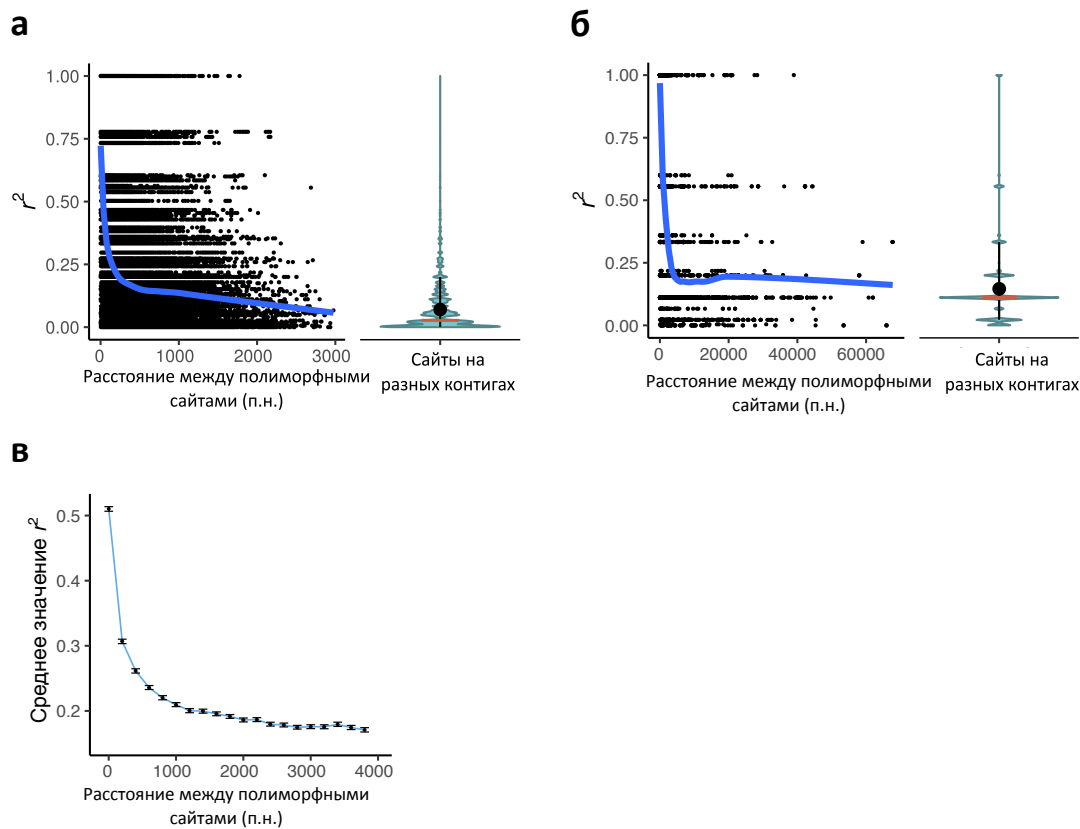
Данный результат свидетельствует об отборной природе понижения дисперсии для несинонимических аллелей в генах, находящихся под действием сильного отрицательного отбора.

Таким образом, присутствие подписи синергических эпистатических взаимодействий у *D. melanogaster* было показано в данной главе для аллелей, вызывающих потерю функции гена, а также для подмножества несинонимических аллелей, попадающих в гены, находящиеся под сильным отрицательным отбором. В отличие от традиционных подходов к поиску эпистатических взаимодействий, примененная нами методика не требует экспериментального измерения приспособленности, что значительно расширяет множество организмов, для которых она может быть использована.

В главе 4 представлены результаты анализа популяционной изменчивости для бделлоидной коловратки вида *A. vaga*. Бделлоидных коловраток ранее рассматривали как наиболее яркий пример группы предположительно древних бесполок видов. Однако до последнего времени отсутствовали исследования, в которых для видов из этой группы на уровне полногеномной популяционной изменчивости были бы показаны подписи бесполого размножения или найдены свидетельства в пользу рекомбинации и обмена генетическим материалом. В этой главе мы проанализировали последовательности геномов 11 особей (L1-L11), принадлежащих к виду бделлоидных коловраток *A. vaga*, и показали, что структура популяционной изменчивости *A. vaga* несовместима с облигатным бесполом размножением.

