

Федеральное государственное автономное  
образовательное учреждение  
высшего образования  
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Институт фундаментальной биологии и биотехнологии  
Кафедра биофизики

УТВЕРЖДАЮ  
Заведующий кафедрой  
\_\_\_\_\_ В.А. Кратасюк  
« \_\_\_\_\_ » \_\_\_\_\_ 2018 г.

**БАКАЛАВРСКАЯ РАБОТА**

06.03.01 Биология  
06.03.01.07 Биофизика

Анализ маркерных последовательностей представителей рода *Armillaria*

Научный руководитель	_____	Ю. А. Путинцева
Научный консультант	_____	д.ф.-м.н., в.н.с М. Г. Садовский
Выпускник	_____	О. С. Зуева

Красноярск 2018

## РЕФЕРАТ

Бакалаврская работа по теме «Анализ маркерных последовательностей представителей рода *Armillaria*» содержит 31 страницу текстового документа, 38 использованных источников, 4 рисунка, 1 таблицу.

### ГЕНЕТИЧЕСКИЕ МАРКЕРЫ, ФИЛОГЕНИЯ, ИДЕНТИФИКАЦИЯ ВИДОВ, МАРКЕРНЫЕ ПОСЛЕДОВАТЕЛЬНОСТИ

Цель работы - филогенетический анализ 28 образцов представителей рода *Armillaria* по маркерным последовательностям ITS, IGS, TEF1 $\alpha$  для видовой идентификации.

Объект исследования - нуклеотидные маркерные последовательности представителей рода *Armillaria*.

Предмет исследования – эволюционные связи исследуемых образцов.

Актуальность настоящего исследования обусловлена тем, что использование генетических маркеров для идентификации видов *Armillaria* на территории Сибири и Дальнего Востока проводилась впервые.

Было идентифицировано 28 сибирских и дальневосточных образцов представителей рода *Armillaria*. Филогенетический анализ позволил с высокой вероятностью определить видовую принадлежность исследуемых образцов.

## СОДЕРЖАНИЕ

РЕФЕРАТ .....	2
ВВЕДЕНИЕ .....	4
1 Обзор литературы .....	6
1.1 Актуальность исследования генетических маркеров грибов .....	6
1.2 Представители рода <i>Armillaria</i> .....	6
1.2.1 <i>Flammulina velutipes</i> эволюционно близкий вид для представителей рода <i>Armillaria</i> .....	7
1.3 Генетические маркеры для видовой идентификации грибов .....	7
1.3.1 Генетический маркер ITS .....	8
1.3.2 Генетический маркер IGS .....	8
1.3.3 Генетический маркер TEF1 $\alpha$ .....	8
1.4 Рестрикционный анализ .....	8
1.5 Строение участка рибосомальной РНК грибов .....	9
1.6 Выравнивание нуклеотидных последовательностей .....	11
1.6.1 Выравнивание двух последовательностей .....	11
1.6.2 Выравнивание нескольких последовательностей .....	12
1.6.3 Обоснование гэпов при оценке эволюционных дистанций .....	12
1.8 Программы для построения выравниваний .....	13
1.9 Филогенетический анализ .....	15
2 Материалы и методы .....	18
2.1 Объекты исследования .....	19
2.2 Методы используемые для филогенетического анализа .....	20
3 Результаты и их обсуждение .....	22
ЗАКЛЮЧЕНИЕ .....	26
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ .....	27

## ВВЕДЕНИЕ

Среди представителей рода *Armillaria* (*Physalacriaceae*, *Agaricales*, *Basidiomycota*) хорошо известны как важные патогены растений, которые могут вызывать серьезные корневые заболевания у разных деревьев и древесных растений, так и сапротрофы, которые играют важную экологическую роль в экосистемах. Однако у разновидностей *Armillaria* наблюдаются значительные различия в патогенности. Например, *A. ostoyae*, *A. solidipes* и *A. mellea* как правило, считаются сильно патогенными, в то время как такие виды, как *A. sinapina*, *A. cepistipes* считаются менее патогенными, а также сапротрофы, предпочитающие рост на мертвой древесине (*A. gallica*). Поэтому разграничение видов *Armillaria* имеет решающее значение для оценки риска заболевания [1, 2]. Грибы рода *Armillaria* играют колоссальную роль в экосистеме, без них леса были бы просто захламлены слаборазлагающейся древесиной. Типичные сапрофиты не влияют сильно пагубно на здоровые деревья, но в настоящее время все чаще и чаще начинают поражать здоровые деревья. *A. borealis* является сапрофитом и в меньшей степени патогеном, а вот с изменением климата возможно приобретает патогенные формы. Понимание филогении и эволюционного происхождения этих видов является важным этапом в изучении патогенности видов *Armillaria*.

28 образцов рода *Armillaria* были взяты на Дальнем Востоке, юге Красноярского края и Республике Хакасия. В Германии сотрудниками лаборатории лесной геномики Сибирского федерального университета были секвенированы маркерные последовательности на капиллярном секвенаторе. Для построения филогенетического дерева также были включены нуклеотидные последовательности представителей рода *Armillaria*, взятые из международной базы данных GenBank ([www.ncbi.nlm.nih.gov/genbank/](http://www.ncbi.nlm.nih.gov/genbank/)).

Целью данного исследования является филогенетический анализ 28 образцов представителей рода *Armillaria* по маркерным последовательностям ITS, IGS, TEF1 $\alpha$  для видовой идентификации.

Для достижения данной цели были поставлены следующие задачи:

1. Сформировать базу генетических последовательностей;
2. Изучить методы построения выравниваний и филогении;
3. Построение филогенетического дерева для видовой идентификации образцов.

Актуальность данной работы заключается в том, что определение видов грибов только по морфологическим признакам не всегда возможно, поскольку они не всегда могут обеспечивать точные группировки на уровне видов, поэтому сейчас широко используются генетические маркеры для идентификации видов. Они позволяют изучать высоко- и слабопатогенные таксоны грибов и осуществлять генетический мониторинг уровня патогенности в популяциях. Использование генетических маркеров для идентификации видов *Armillaria* на территории Сибири и Дальнего Востока проводилось впервые.

## 1 Обзор литературы

### 1.1 Актуальность исследования генетических маркеров грибов

Молекулярно-генетическая идентификация особенно важна в тех случаях, когда определение видов грибов по морфологическим признакам не всегда возможно. Кроме того, возможно выявление новых видов грибов, имеющих сходные морфологические признаки с уже известными видами.

Определение сибирских и дальневосточных видов *Armillaria* с использованием генетических маркеров ранее не проводилось. Также на сегодняшний день филогения сибирских и дальневосточных штаммов совершенно не изучена.

### 1.2 Представители рода *Armillaria*

Грибы образуют одно из самых больших и самых разнообразных царств живых организмов. Установление родственных связей между организмами исходит из филогенетического анализа, основанных на генной информации, построенных обычно на основе степени различий в белках или последовательностях ДНК небольшого числа высококонсервативных генов методом множественного выравнивания последовательностей (MSA).

*Armillaria* (Опёнок) – это широко распространенный род грибов, состоящий в основном из патогенов растений [3]. Представители данного рода играют важную роль в экосистеме и являются производителями ризоморфов или вегетативных структур, которые при их обнаружении часто ассоциируются с инфекцией. Из-за роли этих видов грибов как возбудителей болезней растений, важным является понимание филогении и эволюционного происхождения этих видов [4].

