

Федеральное государственное бюджетное учреждение «Детский научно-клинический центр инфекционных болезней Федерального медико-биологического агентства»

На правах рукописи

Цветкова Ирина Анатольевна

**Генотипическая характеристика *Streptococcus pneumoniae*,  
принадлежащих к эпидемическим генетическим линиям**

1.5.11 – микробиология

Диссертация на соискание ученой степени кандидата  
биологических наук

Научный руководитель:  
доктор медицинских наук, профессор Сидоренко Сергей Владимирович

Санкт-Петербург – 2021

## СОДЕРЖАНИЕ

<b>ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ</b> .....	7
Актуальность темы исследования .....	7
Степень разработанности темы исследования.....	7
Цель исследования .....	9
Задачи исследования .....	9
Научная новизна.....	9
Теоретическая и практическая значимость работы .....	10
Перспективы дальнейшего изучения .....	11
Степень достоверности и апробация результатов .....	11
Методология и методы работы.....	12
Основные положения, выносимые на защиту.....	12
Личное участие автора в получении результатов .....	13
Соответствие диссертации паспорту научной специальности.....	13
Объем и структура диссертационной работы.....	14
<b>ГЛАВА 1. ОБЗОР ЛИТЕРАТУРЫ</b> .....	15
1.1. Современная таксономическая классификация стрептококков.....	15
1.2. Потребности в питании и метаболические характеристики <i>S. pneumoniae</i> ....	15
1.3. Молекулярная эпидемиология <i>Streptococcus pneumoniae</i> .....	18
1.3.1. Задачи молекулярной эпидемиологии .....	18
1.3.2. Методы молекулярной эпидемиологии .....	18
1.4. Биологические процессы, лежащие в основе эволюции <i>S. pneumoniae</i> .....	20
1.4.1. Горизонтальный перенос генов .....	21
1.4.1.1. Трансформация у пневмококков .....	21

1.4.1.2. МГЭ-опосредованный перенос генетического материала между бактериями .....	22
1.4.1.2.1. Конъюгация .....	23
1.4.1.2.2. Трансдукция .....	25
1.4.2. Роль систем рестрикции-модификации в эволюции <i>S. pneumoniae</i> .....	26
1.4.2.1. Распространенность систем рестрикции-модификации .....	26
1.4.2.2. СРМ у <i>S. pneumoniae</i> .....	26
1.4.2.3. Фенотипы, ассоциированные с особенностями фазо-вариабельной системы типа I.....	29
1.4.3. Короткие диспергированные повторы <i>S. pneumoniae</i> .....	30
1.4.4. Ассоциация геномных островков с генетическими линиями .....	31
1.5. <i>cps</i> -локус .....	33
1.5.1. Эволюция генов <i>cps</i> -локуса .....	34
1.5.2. Переключение серотипов (факты, механизмы).....	35
1.6. Факторы вирулентности <i>S. pneumoniae</i> .....	36
1.6.1. Описание некоторых факторов вирулентности .....	38
1.7. Распространение резистентности к бета-лактамам антибиотикам у <i>S. pneumoniae</i> .....	44
1.8. Механизмы резистентности к бета-лактамам антибиотикам у <i>S. pneumoniae</i> .....	44
1.9. Резистентность к антибиотикам других классов в популяциях <i>S. pneumoniae</i> .....	45
1.10. Механизмы и детерминанты резистентности к антибиотикам других классов .....	46
1.11. Заключение по обзору литературы.....	48
<b>ГЛАВА 2. МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ.....</b>	<b>50</b>
2.1. Общий план работы .....	50

2.2. Формирование выборки .....	50
2.3. Характеристика изолятов <i>S. pneumoniae</i> из музейной коллекции ФГБУ ДНКЦИБ ФМБА России (Санкт-Петербург).....	51
2.4. Филогенетический анализ по конкатенатам последовательностей генов MLST .....	52
2.4.1. Выравнивание последовательностей .....	52
2.4.2. Анализ структуры популяции.....	52
2.4.3. Парные оценки генетических расстояний между идентифицированными в популяции группами.....	53
2.5. Анализ полногеномных данных .....	53
2.5.1. Общедоступные базы данных .....	53
2.5.2. Сборка геномов <i>de novo</i> и аннотация .....	53
2.6. Анализ ядерной части генома .....	53
2.6.1. Идентификация ядерной части генома и исключение сайтов рекомбинаций с помощью программного пакета Harvest tools Parsnp.....	53
2.6.2. Идентификация ядерной части геномов с помощью модуля GenomeComparator в программе BIGSdb (Bacterial Isolate Genome Sequence Database) и исключение рекомбинаций с помощью программы Gubbins .....	54
2.7. Поиск переменных генов, вариации в которых участвуют в формировании кладов A/B1/B2, сиквентс-кластеров и серотипов, резистентности к бета-лактамам и макролидам, или ассоциируются с инвазивностью .....	55
2.7.1. Идентификация аллелей генов .....	55
2.7.2. Множественный анализ соответствий (Multiple correspondens analisys, MCA).....	55
2.7.3. Классификация с помощью алгоритма машинного обучения, основанного на методе «Случайного леса» (Random Forest).....	55
2.7.4. Предсказание генов, формирующих анализируемые группы с помощью алгоритма машинного обучения, основанного на моделях градиентного бустинга для линейных основанных на деревьях моделей XGBoost .....	56
2.8. Анализ наличия систем рестрикции-модификации .....	56