Анализ однонуклеотидных полиморфизмов, найденных среди особей L1-L11, выявил присутствие двух генетических кластеров: L1-L3 (маленький кластер) и L4-L11 (большой кластер). На первом этапе мы провели анализ зависимости неравновесия по сцеплению (далее коротко – LD от англ. термина linkage disequilibrium) от физического расстояния между сайтами в большом кластере. Для этого была проведена локальная реконструкция последовательностей гаплотипов (фазирование). В противоположность тому, что бы ожидалось в отсутствие рекомбинации, оказалось, что в участках генома *A. vaga*, в которых удалось восстановить гаплотипы для всех восьми коловраток из большого кластера (L4-L11), LD быстро убывает с увеличением расстояния между полиморфными сайтами (Рисунок 6а).

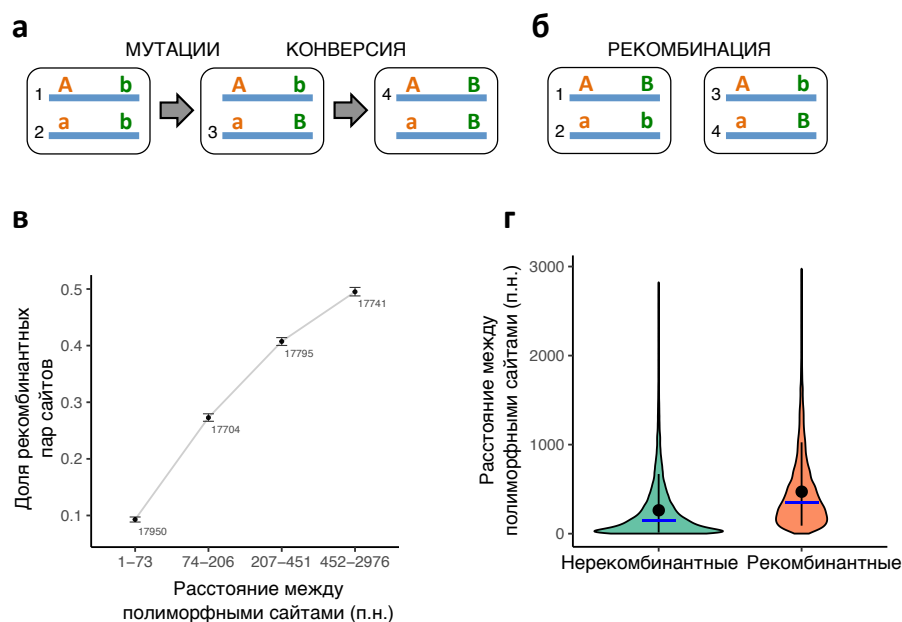
Учитывая то, что падение LD с расстоянием может быть связано с ошибками фазирования, мы оценили частоту ошибочного фазирования и показали, что распад LD у *A. vaga* не может быть объяснен ошибками реконструкции гаплотипов. Кроме того, распад LD был выявлен и на основании нефазированных данных по генотипам L4-L1 (Рисунок 6б и Рисунок 6в).



**Рисунок 6.** Распад неравновесия по сцеплению (LD) с физическим расстоянием у *A. vaga*. **а**, **б** LD выражено как  $r^2$ . Распад  $r^2$  с физическим расстоянием оценивали с использованием фазированных данных (восстановленных гаплотипов, **а**) и нефазированных данных (генотипов, **б**). Синие кривые показывают аппроксимацию падения  $r^2$  с физическим расстоянием с помощью локальной регрессии (LOESS) полиномами второй степени (с коэффициентом сглаживания 0.4). Скрипичные графики показывают распределение значений  $r^2$  для пар полиморфных сайтов на разных контигах. Концы усов соответствуют 10-й и 90-й перцентиллям. Среднее значение и медиана показаны с помощью черной точки и красной горизонтальной линии соответственно. **(а)** Значения  $r^2$  рассчитывали для биаллельных сайтов, находящихся внутри сегментов референтного генома, в которых гаплотипы были восстановлены для всех индивидуумов большого кластера (L4-L11). **(б)** Оценки  $r^2$ , полученные с использованием нефазированных данных (генотипов). Значения  $r^2$  рассчитывали для биаллельных сайтов гомозиготных в геномах всех индивидуумов большого кластера. **(в)** LD выражено как квадрат коэффициента корреляции между генотипами. Квадрат коэффициента корреляции рассчитывали между 10,000 случайно выбранными биаллельными сайтами и всеми остальными биаллельными сайтами (с частотой минорного варианта среди L4-L11  $\geq 4$ ). Пары сайтов разделяли на бины по расстоянию, разделяющему сайты, с шагом 200 п.н.; на графике представлены данные для пар сайтов, находящихся на расстоянии  $\leq 4000$  п.н. Чёрные точки показывают среднее значение квадрата коэффициента корреляции для разных бинов (координаты по оси X соответствуют нижним границам бинов). Планки погрешности соответствуют 95% доверительным интервалам, рассчитанным на основе 1000 бутстреп-реплик.