Изучение видов рода *Armillaria* важно как с экономической, так и с экологической точки зрения. Некоторые виды данного рода являются одними из самых серьезных патогенов бореальных и умеренных лесов и атакуют сотни видов растений в обоих полушариях при разных климатических

условиях, в то время как другие виды являются сапрофитами и играют важную роль в экосистемах. Видовое разнообразие данного рода представляет большой интерес для понимания эволюционных процессов и того, как его представители получили распространение во всем мире. Одним из инструментов для описания межвидовых отношений является филогенетика. В современной молекулярной биологии подавляющее большинство филогенетических исследований основывается на сравнении нуклеотидных последовательностей. Основным инструментом здесь – построение так называемых филогенетических деревьев. При этом построение таких деревьев проводится сравнением не полных геномов, а лишь определенного набора последовательностей [5]

### **1.2.1 *Flammulina velutipes* эволюционно близкий вид для представителей рода *Armillaria***

К грибам-паразитам, поселяющимся на древесине и очень быстро ее разрушающим, относятся грибы семейства *Physalacriaceae*. Представители рода *Armillaria* и *F. velutipes* хорошо известны как патогены растений. *Flammulina velutipes* – близкий вид для представителей рода *Armillaria* может применяться при анализе родственных связей.

Как типичный гриб белой гнили, *F. velutipes* можно найти на мертвой древесине. Его распространение ограничено умеренными климатическими зонами, потому что для плодоношения требуется холодный период [6, 7].

### **1.3 Генетические маркеры для видовой идентификации грибов**

Маркеры – определенные участки гена или любой другой участок хромосомы. Генетический маркер представляет собой фрагмент ДНК, для которого известна точная хромосомная локализация.

Генетические маркеры грибов успешно применяются в эволюционных и филогенетических исследованиях [8].

### **1.3.1 Генетический маркер ITS**

ITS (Internal transcribed spacer) – внутренний транскрибированный спейсер ядерной ДНК хорошо подходит для идентификации грибов как отдельных, так и смешанных таксонов. Дает информацию об относительной численности таксонов, выявляет разнообразие, а также дает наивысшую вероятность правильной идентификации для широкого спектра анализируемых грибов. ITS могут быть использованы для филогенетического анализа только родственных таксонов [9, 10].

Последовательности ITS высоко вариабельны и позволяют легко отличать близкородственные организмы. Высокая степень варьирования обусловлена тем, что данные последовательности являются некодирующими, следовательно, могли не подвергаться давлению отбора, что очень важно для филогенетических исследований [11].

### **1.3.2 Генетический маркер IGS**

IGS (Intergenic spacer) – межгенный спейсер, является некодирующей областью ДНК, расположенной между генами. Этот термин используется, в частности, для спейсеров между многими тандемно-повторяющимися копиями генов рРНК [12].

### **1.3.3 Генетические маркер TEF1 $\alpha$**

TEF1 $\alpha$  (translation elongation factor1 $\alpha$ ) – отвечает за 2 основных процесса, происходящих при синтезе белка на рибосоме. Он несет ответственность за выбор и связывание родственной тРНК. TEF1 $\alpha$  является весьма полезным для филогенетического анализа именно рода *Armillaria* [13].

## **1.4 Рестрикционный анализ**



Рестрикционный анализ ДНК широко используется в молекулярно-биологических исследованиях и прикладных работах и является одним из наиболее важных инструментов при изучении ДНК.

Рестрикционный анализ амплифицированной области внутренних транскрибируемых спейсеров, включающей ген 5,8s малой субъединицы рДНК (ITS1-5,8s-ITS2) широко применяется для идентификации видов грибов. Он основан на способности эндонуклеаз рестрикции расщеплять ДНК в определенных сайтах. Рестриктазы эндонуклеазы расщепляют нуклеиновые кислоты не с конца молекулы, а в середине. При этом каждая рестриктаза узнаёт определённый участок ДНК длиной от четырёх пар нуклеотидов и расщепляет нуклеотидную цепь внутри участка распознавания или вне него.

Полученные с помощью рестриктаз фрагменты ДНК служат генетическими маркерами [14, 15].

### **1.5 Строение участка рибосомальной РНК грибов**

В последние годы имеют довольно широкое применение методы, основанные на использовании фрагментов ДНК в качестве молекулярно-генетических маркеров. Большинство филогенетических анализов грибов сделано на основе кластера рибосомальных генов, обнаруженных в ядерных и митохондриальных геномах.

Кластер ядерных генов рРНК представлен структурными генами большой и малой субъединиц рРНК (18s и 28s), внутренними транскрибируемыми спейсерами (ITS) и межгенными спейсерами (IGS). Между двумя этими внутренними спейсерами ITS1 и ITS2, разделяющими последовательности малой и большой субъединиц рРНК, находится ген 5,8s. Протяженная межгенная последовательность IGS разделяет единицы транскрипции кластера генов рРНК. Так как накопление изменений в нуклеотидных последовательностях структурных генов в участках внутренних и межгенных спейсеров происходит с разной скоростью, их

можно использовать при разделении таксонов. Структурные гены большой и малой субъединиц более консервативны, поэтому их последовательности используют для установления родственных связей между организмами.

Маркерные гены 18S, 5.8S и 28S рРНК высококонсервативны и это позволяет изолировать ITS и IGS последовательности, которые располагаются между кодирующими регионами, амплифицировать с помощью ПЦР и выявлять нуклеотидные последовательности уникальные для гриба [16].

Для филогенетических исследований грибов используют шесть универсальных генных областей, консервативных и переменных в достаточной мере для межвидового разграничения: гены малой и большой субъединиц рРНК и межгенные спейсерные участки (18S рРНК, 28S рРНК, 5.8S рРНК, IGS, ITS), фактор элонгации (TEF1 $\alpha$ ). Наиболее часто используемыми областями в таксономических и филогенетических исследованиях для межвидовых и внутривидовых сравнений являются внутренние транскрибируемые спейсерные области кластера генов рРНК ITS1 и ITS2, фланкированные генами 18S рРНК и 28S рРНК, и геном 5.8S рРНК, прерывающим ITS последовательность. Структура рибосомального кластера генов в области ITS показана на рисунке 1. Именно по этим генным последовательностям накоплены большие информационные базы данных в GenBank ([www.ncbi.nlm.nih.gov/genbank/](http://www.ncbi.nlm.nih.gov/genbank/)).



Рисунок 1 - Схема области кластера генов рРНК с участком локализации ITS [17, 18].

Более того, в связи с малым числом различий в генах 18S рРНК среди родственных видов грибов идентификация обычно ограничивается на уровне

рода или семейства. Для того чтобы увеличить уровень идентификации, используют внутренние транскрибируемые спейсеры (ITS), которые находятся между 18S рРНК и 28S рРНК.

Самым распространенным маркерным геном грибов является ген 18s рРНК. Ген 18S-рРНК находится в малой субъединице рРНК грибов. Он несет как консервативные, так и переменные участки нуклеотидной последовательности, что позволяет использовать его как для определения рода, так и видовой идентификации микроорганизмов. Консервативные участки идентичны у всех грибов и используются для отжига праймеров при ПЦР, а переменные или видоспецифичные участки используются для идентификации грибов [19].

## **1.6 Выравнивание нуклеотидных последовательностей**

Филогенетический анализ устанавливает взаимосвязи между участками генома (например, генами или их фрагментами). Для этого важно, чтобы гомологичные сайты сравнивались друг с другом.