2.9. Анализ содержания профагов в геномах .....	56
2.10. Анализ рекомбинаций в <i>cps</i> -локусе изолятов, принадлежащих генетическим линиям CC90, CC236 / CC271 / CC320 .....	56
2.11. Анализ рекомбинаций в генах пенициллин-связывающих белков анализируемых изолятов и других видов стрептококков .....	57
2.12. Поиск значимых однонуклеотидных полиморфизмов, ассоциированных с резистентностью к бета-лактамам антибиотикам, у изолятов CC320.....	57
2.13. Оценка полиморфизма индивидуальных белков у резистентных к пенициллину изолятов, принадлежащих клональному комплексу CC320.....	58
2.14. Оценка топологии филогенетических деревьев .....	58
2.15. Аннотация метаболических путей.....	58
2.16. Анализ наличия детерминант резистентности к макролидным антибиотикам и тетрациклинам .....	58
<b>ГЛАВА 3. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ .....</b>	<b>59</b>
3.1. Оценка структуры популяции <i>S. pneumoniae</i> , циркулирующих в России .....	59
3.1.1. Филогенетический анализ на основании сравнения локусов схемы MLST	59
3.1.2. Анализ популяции <i>S. pneumoniae</i> по ядерной части генома .....	73
3.1.3. Оценка вклада генов ядерного генома в формирование структуры популяции <i>S. pneumoniae</i> (групп А/В1/В2, сиквенс-кластеров, серотипов, инвазивных штаммов резистентных к антибиотикам штаммов) .....	76
3.1.3.1. Гены, формирующие группы А/В1/В2А .....	84
3.1.3.2. Гены, участвующие в формировании сиквенс-кластеров (SC_MLST).....	88
3.1.3.3. Гены, участвующие в дифференциации на серотипы .....	92
3.1.3.4. Гены, ассоциированные с делением популяции <i>S. pneumoniae</i> на штаммы, чувствительные и резистентные к бета-лактамам антибиотикам .....	96
3.1.3.5. Взаимодействие генов, участвующих в делении популяции на группы А/В1/В2, SC_MLST и серотипы, инвазивные, резистентные и чувствительные к антибиотикам штаммы .....	97
3.1.4. Оценка вклада рекомбинаций в структуру популяции .....	102

3.1.5. Системы рестрикции анализируемых изолятов .....	104
3.1.6. Анализ присутствия профагов в геномах анализируемых изолятов .....	108
Выводы по разделу: .....	108
3.2. Анализ детерминант резистентности к антибиотикам изолятов <i>S. pneumoniae</i> , относящихся к эпидемическим генетическим линиям .....	110
3.2.1. Анализ детерминант резистентности в геномах российских и референсных штаммов .....	110
3.2.2. Динамика циркуляции генетических линий в России в различные периоды с 1980 по 2017 гг. Ассоциация с резистентностью к антибиотикам .....	112
3.2.3. Значимые однонуклеотидные полиморфизмы, ассоциированные с резистентностью к бета-лактамам антибиотикам у изолятов СС320 .....	116
Выводы по разделу: .....	122
3.3. Характеристика геномных локусов изолятов <i>S. pneumoniae</i> эпидемических генетических линий, детерминирующих серотиповую принадлежность и устойчивость к бета-лактамам антибиотикам .....	124
Выводы по разделу: .....	127
3.4. Сравнительный анализ генов пенициллин-связывающих белков изолятов <i>S. pneumoniae</i> , относящихся к эпидемическим генетическим линиям .....	127
Выводы по разделу: .....	130
3.5. Гены вирулентности, продукты которых могут быть потенциальными мишенями для создания белковой антипневмококковой вакцины .....	130
Выводы по разделу: .....	132
<b>ЗАКЛЮЧЕНИЕ И ВЫВОДЫ .....</b>	<b>133</b>
Рекомендации по использованию результатов исследования .....	136
Список публикаций по теме диссертации .....	137
Список используемых сокращений и обозначений .....	140
Список использованной литературы .....	141

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

### Актуальность темы исследования

Несмотря на применение антипневмококковых полисахаридных конъюгированных вакцин (ПКВ) и антибиотиков, пневмококковая инфекция (ПИ) остается частой причиной заболеваемости и смертности во всем мире [1, 2]. В настоящее время много известно о циркулирующих в мире эпидемически значимых клонах, молекулярных маркерах резистентности к антибиотикам и инвазивности, а также об истории эволюции некоторых генетических линий [1-10]. Однако много остается неясного в механизмах, с помощью которых пневмококковая популяция отвечает на селективное давление вакцин и антибиотиков. В частности, прогноз реакции пневмококковой популяции на применение ПКВ до сих пор не ясен, поскольку появляются данные о росте случаев инвазивных ПИ, вызванных невакцинными серотипами [1, 2, 6-10]. Для решения проблемы инвазивных ПИ необходимо не только иметь полное представление об эпидемиологической ситуации в России и мире, но также понимать генетические процессы в популяции *S. pneumoniae*, которые происходили как до, так и на фоне вакцинации, а также механизмы, способствующие увеличению вирулентности пневмококка.

### Степень разработанности темы исследования

Введение вакцинации ПКВ7 и ПКВ13 во многих странах привело к снижению распространенности инвазивных ПИ после 2010 г. Однако в последние годы показатели регистрируемой в мире заболеваемости инвазивными ПИ увеличиваются [1, 2].

Наиболее частые серотипы пневмококка, вызывающие ПИ, соответствуют спектру распространенных циркулирующих серотипов в регионах [3-5]. Согласно опыту иммунизации детей против *S. pneumoniae* в других странах, вакцинация приводит к изменению спектра циркулирующих серотипов, а также к повышению резистентности к антибиотикам невакцинных серотипов пневмококка [6, 7]. Так, согласно данным ECDC, в 2017 г. в странах Евросоюза, было зарегистрировано 23886 случаев инвазивных заболеваний и 68% этих случаев ассоциировались с 10 наиболее распространенными серотипами (почти все не покрываются ПКВ): 8, 3, 22F, 19A, 12F, 9N, 15A, 10A, 11A и 23B (в порядке убывания частоты) [2]. По сравнению с 2013 г., частоты распространения

серотипов 8, 12F и 9N увеличились на 120%, 87% и 85%, соответственно (по данным ежегодных отчетов, предоставленных странами Австрии, Кипра, Чешской Республики, Дании, Эстонии, Финляндии, Франции, Греции, Венгрии, Исландии, Ирландии, Италии, Латвии, Литвы, Нидерландов, Норвегии, Португалии, Словакии, Словении, Испании, Швеции и Великобритании). По данным 2017 г., случаи инвазивных ПИ у детей в возрасте до 5 лет были вызваны: 8% - пневмококками серотипов ПКВ7 (4, 6А, 6В, 9V, 14, 18С, 19F и 23F); 1% - серотипами ПКВ10 / не покрываемыми ПКВ7 (1, 5 и 7F); 15% - серотипами ПКВ13 / не покрываемыми ПКВ10 (3 и 19А); 76% - серотипами, не покрываемыми какой-либо ПКВ вакциной [2]. Число ассоциируемых с инвазивными ПИ серотипов, не относящихся к ПКВ, выросло с 63% до 76% за 2016 - 2017 гг. [2]. В настоящее время в России, даже в ранний период после начала вакцинации, сохраняется доминирование вакцинных серотипов 19F, 6А/В, 3, 23F, но наблюдается распространение невакцинных 11А/Д, 15А/Ф и 23А [8-10].