Необходимо отметить, что даже в отсутствие ошибок реконструкции гаплотипов, сигнал распада LD может возникать без реципрокной рекомбинации в результате действия генной конверсии на аллельные участки генома. Чтобы систематически проверить, может ли распределение полиморфных вариантов по гаплотипам в популяции *A. vaga* быть объяснено

исключительно действием генной конверсии, мы применили модифицированный вариант четырехгаметного теста Хадсона. При проведении данного теста мы исходили из того, что действие генной конверсии может превратить гетерозиготный генотип в гомозиготный, но не наоборот. Таким образом, пара локусов одновременно гетерозиготных в двух особях и представленных всеми четырьмя возможными гаплотипами в этих двух особях не может появиться исключительно в результате генной конверсии (Рисунок 7а и 7б). В то же время такая пара локусов может очевидным образом появиться в результате реципрокной рекомбинации. Мы показали, что среди пар сайтов одновременно гетерозиготных в двух особях доля сайтов, представленных в этих особях четырьмя возможными гаплотипами, растет с увеличением физического расстояния между сайтами (Рисунок 7в–г). Данное наблюдение несовместимо с действием генной конверсии как единственной причиной распада LD с расстоянием и, вероятно, указывает на реципрокную рекомбинацию у *A. vaga*.



**Рисунок 7.** Результаты модифицированного четырехгаметного теста указывают на вероятное существование реципрокной рекомбинации у *A. vaga*. (а) Возникновение четырех гаплотипов для пары биаллельных сайтов без реципрокной рекомбинации и обмена генетическим материалом за счет мутаций и конверсии. Прямоугольники символизируют особей. (б) Схематическое изображение рекомбинантной пары сайтов, которая не могла возникнуть исключительно за счет генной конверсии: пара полиморфных сайтов представлена всеми четырьмя возможными гаплотипами в двух особях. Такие пары сайтов проходят модифицированный четырехгаметный тест. (в) Зависимость доли пар сайтов, проходящих модифицированный четырехгаметный тест, от физического расстояния между сайтами в паре. Для каждого попарного сравнения особей L4-L11 рассматривали только те пары полиморфных сайтов, которые одновременно гетерозиготны в этих двух особях. Пары сайтов, удовлетворяющие условиям модифицированного четырехгаметного теста, были разделены на 4 бина по расстоянию между сайтами в паре с примерно равным количеством наблюдений в разных бинах. Черные точки обозначают доли рекомбинантных пар сайтов для разных бинов. Общее число проанализированных пар сайтов для каждого бина указано

рядом с соответствующей точкой. Планки погрешности соответствуют 95% доверительным интервалам, рассчитанным на основе 1000 бутстреп-реплик для каждого бина. Разница в долях рекомбинантных пар сайтов значима для всех попарных сравнений между бинами (двустороннее Р-значение  $< 6 \times 10^{-4}$ , пермутационный тест). (г) Скрипичные графики показывают распределение расстояний между сайтами в паре для нерекомбинантных ( $n = 48,657$ ) и рекомбинантных пар сайтов ( $n = 22,533$ ). Концы усов соответствуют 10-й и 90-й перцентилям. Среднее значение и медиана для каждой группы показаны с помощью черной точки и синей горизонтальной линии соответственно. Для нерекомбинантных пар сайтов среднее расстояние между сайтами в паре 261.7, медиана 146 п.н.; для рекомбинантных пар сайтов среднее расстояние между сайтами в паре 470.6, медиана 353 п.н. Разница в средних расстояниях между сайтами для нерекомбинантных и рекомбинантных пар значима (двустороннее Р-значение  $< 1 \times 10^{-4}$ , пермутационный тест).