С биологической точки зрения, выравнивание — это гипотеза о гомологичности множества нуклеотидных оснований в последовательности ДНК или аминокислот в последовательности белка.

Существует несколько методов для выравнивания нуклеотидных последовательностей.

### **1.6.1 Выравнивание двух последовательностей**

Метод парного выравнивания последовательностей был предложен Нидлманом и Вуншем в 1970 году. Выравнивание парной последовательности используется между двумя последовательностями для идентификации индекса сходства, и выбирается такое выравнивание двух последовательностей, при котором индекс сходства будет максимальным. Позже Селлерс предложил метод, в котором измеряется коэффициент дистанции между двумя последовательностями и выбирается выравнивание,

которое минимизирует эту дистанцию. Однако, эти два метода практически одинаковы и дают в большинстве случаев одинаковые результаты, которые могут указывать на функциональные, структурные и эволюционные связи между двумя последовательностями.

### **1.6.2 Выравнивание нескольких последовательностей**

Множественное выравнивание представляет собой выравнивание трех и более последовательностей одинаковой длины, также его еще называют прогрессивным множественным выравниванием. В этом алгоритме сначала выравниваются пары последовательностей с маленькими расстояниями, и выравнивание наиболее удаленных последовательностей делается постепенно. Все последовательности попарно выравниваются по вышеописанному механизму. Далее выравниваются группы последовательностей друг с другом, по алгоритму, который схож с выравниванием двух последовательностей, за исключением того, что средние дистанции теперь считаются, рассматривая все нуклеотиды в каждом положении двух групп последовательностей. Из результатов можно вывести гомологию и эволюционные соотношения между изученными последовательностями. Множественное выравнивание последовательностей ДНК, РНК и белков является важным методом в молекулярной биологии, включая филогенетическую оценку деревьев. Объединение алгоритмов множественного выравнивания позволяет улучшить скорость, качество и возможности множественного выравнивания для обработки большого количества последовательностей [20].

### **1.6.3 Обоснование гэпов при оценке эволюционных дистанций**

При выравнивании нуклеотидных последовательностей присутствие гэпов вносит некоторые сложности при оценке эволюционных дистанций. Более того, в последовательностях могут присутствовать сайты, последовательность которых неизвестна из-за неотсеквенированных

участков, создавая такие же проблемы, как и гэпы. Эти сайты обычно выбрасываются из рассмотрения при оценке дистанций двумя способами:

Первый способ – вариант полного удаления гэпов. Исключение этих сайтов из анализа. И лучше всего использовать этот способ так как разные области ДНК часто эволюционируют по-разному.

Второй способ – это вариант попарного удаления гэпов. В случае, когда число нуклеотидов в гэпах мало, и гэпы распределены случайным образом, то можно рассчитать дистанцию между каждой парой последовательностей игнорируя только гэпы, имеющиеся в двух сравниваемых последовательностях.

## 1.7 Программы для построения выравниваний

Современное программное обеспечение для построения множественного выравнивания:

- **ClustalW** – до сих пор считается самым популярным ПО, но уже устаревшим. ClustalW был введен в 1994 году и стал методом для множественных выравниваний последовательностей. Этот метод продемонстрировал довольно резкое увеличение качества и скорости выравнивания по сравнению с предшествующими алгоритмами. При использовании данного метода выравнивание состоит из трех этапов: парные выравнивания, построение направляющего дерева и множественное выравнивание [21]. Прогрессивное выравнивание ClustalW использует парные выравнивания последовательностей, следуя порядку ветвления дерева, ранее построенного методом Neighbor-joining (NJ).

Выравнивание выполняется между всеми последовательностями в сравниваемой группе. Оценки выравнивания используются для построения матрицы расстояния. Вычисляя матрицу расстояний, программа учитывает расхождение последовательностей. [22].

- **MUSCLE** – новый алгоритм выравнивания множественных последовательностей. Этот метод основан на прогрессивном подходе к

множественному выравниванию. MUSCLE использует два расстояния меры. Скорость и точность MUSCLE сравниваются с T-COFFEE, MAFFT и CLUSTALW. И MUSCLE достигает средней точности, статистически неотличимой от T-COFFEE и MAFFT, и является самым проверенным методом для большого количества последовательностей [23].

- **T-COFFEE** – относительно новый метод быстрого и точного множественного выравнивания последовательностей. Этот метод также основан на прогрессивном алгоритме, как и MUSCLE, но избегает серьезных ошибок, вызванных характером этого алгоритма.

В прогрессивном выравнивании парные выравнивания завершаются сначала для создания матрицы расстояния. Затем эта матрица используется для создания дерева, затем это дерево используется для группировки последовательностей во время процесса выравнивания нескольких последовательностей. Самые близкие две последовательности на дереве сначала выравниваются с использованием обычного метода динамического программирования. Это делается для выравнивания остатков в двух последовательностях. Следующие две ближайшие последовательности, предложенные деревом направляющих или предварительно определенной группой последовательностей, всегда соединяются. Это продолжается до тех пор, пока все последовательности не будут выровнены. Чтобы выровнять две группы предварительно определенных последовательностей, используются оценки из расширенной библиотеки; однако принимаются средние оценки библиотек в каждом столбце существующего выравнивания. T-COFFEE повышает точность выравнивания на 5-10% по сравнению с ClustalW; однако алгоритм имеет недостатки, такие как слабая масштабируемость. T-COFFEE может выровнять максимум 100 последовательностей без потери точности [24]. T-COFFEE позволяет комбинировать выходные данные любых других выравниваний. Самым последним улучшением этого алгоритма стала разработка концепции множественного высокоточного выравнивания последовательностей на основе шаблонов [25].

## 1.8 Филогенетический анализ

В настоящее время филогенетический анализ стал необходимой частью любого филогенетического исследования, для решения различных биологических вопросов, таких как отношения между видами или генами. Филогенетический анализ позволяет устанавливать родственные связи между живыми организмами [26].

Филогенетический анализ последовательностей ДНК стал важным в изучении эволюционных взаимоотношений между организмами. Так как скорости эволюции последовательностей варьируются в зависимости от гена или сегмента ДНК, можно изучать эволюционные взаимоотношения практически всех уровней классификации организмов.

В данной работе используются маркерные последовательности ДНК: ITS (internal transcribed spacer), IGS (intergenic spacer) TEF1 $\alpha$  (translation elongation factor). Результатом филогенетического анализа является построение филогенетических деревьев на основе трех генетических маркеров. Дерево представляет собой гипотезу об эволюционных отношениях между организмами. Но необходимо помнить, что филогенетические деревья – это гипотезы, а не окончательные факты.

Наиболее часто используемые методы построения филогений на основе молекулярных данных основываются на моделях эволюции. Один из первых методов в построении филогенетических деревьев является метод максимального правдоподобия (Maximum Likelihood). Для расчета кладограммы, помимо последовательности ДНК, надо выбрать модель замены нуклеотидов, на основании которой будут рассчитываться вероятности. Также в расчет берется длина ветви или эволюционная дистанция между двумя таксонами. Во время анализа рассчитывается, какая длина ветви наиболее вероятна с точки зрения выбранной модели, вероятности всех ветвей кладограммы умножаются, и кладограмма, имеющая наибольшую вероятность, считается правильной [27].