Большинство резистентных к антибиотикам и инвазивных пневмококков до применения конъюгированных вакцин принадлежало ограниченному числу распространенных клонов [11-15]. С началом использования ПКВ-вакцин переключение серотипа сыграло роль в возникновении некоторых успешных клонов, ассоциированных с невакцинными серотипами [16-18]. Однако полагают, что в большинстве случаев подобные клоны циркулировали до введения вакцин и получили преимущество для распространения на фоне вакцинации [19, 20]. Тем не менее, не любая комбинация генотип-серотип становится распространенным клоном [17, 18], [21, 22]. Решающее значение для способности штамма к колонизации и вирулентности имеет не только серотип, но и его общий генетический фон [22]. Таким образом, идентификация и характеристика распространенных клонов необходима для понимания динамики популяции пневмококка.

В Российской Федерации ПКВ13 была внедрена в 2015 г. Тем не менее, уже на ранних сроках после начала вакцинации отмечается изменение серотипового состава пневмококковой популяции в России: рост распространенности невакцинных серогрупп 11AD и 15BC среди здоровых детей до 5 лет [9], а также увеличение частоты невакцинных серотипов, ассоциированных с инвазивными ПИ у взрослых [10]. По данным 2019 г., наблюдалась низкая перекрываемость клинических изолятов *S. pneumoniae*, полученных от взрослых: ППВ-23 выше 60%, а ПКВ-13 – не более 50%



(острый средний отит – 50,0%; внебольничная пневмония - 45,6%; другие инвазивные ПИ – 46,7%) [10]. В связи с этим, для оценки эффективности вакцинации и мониторинга степени роста распространенности инвазивных ПИ, вызванных невакцированными серотипами, необходимо проведение эпидемиологических исследований в РФ во всех возрастных группах пациентов, а также среди носителей.

### **Цель исследования**

Дать генотипическую характеристику изолятов *Streptococcus pneumoniae*, относящихся к эпидемическим генетическим линиям, циркулирующим в Российской Федерации.

### **Задачи исследования**

1. Оценить популяционную структуру и провести сравнительный биоинформатический анализ данных полногеномного секвенирования изолятов *S. pneumoniae*, относящихся к эпидемическим генетическим линиям.
2. Проанализировать детерминанты резистентности к антибиотикам изолятов *S. pneumoniae*, относящихся к эпидемическим генетическим линиям.
3. Охарактеризовать геномные локусы изолятов *S. pneumoniae* эпидемических генетических линий, детерминирующих серотиповую принадлежность и устойчивость к бета-лактамам антибиотикам.
4. Провести сравнительный биоинформатический анализ генов пенициллин-связывающих белков у изолятов *S. pneumoniae*, относящихся к эпидемическим генетическим линиям.
5. Провести сравнительный анализ генов вирулентности, выявить потенциальные мишени для создания белковой антипневмококковой вакцины.

### **Научная новизна**

В результате филогенетического анализа выявлено, что мировая популяция *S. pneumoniae* представлена тремя глобальными группами А, В1 и В2. Наибольший вклад в деление на данные группы вносят гены, кодирующие: сигнальную пептидазу 1, участвующую в процессинге секретлируемых белков; гены глюкокиназы и глюкозо-6-фосфат-1-дегидрогеназы, регулирующие поток глюкозы в клетку. В группе А доминируют серотипы 23F, 19F, 14, 23А, в группе В1 – серотипы 11А, 19F, 19А, 1, 9N, а в группе В2 – серотипы 6А/В/Е, 3, 19А, 7F, 5. Группа В2 ассоциируется преимущественно с серотипами, не входящими в состав вакцины ПКВ7.

Установлено, что происходившие в 2000 – 2010 гг. изменения в структуре популяции пневмококков, циркулирующих в Российской Федерации, были обусловлены изменениями в структуре глобальной популяции в ответ на массовую вакцинацию ПКВ7 в различных регионах мира. Происходивший в этот период рост устойчивости к бета-лактамам и макролидным антибиотикам связан с преимущественным глобальным распространением серотипа 19А, не входящего в состав ПКВ7.

Выявлена глобальная тенденция к распространению после 2000 г. генетических линий, ассоциированных с инвазивностью (группа В2).

Установлено, что особенности метаболизма углеводов и ароматических аминокислот (пневмококки способны синтезировать фенилаланин, тирозин и триптофан) детерминируют регуляцию клеточных процессов пневмококка и фенотипические особенности, в том числе вирулентный потенциал пневмококка. Серотип-специфичная вакцинация, приводящая к исключению из циркуляции распространенных генетических линий пневмококка, обладающих большей способностью к адаптации за счет энерго-обеспечивающих систем (АТФ, НАДФН, фосфотрансферазные системы), может приводить к распространению генетических линий, характеризующихся большей вирулентностью.

Показана ассоциация варибельности гена *strH*, кодирующего экспрессируемую на поверхности клеточной стенки экзо- $\beta$ -D-N-ацетилглюкозаминидазу, с инвазивностью. Установлено, что белок StrH – потенциальный кандидат в мишени для белковой антипневмококковой вакцины.

### **Теоретическая и практическая значимость работы**

В целом полученные в ходе проведенного исследования новые научные данные значительно дополняют и уточняют существующие представления об эволюции *S. pneumoniae*, его метаболических особенностях и возможностях адаптации.

Получена так называемая «точка отсчета», отражающая состояние структуры популяции *S. pneumoniae* на момент начала антипневмококковой вакцинации в России. Полученные результаты позволят в дальнейшем получить информацию об изменениях, касающихся как эпидемиологической ситуации в России, так и генетических процессов, детерминирующих ответ популяции пневмококка на методы профилактики и лечения.

Используемые в исследовании биоинформатические методы филогенетического анализа клинических штаммов бактерий были внедрены в ФГБУ ДНКЦИБ ФМБА России.