Затем мы проанализировали изменчивость среди индивидуумов большого кластера, L4-L11, на уровне отдельных биаллельных сайтов. При облигатном бесполом размножении два аллеля в одном локусе накапливают мутации независимо друг от друга. В популяции конечного размера это создает избыток гетерозигот по сравнению с ожидаемым числом гетерозигот при равновесии Харди-Вайнберга, что соответствует отрицательным значениям коэффициента инбридинга  $F_{IS}$ . В то время как ожидаемое значение  $F_{IS}$  при равновесии Харди-Вайнберга равно 0, ожидаемое значение  $F_{IS}$  при облигатном бесполом размножении  $-1$ . Мы проанализировали распределение значений  $F_{IS}$  в сайтах биаллельных среди L4-L11 и обнаружили, что это распределение сосредоточено вокруг 0 (медиана = 0.0, среднее значение =  $-0.03$ ), что, вероятно, указывает на то, что рассматриваемая популяция *A. vaga* находится близко к равновесию Харди-Вайнберга.

Независимое свидетельство в пользу обмена генетическим материалом между индивидуумами *A. vaga* было получено в ходе анализа трехаллельных сайтов. Трехаллельные сайты могут возникать у видов, размножающихся как бесполом, так и половым путем, за счет множественных мутаций, затрагивающих один и тот же сайт. В гипотетическом полиморфном сайте с тремя аллелями А, В и С в принципе возможны три различных гетерозиготных генотипа: А/В, А/С и В/С. К появлению таких сайтов может приводить обмен генетическим материалом. В бесполой популяции все три гетерозиготных генотипа могут возникнуть только за счет рекуррентных или обратных мутаций, которые происходят относительно редко. Мы оценили, что в случае отсутствия обмена генетическим материалом ожидаемое число трехаллельных сайтов, одновременно представленных в L4-L11 тремя гетерозиготными генотипами, составляет 83.5. Однако в данных по полиморфизму L4-L11 мы идентифицировали 1839 таких сайтов (Р-значение  $< 2.2 \times 10^{-16}$ , одновыборочный Z-тест), что соответствует 22-кратному превышению над ожиданием. По всей видимости, это наблюдение указывает на существование обмена генетическим материалом в популяции *A. vaga*.

Ранее в литературе в качестве возможного механизма обмена генетическим материалом у бделлоидных коловраток рассматривались половое размножение с классическим или атипичным (по типу *Oenothera*) мейозом и