Другой наиболее популярный метод, используемый в построении филогенетических деревьев - Байесовский подход. Этот метод максимального правдоподобия, поскольку он основывается на модели и длине ветвей. Но отличается он тем, что берется в расчет еще один фактор – апостериорная вероятность, которая рассчитывается как на основании исходных данных, так и полученных результатов анализа, создающая наиболее вероятное филогенетическое дерево. Байесовский подход стал популярным благодаря достижениям в скорости вычислений и интеграции алгоритмов марковской цепи и метода Монте-Карло (MCMC). Методы выбора Байесовской модели значительно улучшились.

Несмотря на все плюсы двух последних методов, тут можно найти некоторые сложности. Главная слабость методов в том, что необходимо выбирать модели замен самостоятельно. Но существуют программы, которые могут помочь подобрать подходящую модель [28].

Эволюционная дистанция используется при построении филогенетических деревьев и оценке времени дивергенции. В нуклеотидных последовательностях дистанция обычно измеряется как число нуклеотидных замен.

Подбор моделей нуклеотидных замен характерен для каждого гена и каждой последовательности в филогенетическом анализе. Вероятность замен определяется как функция ряда параметров: частот нуклеотидов и скорости замен [29].

Модель Jukes-Cantor (JC69, 1969) является самой простой моделью нуклеотидных замен. Эта модель предполагает, что нуклеотидные замены происходят в каждом сайте нуклеотида с одинаковой частотой и скоростью мутаций.

Модель Felsenstein (F81, 1981) является расширением модели Jukes-Cantor (JC69), в которой частоты нуклеотидов могут отличаться от 0,25 и все нуклеотидные замены одинаково вероятны, что позволяет применять эти две модели при более широких условиях, чем многие другие модели.



Модель Kimura (K80, 1980) предполагает одинаковые частоты встречаемости для всех нуклеотидов, но вероятность нуклеотидных замен различна для транзиций и трансверсий.

Транзиция - замена пурина (аденин и гуанин) на другой пурин или замена пиримидина (тимин или цитозин) на другой пиримидин. Трансверсия – замена пурина на пиримидин и наоборот.

Модель Hasegawa, Kishino and Yano (HKY85, 1985) различает скорость транзиций и трансверсий и позволяет использовать неодинаковые частоты нуклеотидов, также можно рассматривать как сочетание расширений, сделанных в моделях K80 и F81.

Модель T92 (Tamura, 1992) представляет собой математический метод, разработанный для оценки количества нуклеотидных замен на сайт между двумя последовательностями ДНК, путем расширения модели Kimura (K80) для случая низкого или высоко GC-содержания.

Модель Tamure-Nei (TrN, 1993) представляет собой разные частоты встречаемости нуклеотидов. Одной из моделей филогенетической оценки методом наибольшего правдоподобия является HKY модель, которая учитывает разницу между транзициями, трансверсиями и GC-содержания.

Модель General time reversible (GTR, 1986) считается наиболее сложной. Она использует разные частоты нуклеотидов (4 параметра), и разные частоты замен между нуклеотидами (6 параметров).

При построении филогенетических деревьев можно сделать вывод, что сложные модели не всегда более эффективны для получения правильной топологии чем простые, даже если модель более точно передает данные. Для оценки длин ветвей филогенетического дерева, модель, лучше описывающая данные, дает более правдоподобные результаты [30, 31].

## 2 Материалы и методы

Секвенирование 28 образцов рода *Armillaria* проводилось в Германии на капиллярном секвенаторе сотрудниками лаборатории лесной геномики Сибирского федерального университета. Эти образцы были взяты на территории Дальнего Востока, на юге Красноярского края и в Республике Хакасия. Филогенетический анализ широко применяется при необходимости провести идентификацию видов.

Благодаря филогенетическим исследованиям, проведенным за последние 10 –15 лет разными группами исследователей в разных странах, был накоплен большой массив данных по маркерным последовательностям ДНК разных видов *Armillaria*. Эти данные хранятся в открытом доступе в международной базе данных GenBank ([www.ncbi.nlm.nih.gov/genbank/](http://www.ncbi.nlm.nih.gov/genbank/)) и доступны для изучения, что создает возможность использования всего массива накопленной информации для переоценки существующих представлений о филогении *Armillaria*.

Для данного исследования использовались нуклеотидные последовательности представителей рода *Armillaria* для построения базы, взятые из международной базы данных GenBank. В анализ были включены данные 8 видов *Armillaria* и 1 вида *Flammulina*. В результате, в полученной матрице многие виды были представлены несколькими образцами. Использование маркерных последовательностей из GenBank позволяет сравнивать полученные последовательности с уже имеющимися, и тем самым идентифицировать наши 28 образцов.

Всего в анализ были включены 28 образцов рода *Armillaria*, 56 последовательностей ITS, 56 последовательностей IGS, и 56 последовательностей TEF1 $\alpha$  из 8 видов *Armillaria* и 1 вида *F. velutipes*. В качестве внешней группы (outgroup) была выбрана *F. velutipes*, как уже известно это эволюционно близкий вид для представителей *Armillaria*.

Таблица 1 - Количество маркерных последовательностей для разных видов *Armillaria* по результатам поиска в GenBank

Вид	ITS	IGS	TEF1 $\alpha$
<i>A. borealis</i>	6	6	6
<i>A. gallica</i>	11	11	11
<i>A. cepistipes</i>	12	12	12
<i>A. mellea</i>	6	6	6
<i>A. tabescens</i>	5	5	5
<i>A. sinapina</i>	5	5	5
<i>A. ectypa</i>	1	1	1
<i>A. sinapina</i>	5	5	5
<i>A. ostoyae</i>	9	9	9
<i>F. velutipes</i>	1	1	1

На основе полученных множественных выравниваний были построены филогенетические деревья методом максимального правдоподобия в программе IQ-TREE. В качестве внешней группы были использованы последовательности генетических маркеров рода *Flammulina*, взятые из GenBank. Выравнивания обеспечивают нас знанием о том, какие участки ДНК гомологичны друг с другом, и именно на основе различий в нуклеотидных заменах гомологов мы сможем затем реконструировать филогенетическое дерево.

## 2.1 Объекты исследования

Объектами исследования являлись грибы комплекса *Armillaria mellea sensu lato*. В данном исследовании использовалось 28 образцов, 8 видов представителей рода *Armillaria* и 1 эволюционно близкий вид *F. velutipes*, соответствующих маркерам ITS, IGS, и TEF-1 $\alpha$ .

*A. mellea* (опенок осенний) – является видом, распространенным по всему северному полушарию. Был найден в Северной Америке, Европе и

Северной Азии. Растет паразитарно на большом количестве широколиственных деревьев [3];

*A. borealis* (опенок северный) – встречается повсеместно на Евразийском континенте. Филогенетический анализ позволил определить близость данного вида к виду *A. gemina* [34];

*A. cepistipes* (опенок луковичноногий или серый) – общий древесный паразитирующий гриб, слабопатогенный вид растений, который встречается в большинстве лесов Центральной Европы [32];

*A. gallica* (опенок толстоногий) – важный патоген растений, встречающийся в умеренных регионах Азии, Северной Америки и Европы [33];

*A. ectypa* (опенок чеканный) – встречается в Евразии и крайне редко встречается в Европе [34];

*A. tabescens* – патоген растений, встречающийся в Северной Америке [32];

*A. sinapina* – встречается в Азии и Северной Америке [32].