Результаты использования предложенной методики исследования используются при чтении лекций и проведении практических занятий в ФГБОУ ВО КрасГМУ им. проф. В.Ф. Войно-Ясенецкого Минздрава России.

Депонирование последовательностей сиквенированных геномов *S. pneumoniae* в GenBank (международный уровень внедрения).

### **Перспективы дальнейшего изучения**

Изучение и характеристика популяции пневмококка на фоне вакцинации в России позволит прогнозировать распространение эпидемически значимых клонов, несущих детерминанты резистентности к антибиотикам и вирулентности.

Понимание механизмов вирулентности и резистентности будет способствовать разработке эффективных белковых вакцин и новых антибиотиков.

### **Степень достоверности и апробация результатов**

Достоверность полученных результатов определяется репрезентативным объемом выборки изолятов *S. pneumoniae*. Выводы и рекомендации, сформулированные в работе, основаны на полученных результатах исследования.

Результаты исследований по теме диссертации доложены и обсуждены на 11 конференциях и научных форумах, в том числе международных; опубликовано 4 статьи в рецензируемых ВАК журналах.

Результаты и методики исследований по теме диссертации (алгоритм проведения филогенетического анализа) были внедрены в Федеральном государственном бюджетном учреждении «Детский научно-клинический центр инфекционных болезней Федерального медико-биологического агентства», отдел медицинской микробиологии и молекулярной эпидемиологии, г. Санкт-Петербург, ул. Профессора Попова 9 (научно-практическая деятельность) и Федеральном государственном бюджетном образовательном учреждении высшего образования «Красноярский государственный медицинский университет имени профессора В.Ф. Войно-Ясенецкого» Министерства здравоохранения Российской Федерации, Кафедра микробиологии имени доцента Б.М.Зельмановича, г. Красноярск, Партизана Железняка ул., д. 1. (образовательная деятельность). Наименование внедрения: использование биоинформатических методов

филогенетического анализа клинических штаммов *Streptococcus pneumoniae*, основанного на сравнении последовательностей генов схемы мультилокусного сиквенс-типирования, а также на сравнении данных полногеномного секвенирования.

### **Методология и методы работы**

Методология диссертационной работы состояла в организации и выполнении фундаментального исследования по характеристике популяции *S. pneumoniae*, принадлежащих эпидемическим генетическим линиям. Анализ научной литературы проведен формально-логическими методами. В работе использованы молекулярно-биологические, эпидемиологические, биоинформатические, статистические методы исследований.

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

1. Мировая популяция *S. pneumoniae* представлена тремя глобальными группами А, В1 и В2. В настоящее время в популяции пневмококка эволюционирует генетически гетерогенная группа В, представители которой ассоциируются преимущественно с разными серотипами, но тем не менее, имеют характерные общие метаболические особенности. Наибольший вклад в деление на глобальные группы вносят гены, кодирующие: сигнальную пептидазу I, участвующую в процессинге секретируемых белков; гены глюкокиназы и глюкозо-6-фосфат-1-дегидрогеназы, регулирующие поток глюкозы в клетку. В группе А доминируют серотипы 23F, 19F, 14, 23A; в группе В1 – серотипы 11A, 19F, 19A, 1, 9N; в группе В2 – серотипы 6A/B/E, 3, 19A, 7F, 5. Группа В2 ассоциируется преимущественно с серотипами, не входящими в состав вакцины ПКВ7.

2. Происходившие в 2000 – 2010 гг. изменения в структуре популяции пневмококков, циркулировавших в Российской Федерации, были обусловлены изменениями в структуре глобальной популяции в ответ на массовую вакцинацию ПКВ7 в различных регионах мира. Происходивший в этот период рост устойчивости к бета-лактамам и макролидным антибиотикам связан с преимущественным глобальным распространением серотипа 19A, не входящего в состав ПКВ7.

3. Принадлежность к генетическим линиям определяется структурой базовых систем метаболизма: типа системы рестрикции-модификации ДНК; CiaH сенсорной гистидин-киназы (компонент двухкомпонентной сигнальной системы CiaHR, регулятор компетентности / вирулентности); АссС - ацетил-СоА-карбоксилазы (участвует в первых реакциях синтеза жирных кислот). Возможно, АТФ-синтазный комплекс и

компоненты фосфотрансферазной системы детерминируют формирование типов капсулы и метаболический тип *agaD-1* и *atpG-1* является оптимальным для роста и адаптации большинства пневмококков. Серотип-специфичная вакцинация, приводящая к исключению из циркуляции распространенных генетических линий пневмококка, обладающих большей способностью к адаптации за счет энерго-обеспечивающих систем (АТФ, НАДФН, фосфотрансферазные системы), может приводить к распространению генетических линий, характеризующихся большей вирулентностью.

4. Особенности метаболизма углеводов и ароматических аминокислот детерминируют регуляцию клеточных процессов пневмококка и фенотипические особенности, в том числе устойчивость к антибиотикам и вирулентность. Расщепление популяции по вариантам гена *strH* коррелирует с кластеризацией по компонентам углеводного метаболизма (*gnd*, *dexB*), путям синтеза ароматических аминокислот (*aroE*), регуляторному гену *relA* (ГТФ-пирофосфокиназа), генам синтеза пептидогликана и клеточного деления (*murD*, *pbp1A*), регуляторным генам экспрессии полисахаридной капсулы (*wzg-wzh-wze*).

5. Белок StrH (экзо-β-D-N-ацетилглюкозаминидаза) оказывается на стыке метаболических путей и вариабельность гена *strH*, кодирующего экспрессируемую на поверхности клеточной стенки экзо-β-D-N-ацетилглюкозаминидазу, ассоциируется с инвазивностью и источником выделения (кровь, ликвор, жидкость среднего уха, носоглотка). С источником выделения «кровь» ассоциируется ограниченное число вариантов гена *strH* и генетических линий. Белок StrH может быть кандидатом в мишени для белковой антипневмококковой вакцины. Возможно, это позволит модулировать вирулентный потенциал пневмококка.

#### **Личное участие автора в получении результатов**

Все этапы экспериментальной и аналитической работы, а также литературный обзор, выполнены самостоятельно. Отдельные этапы работы выполнены вместе с коллегами отдела медицинской микробиологии и молекулярной эпидемиологии ФГБУ ДНКЦИБ ФМБА России.