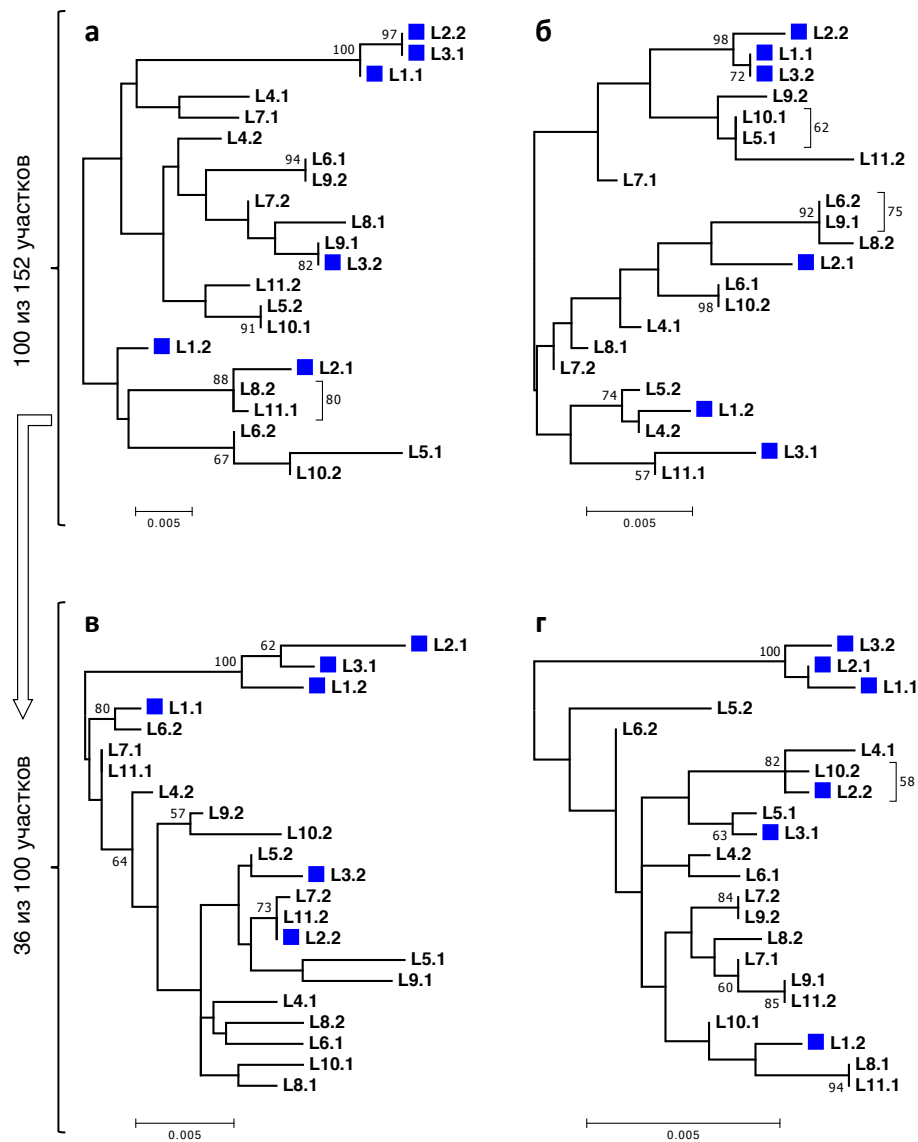


горизонтальный перенос генов. Для того, чтобы попробовать ответить на вопрос о том, какой механизм обмена генетическим материалом более вероятен у *A. vaga*, мы сравнили филогении для двух гаплотипов каждого индивидуума из большого кластера в разных геномных локусах. Этот анализ выявил значительное количество случаев, в которых два гаплотипа одного и того же индивидуума кластеризовались с гаплотипами из разных индивидуумов. При этом в разных геномных локусах для двух гаплотипов одного и того же индивидуума наблюдались различные паттерны инконгруэнтности. Этот результат является дополнительным свидетельством в пользу генетического обмена у *A. vaga* и делает атипичный мейоз маловероятным объяснением генетического обмена у бделлоидных коловраток.

Затем мы провели расширенный анализ филогений гаплотипов, включив в него все 11 особей (L1-L11). Этот анализ выявил, что в большинстве случаев L1, L2 и L3 кластеризовались по одному из двух гаплотипов, но не по второму. Как правило, три гаплотипа (по одному гаплотипу каждого из индивидуумов L1-L3) формировали кладу с хорошей поддержкой, в то время как три оставшихся гаплотипа из L1-L3 были перемешаны с гаплотипами L4-L11 (Рисунок 8а-г). Кроме того, в значительном числе случаев укоренение посередине разделяло дерево на монофилетическую группу, содержащую три гаплотипа из L1, L2 и L3, и остальные гаплотипы (Рисунок 8в-г). В соответствии с этими результатами в большинстве проанализированных сегментов дерева, построенное методом максимального правдоподобия без наложения ограничений на топологию, статистически значимо отличалось от дерева, для которого было задано ограничение на то, чтобы шесть гаплотипов, принадлежащие трем индивидуумам из маленького кластера, формировали монофилетическую группу ( $P$ -значение  $< 0.05$  после поправки Бонферрони). Эти результаты было бы сложно объяснить в рамках механизма горизонтального переноса генов. В то же время они совместимы с классическим мейозом, если мы предположим, что три индивидуума из маленького кластера имеют гибридное происхождение и являются потомством, появившимся в результате скрещивания особей из популяции, генетически близкой к популяции большого кластера, и другой популяции, находящейся на большем генетическом расстоянии от популяции большого кластера. Анализ митохондриальной изменчивости указывает на то, что для того, чтобы объяснить возникновение особей L1-L3, требуется предположить более одного события полового размножения.