## **2.2 Методы используемые для филогенетического анализа**

Множественное выравнивание было проведено с помощью программы SeaView. SeaView - это программа, которая позволяет выполнять полный филогенетический анализ набора гомологичных ДНК или последовательностей белка, отображает алгоритмы построения дерева максимального правдоподобия. Применяется алгоритм выравнивания MUSCLE, который создает высокоточные множественные выравнивания последовательностей [35].

На основе полученных выровненных последовательностей проводилось их конкатенирование, соединение генов в один файл.

Используя PartitionFinder 2, находили лучшие модели эволюционных замен для генов. Каждый ген может эволюционировать по-разному, поэтому для каждого гена можно рассчитать свою модель эволюции. PartitionFinder 2

– предоставляет набор моделей замен нуклеотидов для филогенетических анализов. Программа позволяет использовать два информационных критерия: AIC (информационный критерий Акаике), AICc (информационный скорректированный критерий Акаике) и BIC (Байесовский информационный критерий).

AIC (информационный критерий Акаике) – этот критерий применяется для выборки из нескольких моделей.

BIC (Байесовский информационный критерий) – критерий выбора модели среди других моделей зависит от разного числа параметров. BIC выберет истинную модель, если, среди других предположений, истинная модель входит в число рассмотренных моделей. В этих обстоятельствах AIC не согласуется. Когда истинная модель не входит в число моделей, AIC эффективна, в том, что он будет выбирать любую модель, которая минимизирует среднеквадратичную ошибку прогнозирования, оценки. В этих условиях BIC не эффективен. В отличие от BIC, AIC также обладает минимаксным свойством, поскольку он может минимизировать максимально возможный риск при конечных размерах выборки. В целом, AIC и BIC имеют совершенно разные свойства [36].

PartitionFinder 2 способен анализировать морфологические наборы данных, новые методы анализа наборов данных, новые форматы вывода для облегчения взаимодействия с программным обеспечением и множество новых моделей эволюционных замен [37].

Построение деревьев было выполнено с помощью программы IQ-TREE, основанной на методе максимального правдоподобия. Этот метод основывается на модели и длине ветвей. Во время анализа рассчитывается, какая длина ветви наиболее вероятна с точки зрения выбранной модели. IQ-TREE – программа, которая основана на построении филогенетических деревьев. Филогения каждого гена производится с соответствующей моделью и параметрами [38].

### 3 Результаты и их обсуждение

Были построены филогенетические деревья для идентификации 28 образцов. На рисунке 1 изображено филогенетическое дерево на основе маркеров ITS, IGS, TEF1 $\alpha$ , построенное с помощью метода максимального правдоподобия. Был использован байесовский информационный критерий (BIC).

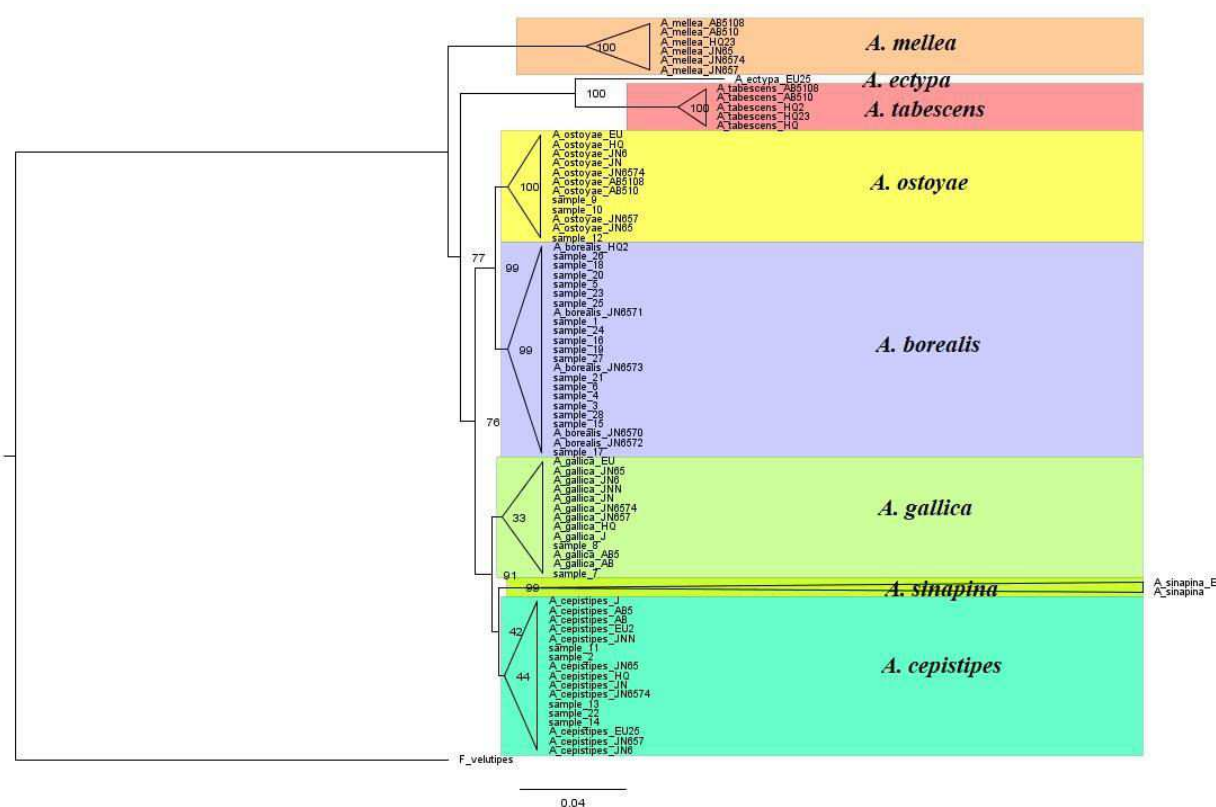


Рисунок 1 – Филогенетическое дерево на основе маркеров ITS, IGS, TEF1 $\alpha$

Построение общего дерева по трем маркерам ITS, IGS, TEF1 $\alpha$  показало, что 28 образцов распределились по 4 клатам с уже известными видами и высоким уровнем поддержки. Новых видов не выявлено. Наибольшее число образцов было отнесено к кладе *A. borealis*, таким образом 18 из 28 образцов относятся к виду *A. borealis*, 5 образцов относятся к *A. cepistipes*, 3 образца – *A. ostoyae* и 2 образца – *A. gallica*. Построение дерева по трем маркерам

довольно сильно отличается, нежели построение деревьев по одному из маркеров.

На рисунке 2 изображено филогенетическое дерево на основе маркера IGS, построенное с помощью метода максимального правдоподобия. Также был использован информационный критерий.