#### **Соответствие диссертации паспорту научной специальности**

Диссертационная работа Цветковой И.А. «Генотипическая характеристика *Streptococcus pneumoniae* эпидемически значимых генетических линий» представлена к защите по специальности 1.5.11 (микробиология) и соответствует формуле

специальности, охватывающей проблемы теоретических основ жизнедеятельности микроорганизмов (наследственности, изменчивости, метаболизма, закономерности взаимоотношения с окружающей средой и живыми организмами, распространения в природе, взаимодействия с факторами внешней среды и живыми организмами) в областях: «Эволюция микроорганизмов, установление их филогенетического положения», «Исследование микроорганизмов на популяционном уровне».

### **Объем и структура диссертационной работы**

Диссертация изложена в двух томах (Том 2 – Приложения). Первый том содержит 175 страниц машинописного текста и состоит из общей характеристики работы, 3 глав, заключения и выводов, списка публикаций по теме диссертации, списка сокращений и списка использованной литературы. Список литературы включает 347 источников литературы. Том 2 содержит 254 страницы машинописного текста и состоит из 9 приложений. Диссертация иллюстрирована 28 таблицами и 119 рисунками (первый том включает 20 таблиц и 43 рисунка, Том 2 включает 8 таблиц и 76 рисунков).

## ГЛАВА 1. ОБЗОР ЛИТЕРАТУРЫ

*S. pneumoniae* является комменсальным микроорганизмом для человека, входит в состав назофарингеальной микробиоты. До эры антибиотиков пневмококк считался одним из важнейших патогенов человека. Открытие сульфаниламида, пенициллина, а затем и других антибиотиков, к которым пневмококки длительное время сохраняли высокую чувствительность, снизило их значимость по сравнению со стафилококками и другими более устойчивыми к антибиотикам возбудителями. Интерес к пневмококковой инфекции повысился в последние десятилетия в связи с изучением заболеваемости, ростом лекарственной устойчивости пневмококков и созданием конъюгированных вакцин.

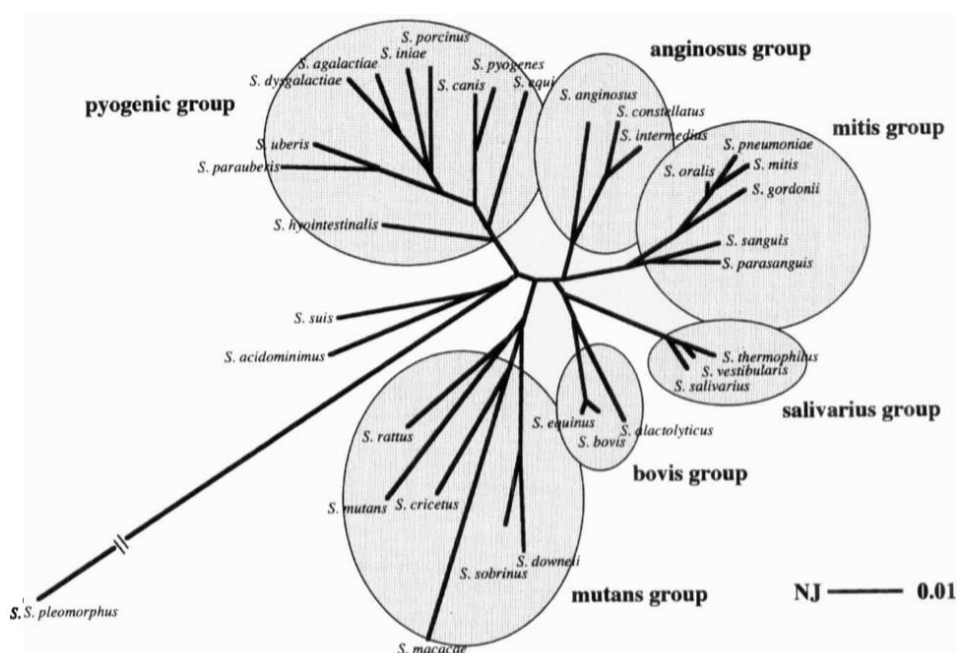
### 1.1. Современная таксономическая классификация стрептококков

Современная классификация стрептококков отражает филогенетические взаимосвязи между видами и основана на сравнении последовательностей гена 16S-рРНК [23, 24]. Стрептококки принадлежат семейству *Streptococcaceae* из филы *Firmicutes: Bacteria / Terrabacteria group / Firmicutes / Bacilli / Lactobacillales / Streptococcaceae / Streptococcus*. Стрептококки группы «viridans» (Рисунок 1) образуют основной компонент микробиоты полости рта человека и могут быть классифицированы на четыре группы: *Anginosus*, *Bovis*, *Mutans* и *Mitis* (*S. pneumoniae*, а также преимущественно комменсальные виды *S. oralis*, *S. mitis*, *S. gordonii*, *S. infantis*, *S. pseudopneumoniae*, *S. sanguinis*, *S. parasanguinis*, *S. cristatus*, *S. peroris*, *S. australis*, *S. sinensis*) [23, 24]. Последовательностей гена 16S-рРНК достаточно, чтобы выделить основные группы стрептококков, но для дифференциации пневмококка и его ближайших родственников *S. pneumoniae*, *S. mitis*, *S. pseudopneumoniae* и *S. oralis* необходимо учитывать также последовательности других консервативных генов [23, 25-27].

### 1.2. Потребности в питании и метаболические характеристики *S. pneumoniae*

Особенностью всех молочнокислых бактерий является отсутствие функциональной респираторной системы и неспособность к аэробному типу дыхания [24]. Представители рода *Streptococcus* являются грамположительными, не способными к образованию спор, каталазо- и оксидазо-отрицательными, цитохром-негативными

факультативными анаэробами. Среди бактерий, способных к росту в аэробных условиях, стрептококки уникальны тем, что они не синтезируют АТФ с участием электронных транспортных систем [24]. В геномах пневмококков отсутствуют гены, кодирующие ферменты цикла Кребса и цепи переноса электронов [24]. У них утрачена способность к синтезу порфиринов, цитохромов и каталазы [24].