Результаты, представленные в данной главе, указывают на то, что изменчивость в природной популяции *A. vaga* значительно рандомизирована и несовместима со строго бесполом способом размножения и отсутствием рекомбинации. Необходимо отметить, что для того, чтобы ответить на вопрос о механизме обмена генетическим материалом у бделлоидных коловраток, нужны дополнительные исследования. Однако некоторые выявленные закономерности позволяют считать, что половое размножение является более вероятным механизмом обмена генетическим материалом у *A. vaga*, чем горизонтальный перенос генов.





**Рисунок 8.** Филогенетический анализ гаплотипов особей L1-L11 намекает на гибридное происхождение особей из маленького кластера. Укорененные посередине филогенетические деревья, построенные для четырех сегментов, в которых фазы полиморфизмов удалось восстановить одновременно для 11 особей из обоих кластеров (L1-L11), и, предположительно, указывающие на гибридное происхождение индивидуумов из маленького кластера. Гаплотипы особей из маленького кластера показаны с помощью синих квадратов. В 100 из 152 сегментов была найдена монофилетическая группа, в состав которой входило три гаплотипа – по одному гаплотипу каждой особи L1, L2 и L3 (как в панелях а–г). Среди этих 100 сегментов в 36 корень, помещенный посередине, отделял монофилетическую группу, включающую три гаплотипа L1, L2 и L3, от всех остальных гаплотипов (как в панелях в и г). Филогенетические деревья, показанные на данном рисунке, были построены для четырех фазированных сегментов, находящихся на разных контигах в диплоидной сборке L1. Длина сегментов составляет 1058 (а), 1452 (б), 1782 (в) и 1824 (г) п.н. Филогенетические деревья были построены с применением метода максимального правдоподобия, реализованного в программе PhyML, с использованием модели GTR+G (1000 бутстреп-реplik). Значения бутстреп-поддержки  $\geq 50\%$  показаны рядом с соответствующими ветвями или рядом со скобками. Значения бутстреп-поддержки округлены до ближайшего целого числа. Для построения филогенетических деревьев были

использованы только те полиморфные сайты, которые прошли все шаги фильтрации и были одновременно фазированы у особей L1-L11 ( $n = 46, 32, 43$  и  $30$  полиморфных сайтов с частотой минорного варианта  $\geq 2$  в панелях **a–г** соответственно); остальные сайты рассматривались как мономорфные. Индексы 1 и 2 обозначают два гаплотипа одной особи.

## Заключение

В данной работе был проведен поиск сигнала отрицательного отбора и рекомбинации на различных геномных данных. Анализ ортологичных интронов в далеких парах видов выявил сигнал отрицательного отбора, продолжающего действовать на имеющие общее происхождение некодирующие участки генома после того, как их последовательности разошлись до неузнаваемости. Так, интроны, несущие сегмент значимого локального сходства или регуляторный элемент в одной паре видов, с повышенной частотой также несут сегмент локального сходства во второй паре видов, несмотря на то, что выравнивание между видами из разных пар отсутствует. Кроме того, последовательность интронов, сохранившихся в одной паре видов, более консервативна во второй паре видов, несмотря на отсутствие сквозного выравнивания между интронами из разных пар видов. Мы предполагаем, что эти результаты могут быть объяснены сохранением предковой функции разошедшихся до неузнаваемости некодирующих последовательностей. Анализ, выполненный в данной работе, проводился на интронах, поскольку общее происхождение интронов может быть установлено даже на больших филогенетических расстояниях на основании выравнивания фланкирующих экзонов. Вероятно, что явление сохранения функции без сохранения сходства последовательностей характерно не только для интронов, но и для межгенных областей генома. Изучение возможного продолжения действия отрицательного отбора на разошедшиеся до неузнаваемости ортологичные межгенные участки может стать направлением для дальнейших исследований, однако установление ортологических соответствий для межгенных участков на значительных филогенетических расстояниях требует разработки отдельного подхода.