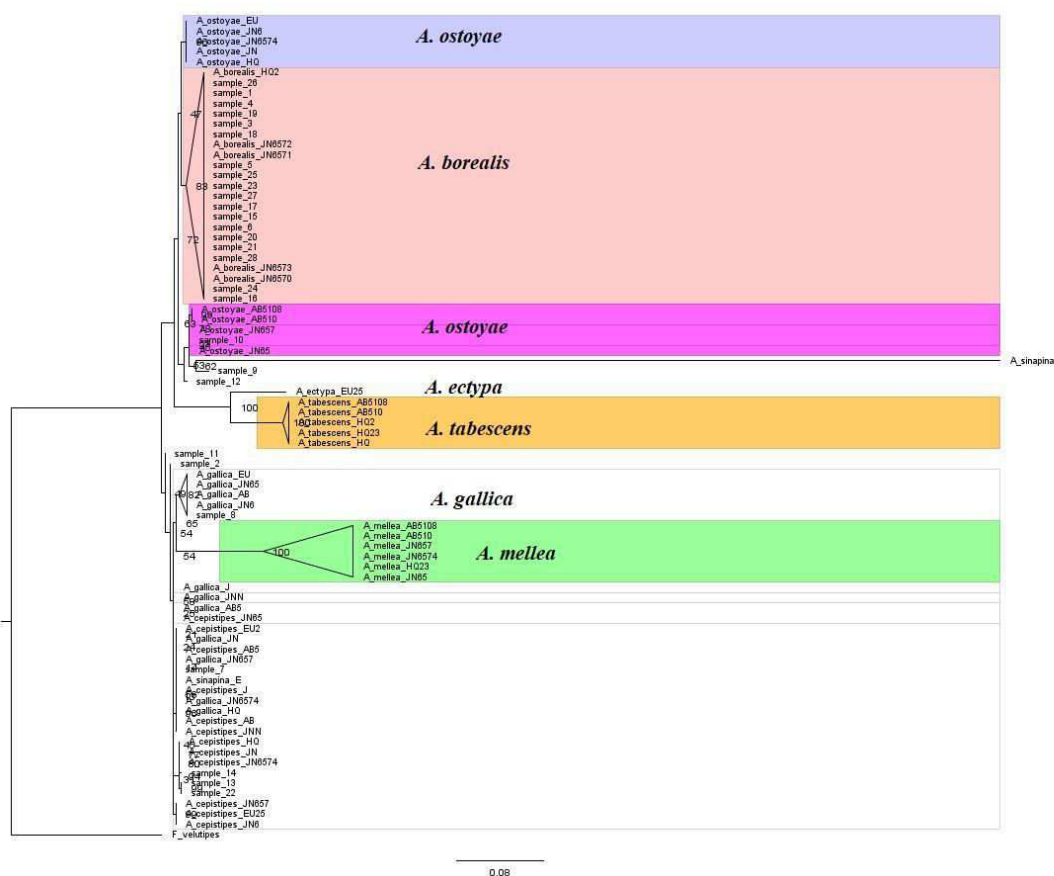


Рисунок 2 – Филогенетическое дерево на основе маркера IGS

На основе полученных результатов на рис. 2 можно сказать, что объединились в одной кладе те же 18 из 28 образцов, относящихся к виду *A. borealis*, как и в общем дереве на основании трех маркеров, что доказывает точную идентификацию образцов из вышеуказанных результатов. 1 образец отнесен к кладе *A. ostoyae* и еще 2 образца отделились. 4 образца относятся к *A. cepistipes*, а один образец также отделился. К кладе *A. gallica* отнесено 3 образца. Можно предположить, что нельзя проводить идентификацию по одному маркеру, возможно влияние организменной дивергенции, что

показывает не очень высокую устойчивость или же необходимо применить другой метод построения филогенетического дерева и другую эволюционную модель замен.

На рисунке 3 изображено филогенетическое дерево на основе маркера ITS.

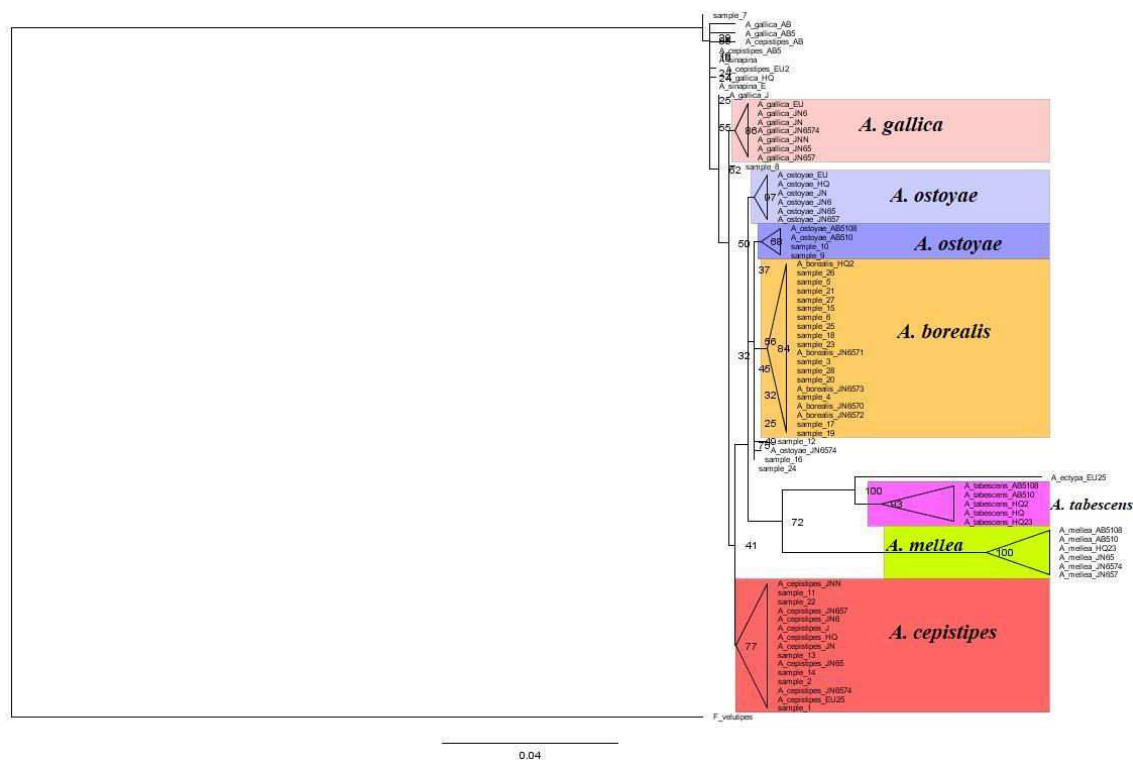


Рисунок 3 – Филогенетическое дерево на основе маркера ITS

Филогенетическое дерево на основе маркера ITS дает нам похожий результат, как и на основе маркера IGS. Все образцы распределились по разным кладам, а некоторые даже не объединились в кладе с уже известными видами. 15 из 28 образцов образовались в кладе *A. borealis*, *A. cepistipes* и *A. gallica* смешались и не образуют общие группы эволюционных организмов.

На рисунке 4 изображено филогенетическое дерево на основе маркера TEF1 $\alpha$ .