**Рисунок 1 – Филогенетические взаимосвязи между 34-мя видами *Streptococcus*, по данным сравнения гена 16S-рРНК из Kawamura Y., Hou X., et al. 1995 [23].**

Энергетические потребности пневмококка удовлетворяются за счет ферментации углеводов. Однако в геноме пневмококков отсутствуют ферменты, кодирующие путь Энтнера-Дудорова, являющийся альтернативным вариантом гликолиза [24]. *S. pneumoniae* характеризуется гомоферментативным типом брожения. Основным катаболитом углеводного обмена в микроаэрофильных условиях у пневмококков является лактат. Лактатдегидрогеназа превращает пируват в лактат и регенерирует NADH в NAD<sup>+</sup>. У пневмококка присутствуют дополнительные ферменты (фосфотрансацетилаза, ацетокиназа и NADH-оксидаза) для превращения пирувата в ацетат в АТФ-синтазной реакции [24].

Пневмококки не способны к синтезу предшественников большинства аминокислот. При анализе генома пневмококков были выявлены неполные пути для биосинтеза Cys, Gly, Gln, Glu и His, Lys и Pro [28, 29]. Потребность пневмококков в азот-содержащих соединениях реализуется либо за счет транспортеров, участвующих в



захвате аминокислот, экскретируемых другими комменсальными бактериями, либо за счет экспрессии на поверхности клетки пептидаз и протеаз, воздействующих на белки тканей хозяина [30-33].

Таким образом, пневмококки требовательны к питательной среде и зависят от представленных в ней углеводов, аминокислот и кофакторов. Пневмококки обладают большим числом ферментов и транспортных систем, участвующих в поглощении углеводов, аминокислот и других соединений [34]. Треть транспортных систем пневмококка, в том числе АТФ-связывающие кассетные транспортеры (АВС) и классические фосфотрансферазные системы, способны участвовать в поглощении около 12 различных углеводов, метаболизирующихся путем гликолиза, а также с помощью пентозофосфатного пути [28, 29, 34]. В многочисленных исследованиях была продемонстрирована важность транспортных систем для адаптации и вирулентности пневмококка.

При культивировании пневмококков используют питательные среды, обогащенные белковыми компонентами и дефибринированной кровью животных. Для первичного выделения и культивирования пневмококков используют кровяной агар с добавлением 5%-й дефибринированной бараньей или лошадиной крови [35]. Рост пневмококков значительно улучшается при добавлении к питательной среде глюкозы (0,1%), сыворотки. Рост пневмококка стимулирует повышенное содержание в атмосфере CO<sub>2</sub> (5-10%). Оптимум рН среды – 7,8. Оптимальная температура роста 37<sup>0</sup>С, максимальная 42<sup>0</sup>С, минимальная 25<sup>0</sup>С [35].

На плотных питательных средах пневмококк образует мелкие (1-2 мкм в диаметре), округлые, блестящие, бесцветные колонии с ровным краем, мягкой консистенции, которые в большинстве случаев через 24 часа инкубации имеют сферическую форму с уплощенным центром, образующимся в результате аутолиза (шероховатая R-форма). На 5% кровяном агаре вокруг колонии пневмококка характерно образование зеленыщей зоны альфа-гемолиза, вызванное действием перекиси водорода, которая образуется при росте пневмококков в присутствии кислорода [36]. При этом происходит окисление ионов железа гемоглобина с формированием метгемоглобина [37]. Наиболее вирулентные капсульные бактерии могут располагаться в виде капель росы, в последующем принимая сливной рост (мукоидная M-форма). Гладкие, компактные, точечные колонии (S-форма) преобладают у авирулентных штаммов. Рост

*S. pneumoniae* в жидких питательных средах характеризуется диффузным помутнением бульона без образования пленки. Пневмококки из жидких питательных сред под микроскопом визуализируются цепочками средней длины.

Биохимическая активность пневмококка во многом совпадает с активностью других зеленеющих стрептококков. Штаммы *S. pneumoniae* разлагают глюкозу, лактозу, мальтозу, сахарозу. Не ферментируют маннит, салицит и сорбит. Не способны восстанавливать нитрат до нитрита.

Видовая специфичность *S. pneumoniae* выражается в чувствительности к оптохину и солям желчных кислот (дезоксихолату и турохолату натрия) [35]. Вместе с тем, в последние годы появились изоляты пневмококка, устойчивые к оптохину [38].

### **1.3. Молекулярная эпидемиология *Streptococcus pneumoniae***

#### **1.3.1. Задачи молекулярной эпидемиологии**

Молекулярная эпидемиология позволяет прогнозировать распространение пневмококка с учетом его генетической изменчивости, а также закономерности эволюции в результате селективных воздействий, таких как использование противомикробных препаратов или вакцинация. С этой целью проводится определение принадлежности патогенного штамма к эпидемически значимым генетическим линиям с помощью методов массового генотипирования, выявление вирулентных и резистентных к антибиотикам штаммов, а также выявление новых клонов, с помощью фенотипических и генотипических методов [18, 39].

#### **1.3.2. Методы молекулярной эпидемиологии**

К фенотипическим относят немоллекулярные методы типирования микроорганизмов: биотипирование (идентификация различий между штаммами по способности использовать в качестве субстрата те или иные питательные вещества), серотипирование (идентификация особенностей антигенной структуры), а также определение чувствительности к бактериофагам, бактериоцинам и антибактериальным препаратам.

Фенотипическая характеристика актуальна и в настоящее время (преимущественно, антибиотикотипирование, серотипирование). Однако штаммы, принадлежащие различным эпидемическим генетическим линиям, могут проявлять одинаковые фенотипические свойства. Поэтому к настоящему времени в рутинной практике типирования широкое распространение получили методы, основанные на

анализе ассоциаций протеомного профиля (в частности, MALDI-TOF, Matrix-assisted laser desorption ionization time-of-flight — времяпролетная лазерная десорбционная ионизация, ассоциированная с матрицей). Данные методы обладают высокой разрешающей способностью при видовой идентификации штаммов, но часто недостаточны для внутривидового типирования.