С помощью анализа распределения мутационной нагрузки в популяциях *D. melanogaster* нами была выявлена подпись синергического эпистатического отбора, действующего на аллели, вызывающие потерю функции гена, у *D. melanogaster*. Ожидается, что в случае синергических эпистатических взаимодействий дисперсия распределения мутационной нагрузки должна быть ниже аддитивной дисперсии. Такой сигнал был выявлен для аллелей, вызывающих потерю функции гена, в двух популяциях *D. melanogaster*, а также для подмножества несинонимических аллелей, попадающих в гены, находящиеся под сильным давлением отрицательного отбора, в замбийской популяции *D. melanogaster*. Свидетельство о существовании распространенных синергических эпистатических взаимодействий между вредными мутациями является важным как с точки зрения возможного объяснения парадокса мутационного груза, так и с точки зрения поиска причин преобладания полового размножения среди эукариот. Действительно, как показано в нескольких теоретических работах, в случае существования синергического

эпистаза мутационный груз половой популяции может быть значительно ниже, чем в случае независимого влияния мутаций на приспособленность<sup>9, 10, 11</sup>. При этом, если скорость вредных мутаций на геном на поколение достаточно высока, в присутствии синергического эпистаза половое размножение может получать преимущество перед бесполом<sup>10, 11</sup>.

Несмотря на то, что точные причины преобладания полового размножения остаются неизвестными, по всей видимости, оно дает существенные преимущества, поскольку отказ от полового размножения обычно приводит к вымиранию. Это делает особенно интересным изучение возможных механизмов, с помощью которых немногочисленные древние группы бесполок видов противостоят долгосрочным последствиям отказа от полового размножения. В данной работе мы провели анализ популяционной изменчивости для вида *A. vaga*, принадлежащего к группе бделлоидных коловраток, которых в течение длительного времени рассматривали как наиболее яркий пример древнего бесполого таксона. Однако сравнение геномов разных особей *A. vaga*, проведенное в данной работе, выявило сигнал рекомбинации и обмена генетическим материалом. Распад неравновесия по сцеплению с увеличением расстояния между полиморфными сайтами указывает на существование рекомбинации у *A. vaga*. При этом результаты модифицированного четырехгаметного теста позволяют заключить, что распад неравновесия по сцеплению у *A. vaga* не может быть объяснен исключительно за счет генной конверсии. Результаты, полученные при анализе отдельных полиморфных сайтов, дают основания полагать, что в популяции бделлоидной коловратки *A. vaga* происходит обмен генетическим материалом. В частности, на обмен генетическим материалом у *A. vaga* указывает избыток над мутационным ожиданием трехаллельных сайтов, представленных всеми тремя гетерозиготными генотипами. Свидетельства в пользу обмена генетическим материалом у *A. vaga* были получены и при анализе филогений гаплотипов разных индивидуумов. Распределение гаплотипов в филогениях, построенных одновременно для особей из большого и маленького кластеров, согласуется с тем, что ожидалось бы в том случае, если бы три особи маленького кластера имели гибридное происхождение. Если наша интерпретация верна, то половое размножение представляется более вероятным объяснением обмена генетическим материалом в популяции *A. vaga*, чем горизонтальный перенос генов. Тем не менее определение механизма обмена генетическим материалом у бделлоидных коловраток безусловно заслуживает дальнейших исследований. Вне зависимости от способа, с помощью которого у бделлоидных коловраток происходит обмен генетическим материалом, по всей видимости, бделлоидные коловратки не относятся к древним группам бесполок видов: структура популяционной изменчивости у *A. vaga* несовместима с исключительно клональным наследованием ДНК. Свидетельства в пользу рекомбинации у бделлоидных коловраток, которых рассматривали как наиболее удивительный пример древней группы бесполок видов, подчеркивают важность рекомбинации для долгосрочного эволюционного успеха.

## Выводы

1. В ортологичных интронах из филогенетически далеких пар видов присутствует сигнал, свидетельствующий о продолжении действия отрицательного отбора на имеющие общее происхождение, но разошедшиеся до неузнаваемости, некодирующие последовательности. Возможное объяснение этого явления заключается в том, что предковая функция некодирующего участка генома, вероятно, может сохраняться дольше, чем сходство последовательностей.

2. Распределение мутационной нагрузки для аллелей *D. melanogaster*, вызывающих потерю функции гена, характеризуется понижением дисперсии относительно аддитивной дисперсии. Кроме того, понижение дисперсии распределения мутационной нагрузки относительно аддитивной дисперсии наблюдается и для подмножества несинонимических аллелей, попадающих в гены, находящиеся под сильным давлением отрицательного отбора, в замбийской популяции *D. melanogaster*. Данная картина соответствует тому, что ожидается в случае существования у *D. melanogaster* синергических эпистатических взаимодействий между вредными аллелями.