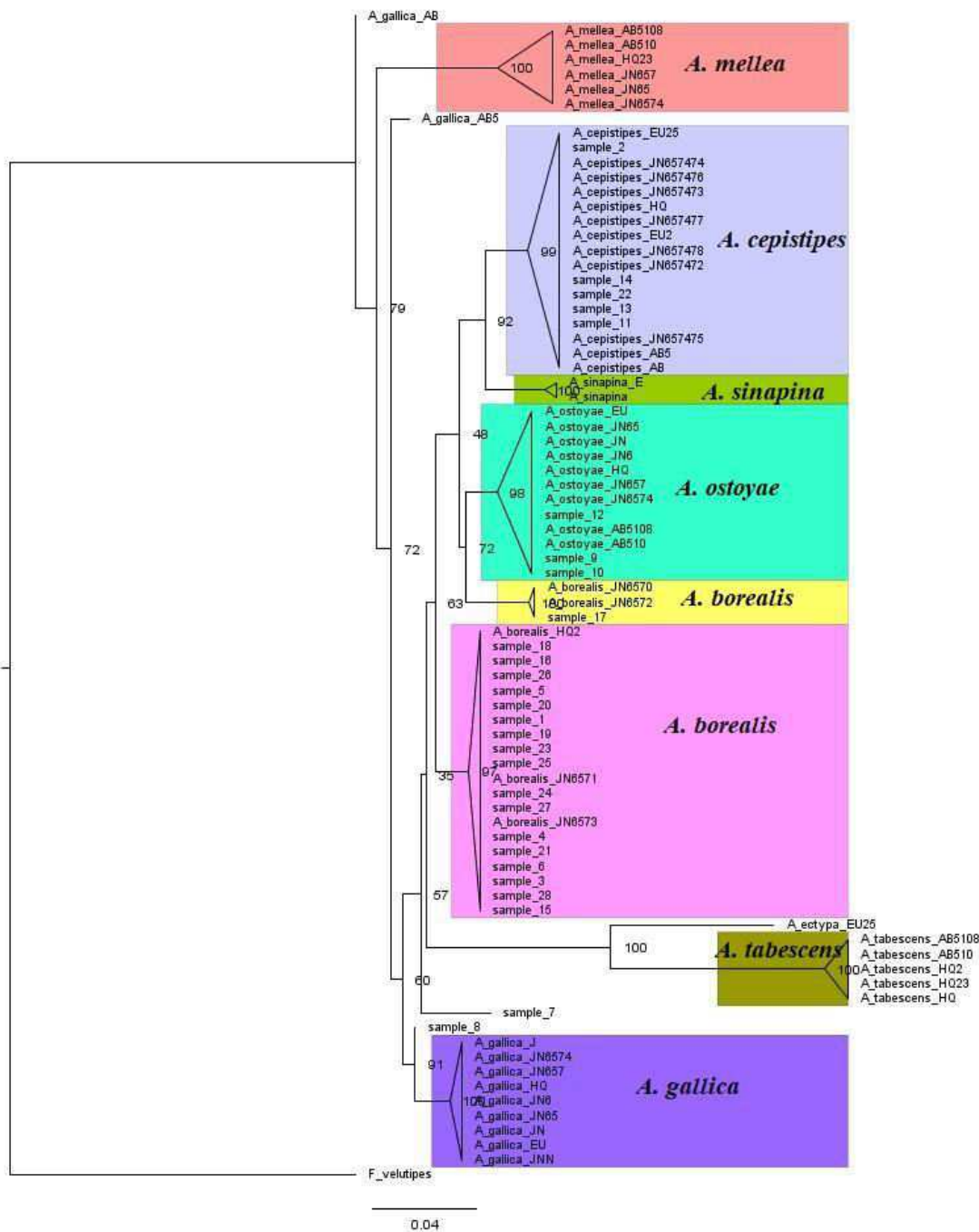


Рисунок 4 – Филогенетическое дерево на основе маркера TEF1 $\alpha$

Филогенетическое дерево на основе маркера TEF1 $\alpha$  дает нам более правдоподобное дерево, с хорошей устойчивостью и хорошим распределением образцов с уже известными видами. Результат данного филогенетического дерева отражает наиболее схожий результат с использованием всех маркеров. 18 из 28 образцов образовались в кладе *A. borealis*, 5 образцов - *A. cepistipes*, и 5 образцов - *A. ostoyae*, однако, *A. gallica* разделились и не образуют общие группы организмов.

## ЗАКЛЮЧЕНИЕ

Впервые была произведена идентификация 28 сибирских и дальневосточных образцов представителей рода *Armillaria*;

Филогенетический анализ позволил с высокой вероятностью определить образцы, относящиеся к представителям *Armillaria*;

Исследуемые образцы относятся к 4 разным видам, 18 из 28 образцов относятся к виду *A. borealis*;

Все образцы попадают в кластеры с уже известными видами.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Guo, T. Phylogenetic analyses of *Armillaria* reveal at least 15 phylogenetic lineages in China, seven of which are associated with cultivated *Gastrodia elata* / T. Guo, H. C. Wang, W. Q. Xue, J. Zhao, Z. Yang // PLoS One. – 2016. – Т. 11. – №. 5. – С. e0154794;
2. Baumgartner, K. Secrets of the subterranean pathosystem of *Armillaria* / K. Baumgarther, M. Coetzee, D. Hoffmeister // Molecular plant pathology. – 2011. – Т. 12. – №. 6. – С. 515-534;
3. Koch, R. A. Resolved phylogeny and biogeography of the root pathogen *Armillaria* and its gasteroid relative, *Guyanagaster* / R. A. Koch, A. W. Wilson, O. Sene, T. W. Henkel, M. C. Aime // BMC evolutionary biology. – 2017. – Т. 17. – №. 1. – С. 33-38;
4. Sipos, G. Genome expansion and lineage-specific genetic innovations in the forest pathogenic fungi *Armillaria* / G. Sipos, A. N. Prasanna // Nature ecology & evolution. – 2017. – Т. 1. – №. 12. – С. 1931-1941;
5. Coetzee, M. P. A. Molecular identification and phylogeny of *Armillaria* isolates from South America and Indo-Malaysia / M. P. A. Coetzee, B. D. Wingfield, P. Bloomer, G. S. Ridley // Mycologia. – 2003. – Т. 95. – №. 2. – С. 285-293;
6. Park, Y. J. Whole genome and global gene expression analyses of the model mushroom *Flammulina velutipes* reveal a high capacity for lignocellulose degradation / Y. J. Park, / R. A. Koch, J. H. Baek, S. Lee, C. Kim, H. Rhee // PloS one. – 2014. – Т. 9. – №. 4. – С. e93560;
7. Joh, J. H. Comparative analysis of expressed sequence tags from *Flammulina velutipes* at different developmental stages / J. H. Jon, K. Y. Kim, E. S. Son, H. R. Park // Journal of microbiology and biotechnology. – 2009. – Т. 19. – №. 8. – С. 774-780;

8. Kim, C. K. Molecular marker database for efficient use in agricultural breeding programs / C.K. Kim, G. S. Lee, J. C. Mo, S. H. Bae, T. H. Lee //Bioinformation. – 2015. – Т. 11. – №. 9. – С. 444-448;
9. Badotti, F. Effectiveness of ITS and sub-regions as DNA barcode markers for the identification of Basidiomycota (Fungi) / F. Badotti, F. Oliveira, C. F. Garcia, A. Vaz, P. Fonseca, L. A. Nahum, G. Oliveira, A. Goes-Neto //BMC microbiology. – 2017. – Т. 17. – №. 1. – С. 42-47;
10. Bellemain, E. ITS as an environmental DNA barcode for fungi: an *in silico* approach reveals potential PCR biases / E. Bellemain, T. Carlsen, C. Brochmann, E. Coissac, P. Taberlet, H. Kauserud //BMC microbiology. – 2010. – Т. 10. – №. 1. – С. 189-198;
11. Матвеева, Т. В. Молекулярные маркеры для видоидентификации и филогенетики растений / Т. В. Матвеева, О. А. Павлова, Д. И. Богомаз //Ecological genetics. – 2011. – Т. 9. – №. 1. – С. 33-35;
12. Cornet, M. Molecular identification of closely related *Candida* species using two ribosomal intergenic spacer fingerprinting methods / M. Cornet, B. Sendid //The Journal of Molecular Diagnostics. – 2011. – Т. 13. – №. 1. – С. 12-22;
13. Ichi-ishi, A., Inoue H. Cloning, nucleotide sequence, and expression of *tef-1*, the gene encoding translation elongation factor 1-alpha of *Neurospora crassa* / A. Ichi-Ishi, H. Inoue //– 1995. – Т. 70. – №. 2. – С. 273-287;
14. Кочкина, Г. А. Галопсихротолерантные грибы рода *Geomyces* из криопэгов и морских отложений Арктики / Г. А. Кочкина, Н. Е. Иванушкина, В. Н. Акимов //Микробиология. – 2007. – Т. 76. – №. 1. – С. 39-47;
15. Иванов, Д. М. Верификация метода рестрикционного анализа рДНК для изучения геномного полиморфизма представителей порядка Boletales / Д. М. Иванов //Вестник Санкт-Петербургского университета. Серия 3. Биология. – 2008. – №. 4. – С. 112-113;