В последние десятилетия также интенсивно развивались молекулярно-генетические методы типирования: полимеразная цепная реакция (ПЦР), анализ длин рестрикционных фрагментов с помощью метода гель-электрофореза в пульсирующем поле (PFGE – Pulsed-Field Gel Electrophoresis), риботипирование, ПЦР с идентификацией повторяющихся последовательностей (MLVA, multiple-locus variable number tandem repeat analysis – мультилокусный анализ варибельности числа тандемных повторов) [18]. Длительное время золотым стандартом для генотипирования пневмококков был метод мультилокусного сиквенс-типирования (multilocus sequence typing, MLST) [18], основанный на секвенировании фрагментов 7 генов домашнего хозяйства (длиной около 500 п.н.). Каждой уникальной комбинации аллелей присваивается уникальный тип последовательности (сиквенс-тип, sequence type, ST). Схема MLST-типирования пневмококка включает гены: *aroE*, *gdh*, *gki*, *recP*, *spi*, *xpt*, *ddl* (кодирующие шикимат-киназу, глюкозо-6-фосфат-1-дегидрогеназу, глюкозо-киназу, транскетолазу, сигнальную пептидазу 1, ксантин-фосфорибозилтрансферазу и D-аланин-D-аланин лигазу, соответственно). Выбор генов для MLST-типирования был основан на том, что данные локусы считались нейтральными с точки зрения отбора, т.е. предполагалось, что изменения в нуклеотидных последовательностях накапливаются медленно [18]. Однако это предположение оказалось верным не для всех локусов. В частности, ген *ddl* задействован рекомбинациями, поскольку расположен в геноме пневмококка на расстоянии всего 783 п.н. ниже гена *pbp2b*, кодирующего пенициллин-связывающий белок 2b и являющегося мишенью для отбора [40]. Изоляты, генотипированные методом MLST, могут быть отнесены к клональным комплексам (clonal complex, CC) с помощью алгоритма eBURST [18, 41].

В настоящее время метод полногеномного секвенирования стал референсным методом генетического типирования штаммов - благодаря высокой разрешающей способности, а также возможности детального анализа генома. В базе данных PubMLST доступно описание более 48000 изолятов пневмококка, более 15200 уникальных ST, а

также коллекция из 9147 геномов (<https://pubmlst.org>). Развитие методов генотипирования привело к необходимости установления критериев, допускающих генетические изменения, произошедшие посредством мутации или рекомбинации, для идентификации штаммов, представляющих один и тот же клон. Для решения этой проблемы были установлены критерии для каждого метода типирования. Например, при MLST-типировании к клону относят однолокусные варианты (single locus variants, SLV). При анализе данных полногеномного секвенирования наиболее часто используемые алгоритмы сфокусированы на участках генома, общих для всех анализируемых штаммов (ядерная часть генома), при этом исключаются области с идентифицированными событиями рекомбинации. Данный подход позволяет сравнивать высоко консервативные последовательности генома, что позволяет установить генетическую линию на основе вертикального наследования генов. Однако необходимо заметить, что даже конкретные однонуклеотидные полиморфизмы (single nucleotide polymorphisms, SNP), а также приобретение дополнительных генетических элементов, не относящихся к ядерному геному, могут привести к значительным изменениям фенотипа. Поэтому в настоящее время стандартным подходом генетического анализа микробной популяции является анализ пангенома, включающего ядерный и дополнительный геном: оценка процентного соотношения консервативного ядерного и дополнительного генома, оценка размера пангенома, а также наличия/отсутствия генов в геномах анализируемой популяции.

В многочисленных исследованиях, анализ данных генотипирования выявил следующие характеристики популяций пневмококка: многие ST принадлежат ограниченному числу больших клонов (clonal complexes, CC), представители которых были выделены в различных регионах мира [42]; изоляты, принадлежащие к одному и тому же CC, часто экспрессируют один и тот же серотип, однако, изоляты одного и того же серотипа могут принадлежать разным CC [21, 43, 44]; представители ряда CC ассоциируются с резистентностью к антибиотикам [42, 45], определенные генотипы могут обладать большим потенциалом вызывать инвазивные ПИ [21, 46, 47].

#### **1.4. Биологические процессы, лежащие в основе эволюции *S. pneumoniae***

Популяция бактерий представлена множеством совместно циркулирующих генетических линий. Генетическое разнообразие лежит в основе фенотипической вариации свойств, таких как резистентность к антибиотикам, вирулентность и

антигенный профиль. Чтобы ответить на вопрос, как возникают такие структуры популяций, необходимо понимание, что движет возникновением разнообразия среди бактерий. В дополнение к значительной вариации аллелей генов ядерного генома, генетические линии *S. pneumoniae* содержат «геномные островки» («genomic islands», GI) – генетические локусы, встречающиеся только в определенной субгруппе популяции [48]. Геномные островки могут переноситься между клетками посредством естественной трансформации, а также вместе с мобильными генетическими элементами (МГЭ).

#### **1.4.1. Горизонтальный перенос генов**

Все три механизма, опосредующие горизонтальный перенос генов у бактерий (трансформация, конъюгация и трансдукция) [49], являются распространенными в пневмококковой популяции и сыграли решающую роль в эволюции генома пневмококка.

##### **1.4.1.1. Трансформация у пневмококков**

Передача хромосомной ДНК посредством трансформации возможна между родственными бактериями, обладающими механизмами естественной компетентности. Геном реципиента может приобрести от единичных до множественных рекомбинационных блоков [50-52].

У пневмококков спонтанное развитие состояния компетентности происходит на ранней фазе экспоненциального роста [53], в результате активации межклеточного сигнального каскада, кодируемого опероном *comABCDE* («гены ранней компетентности») [54, 55]. Стимулирующий компетентность пептид, CSP (ComC) секретируется компетентной клеткой через ABC-транспортер ComAB и связывается со своим рецептором – гистидинкиназой ComD. Фосфорилированный регулятор ответа ComE усиливает транскрипцию оперона *comABCDE* и гена *comX* (кодирует сигма-фактор X, SigX) [54, 55]. Фактор SigX регулирует экспрессию «генов поздней компетентности», кодирующих продукты, связанные с процессингом и репарацией ДНК, а также белки, участвующие в уничтожении некомпетентных клеток [56, 57]. Период полураспада фосфорилированного ComE является коротким, он может оказывать свое влияние только в случае, когда скорость фосфорилирования превышает скорость деградации – состояние, соответствующее концентрации внеклеточного CSP 1-10 нг/мл [58].