3. Данные по внутривидовой изменчивости бделлоидной коловратки вида *A. vaga* указывают на существование рекомбинации и обмена генетическим материалом в популяции этого вида. Подписи рекомбинации, выявленные при анализе геномов *A. vaga*, не могут быть объяснены исключительно за счет генной конверсии и, вероятно, являются результатом реципрокной рекомбинации. Структура изменчивости у *A. vaga* как на уровне отдельных полиморфных сайтов, так и на уровне геномных сегментов несовместима с исключительно клональным способом размножения. Распределение гаплотипов особей *A. vaga* из большого и маленького кластеров в филогениях, построенных для разных геномных локусов, наиболее экономно объясняется в рамках сценария, в котором обмен генетическим материалом у *A. vaga* происходит в результате полового размножения.

## Список публикаций по теме диссертации

По теме диссертации опубликовано три статьи в рецензируемых международных научных журналах, входящих в основные библиометрические базы данных (PubMed, WoS, Scopus, РИНЦ):

1. **Vakhrusheva O. A.**, Bazykin G. A., Kondrashov A. S. Genome-level analysis of selective constraint without apparent sequence conservation // *Genome Biology and Evolution*. 2013. Vol. 5, № 3. P. 532–541. doi: 10.1093/gbe/evt023.

2. Sohail M., **Vakhrusheva O. A.**, Sul J. H., Pulit S. L., Francioli L. C., Genome of the Netherlands Consortium, Alzheimer’s Disease Neuroimaging Initiative, van den Berg L. H., Veldink J. H., de Bakker P. I. W., Bazykin G. A., Kondrashov A. S., Sunyaev S. R. Negative selection in humans and fruit flies involves synergistic epistasis // *Science*. 2017. Vol. 356, № 6337. P. 539–542. doi: 10.1126/science.aah5238.

3. **Vakhrusheva O. A.**, Mnatsakanova E. A., Galimov Y. R., Neretina T. V., Gerasimov E. S., Naumenko S. A., Ozerova S. G., Zalevsky A. O., Yushenova I. A., Rodriguez F., Arkhipova I. R., Penin A. A., Logacheva M. D., Bazykin G. A., Kondrashov A. S. Genomic signatures of recombination in a natural population of the bdelloid rotifer *Adineta vaga* // *Nature Communications*. 2020. Vol. 11, № 1:6421. doi: 10.1038/s41467-020-19614-y.

Кроме того, результаты работы опубликованы в сборниках тезисов международных и российских конференций:

1. Vakhrusheva O. A., Bazykin G. A., Kondrashov A. S. Selective constraint beyond apparent sequence conservation // Информационные технологии и системы 2012 (ИТиС 2012), Петрозаводск, Россия, 19–25 августа 2012. <http://www.itas2012.iitp.ru/pdf/1569601189.pdf>

2. Vakhrusheva O. A., Mnatsakanova E. A., Galimov Y., Gerasimov E. S., Neretina T. V., Penin A. A., Logacheva M. D., Bazykin G. A., Kondrashov A. S. Population genomic data reveal signatures of genetic exchange in the bdelloid rotifer *Adineta vaga* // Информационные технологии и системы 2018 (ИТиС 2018), Казань, Россия, 25–30 сентября 2018. <http://itas2018.iitp.ru/media/papers/1570471700.pdf>

3. Vakhrusheva O. A., Mnatsakanova E. A., Galimov Y., Neretina T. V., Gerasimov E. S., Ozerova S. G., Zalevsky A. O., Yushenova I. A., Arkhipova I. R., Penin A. A., Logacheva M. D., Bazykin G. A., Kondrashov A. S. Signatures of genetic exchange in a natural population of the bdelloid rotifer *Adineta vaga* inferred from whole-genome data // Proceedings of the International Moscow Conference on Computational Molecular Biology 2019 (MCCMB’19), Moscow, Russia, July 27–30, 2019. <http://mccmb.belozersky.msu.ru/2019/thesis/MCCMB2019/abstracts/64.pdf>