16. Гагкаева, Т. Ю. Использование метода ПЦР для выявления и идентификации грибов рода *Fusarium* и трихотеценовых токсинов / Т. Ю. Гагкаева, М. М. Левитин //МОЛЕКУЛЯРНАЯ ГЕНЕТИКА В МЕДИЦИНСКОЙ МИКОЛОГИИ. ГЕНОМЫ ГРИБОВ И ДОСТИЖЕНИЯ В ГЕНОДИАГНОСТИКЕ ГРИБКОВЫХ ЗАБОЛЕВАНИЙ ЧЕЛОВЕКА. – 2005. – С. 12;
17. Шнырева, А. А. Грибы рода *Pleurotus*: генотипирование и анализ локусов половой совместимости : дис. ... канд. биол. наук : 03.02.2012 / Шнырева Анастасия Андреевна //МГУ, Москва. – 2015;
18. Кочкина, Г. А. Галопсихротолерантные грибы рода *Geomyces* из криопэггов и морских отложений Арктики / Г. А. Кочкина, Н. Е. Иванушкина, В. Н. Акимов //Микробиология. – 2007. – Т. 76. – №. 1. – С. 39-47;
19. Кокаева, Л. Ю. МИКОБИОТА ПОРАЖЕННЫХ ЛИСТЬЕВ *SOLANUM TUBEROSUM L.*, *S. LYCOPERSICUM L.* И *S. DULCAMARA L.*: дис. ... канд. биол. наук : 03.02.12 / Кокаева Людмила Юрьевна // МГУ, Москва. – 2016;
20. Daugelaite, J., O'Driscoll, A., Sleator, R. D. An overview of multiple sequence alignments and cloud computing in bioinformatics //ISRN Biomathematics. – 2013. – Т. 2013;
21. Бутвиловский, А. В. Выравнивание аминокислотных и нуклеотидных последовательностей / А. В. Бутвиловский, Е. В. Барковский, В. Э. Бутвиловский // Медицинский журнал. – 2007. – №. 1. – С. 45-54;
22. Daugelaite, J., O'Driscoll, A., Sleator, R. D. An overview of multiple sequence alignments and cloud computing in bioinformatics //ISRN Biomathematics. – 2013. – Т. 2013;
23. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput / R. C. Edgar //Nucleic acids research. – 2004. – Т. 32. – №. 5. – С. 1792-1797;

24. Notredame, C. T-coffee: a novel method for fast and accurate multiple sequence alignment / D. G. Higgins, J. Heringa // *Journal of molecular biology*. – 2000. – Т. 302. – №. 1. – С. 205-217;
25. Tommaso, P. D. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension / P. D. Tommaso, S. Moretti, I. Xenarios // *Nucleic acids research*. – 2011. – Т. 39. – №. 2. – С. 13-17;
26. Yang, Z. *Molecular phylogenetics: principles and practice* / Z. Yang, B. Rannala // *Nature Reviews Genetics*. – 2012. – Т. 13. – №. 5. – С. 303-312;
27. Lin, Y. Maximum likelihood phylogenetic reconstruction from high-resolution whole-genome data and a tree of 68 eukaryotes / Y. Lin, F. Hu, J. Tang // *Biocomputing 2013*. – С. 285-296;
28. Ronquist, F. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space / F. Ronquist, M. Teslenko, D. L. Ayres // *Systematic biology*. – 2012. – Т. 61. – №. 3. – С. 539-542;
29. Liò, P. Goldman N. Models of molecular evolution and phylogeny / P. Liò, N. Goldman // *Genome research*. – 1998. – Т. 8. – №. 12. – С. 1233-1244;
30. Шнырева, А. В. Геносистематика и проблема вида у грибов: подходы и решения / А. В. Шнырева // *Микология и фитопатология*. – 2011. – Т. 45. – №. 3. – С. 209-220;
31. Appel, D. Relationships among pathogenic and nonpathogenic isolates of *Fusarium oxysporum* based on the partial sequence of the intergenic spacer region of the ribosomal DNA / D. J. Appel, T. R. Gordon // *MPMI-Molecular Plant Microbe Interactions*. – 1996. – Т. 9. – №. 2. – С. 125-138;
32. Heinzelmann, R. Population genetics of the wood-rotting basidiomycete *Armillaria cepistipes* in a fragmented forest landscape / R. Heinzelmann, R. Rigling, S. Prospero // *Fungal biology*. – 2012. – Т. 116. – №. 9. – С. 985-994;

33. Brazeo, N. J. Genotypic diversity of *Armillaria gallica* from mixed oak forests in Massachusetts / N. J. Brazeo, R. E. Marra, R. L. Wick // *Mycologia*. – 2012. – T. 104. – №. 1. – C. 53-61;
34. Klopfenstein, N. B. Insights into the phylogeny of Northern Hemisphere *Armillaria*: Neighbor-net and Bayesian analyses of translation elongation factor 1- $\alpha$  gene sequences / N. B. Klopfenstein, J. E. Stewart, Y. Ota, J. W. Hanna // *Mycologia*. – 2017. – T. 109. – №. 1. – C. 75-91;
35. Gouy, M. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building / M. Gouy, S. Guindon, O. Gascuel // *Molecular biology and evolution*. – 2009. – T. 27. – №. 2. – C. 221-224;
36. Vrieze, S. I. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) / S. I. Vrieze // *Psychological methods*. – 2012. – T. 17. – №. 2. – C. 228-243;
37. Lanfear, R. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses / R. Lanfear, P. B. Frandsen, A. M. Wright // *Molecular Biology and Evolution*. – 2016. – T. 34. – №. 3. – C. 772-773;
38. Nguyen, L. T. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies / L. T. Nguyen, H. A. Schmidt, B. Q. Minh // *Molecular biology and evolution*. – 2014. – T. 32. – №. 1. – C. 268-274.

Федеральное государственное автономное  
образовательное учреждение  
высшего образования  
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Институт фундаментальной биологии и биотехнологии  
Кафедра биофизики

УТВЕРЖДАЮ

Заведующий кафедрой

*В.А. Кратасюк*  
В.А. Кратасюк  
« 9 » 06 2018 г.

**БАКАЛАВРСКАЯ РАБОТА**

06.03.01 Биология

06.03.01.07 Биофизика

Анализ маркерных последовательностей представителей рода *Armillaria*

Научный  
руководитель

*Ю. А. Путинцева*

Ю. А. Путинцева

Научный  
консультант

*М. Г. Садовский*  
14.06.2018

д.ф.-м.н., в.н.с М. Г. Садовский

Выпускник

*О. С. Зуева*

О. С. Зуева

Красноярск 2018