В целом, в ответ на повышенный уровень внеклеточного CSP изменяется экспрессия более 180 генов [59]. Сообщалось о совместной колонизации пневмококков, экспрессирующих различные варианты CSP (феротипы), и частых событиях горизонтального переноса генов между ними [60, 61]. Считают, что при совместной колонизации пневмококки наиболее распространенного феротипа первыми достигают состояния компетентности и это приводит к лизису остальных типов клеток (показано в экспериментах *in vitro*) [62].

Для лизиса некомпетентных клеток компетентные клетки экспрессируют белок CbpD (амидаза/пептидаза, расщепляет пептидные мостики пептидогликана клеточной стенки [63]), а также CibAB (двухпептидный бактериоцин, увеличивает проницаемость клеточных мембран). Компетентные клетки защищены от литической активности CbpD продукцией белка ранней фазы компетентности ComM [62], от активности бактериоцина – продукцией белка CibC [64]. Гидролазы клеточной стенки LytA (амидазная активность) и LytC (лизоцимная активность) также участвуют в лизисе некомпетентных клеток.

Процесс гомологичной рекомбинации происходит в несколько стадий. Нативная донорская ДНК связывается с поверхностью компетентной клетки через ДНК-связывающий белок, разрезается на двухцепочечные фрагменты размером от 3 до 9 тыс. п.н. [65-67]. Однонитевая ДНК транслоцируется через мембрану по белковому каналу ComEC, этому процессу способствует ДНК-транслоказа ComFA [68]. После интернализации, около 75% одноцепочечных фрагментов в клетке деградируют [69], тогда как другие участвуют в «атаке» на хромосому, вытесняя гомологичные сегменты [70]. Неразрушенная однонитевая ДНК защищена от клеточных нуклеаз комплексом с белком SsbB – «комплекс затмения», eclipse complex. ДНК в комплексе с SsbB не способна к рекомбинации. Внутри клетки может присутствовать до 70000 молекул SsbB, защищающих до ~1,15 Мб интернализированной ДНК [71]. Считают, что такой объем интернализированных нуклеиновых кислот может служить резервуаром для последовательных циклов рекомбинации. Освобождение однонитевой ДНК происходит через связывание с белками DprA и RecA [72].

#### **1.4.1.2. МГЭ-опосредованный перенос генетического материала между бактериями**

МГЭ-опосредованный перенос осуществляется с помощью таких событий как трансдукция и конъюгация [49, 53]. У *S. pneumoniae* были охарактеризованы три типа

МГЭ: фаги [73, 74]; плазмиды [75, 76] (из которых только два примера известны у пневмококков); а также интегративные конъюгативные элементы (ИКЭ) [58, 77].

#### 1.4.1.2.1. Конъюгация

##### Геномные островки, переносимые ИКЭ пневмококка

У пневмококков, принадлежащих различным серотипам и сиквенс-типам, детерминанты резистентности к антибиотикам часто ассоциированы с *Tn5253* и *Tn916* [78, 79]. Мобильный элемент *Tn916* содержит ген *tet(M)*, обеспечивающий устойчивость к тетрациклину [79]. В геномах резистентных к эритромицину изолятов были идентифицированы *Tn916*-подобные мобильные элементы *Tn2010* и *Tn6002*, содержащие ген *erm(B)*, обеспечивающий резистентность к макролидам, линкозамидам и стрептограмину В. Мобильный элемент *Tn2010* содержит генную кассету *MEGA*, кодирующую гены эффлюкса макролидов [80, 81].

В геномах пневмококков, принадлежащих клону Spain23F-1, был идентифицирован *Tn5253*-подобный элемент, размером ~ 81 т.п.н. [77]. *Tn5253* представляет собой сложносоставной мобильный элемент, сформированный путем интеграции *Tn916*-подобного элемента в *Tn5252*. При этом *Tn5252* делится *Tn916*-вставкой на большую проксимальную область и меньшую дистальную область. Данная комбинация ИКЭ распространена у стрептококков [77]. Сайт интеграции *Tn5253*, как и многих других ИКЭ стрептококков, часто находится рядом с 3'-концом гена *rplL* [77]. В геномах многих пневмококков идентифицированы остатки дистальной области элемента *Tn5253* и фага φMM1, содержащие ген цитозинметилтрансферазы [77]. Возможно, присутствие в структуре ИКЭ гена цитозинметилтрансферазы позволяет данным мобильным элементам уклоняться от рестрикционных систем реципиента.

Гены, переносимые *Tn5252*, могут восстанавливать некоторые функции SOS-ответа у пневмококков. В ядерном геноме пневмококка отсутствует компонент системы SOS-ответа – ген хеликазы *uvrD*. Однако гены *uvrABC* присутствуют у всех пневмококков и захват *uvrD* путем горизонтального переноса с элементом *Tn5252* потенциально может воссоздать механизм эксцизионной репарации нуклеотидов [77]. Этот механизм важен для восстановления повреждений ДНК, вызванных перекисью водорода [82]. Приобретение гена *uvrD* может способствовать увеличению толерантности пневмококка к активным формам кислорода и колонизации носоглотки [83]. Другой важной частью SOS-ответа является репарация мутаций, которая требует

снижения точности репликации ДНК-полимеразы III, вызываемого взаимодействием с белковым комплексом UmuC-UmuD. *Tn5252* несет оперон, содержащий гены *umuC* / *umuD* [77, 78], экспрессия которого способствует повышению устойчивости бактерий-хозяев к ультрафиолету [84].

В геномах стрептококков были описаны островки патогенности PPI-1 (putative pathogenicity island 1), которые, возможно, когда-то были частью *Tpn5253* (3'-область, ~200 п.н., идентична *Tn5253* > 90%) [77, 85, 86]. У штамма *S. pneumoniae* ATCC 700669 эта область характеризуется низким содержанием GC и соответствует размеру около 30 т.п.н., содержит систему токсин-антитоксин системы омега-эпсилон-зета *pezAT*, а также кластер из трех генов механизма конъюгации *Tn5252*, включая релаксазу и MobC-домен-содержащий белок [77]. Модуль генов, ассоциированных с конъюгативным механизмом, был в различной степени утрачен в этом локусе разными штаммами. Напротив, 5'-область PPI-1, содержащая *pit*-оперон, кодирующий транспорт железа, имеющий важное значение для вирулентности *S. pneumoniae* [87], сохраняется у многих штаммов [77]. Предполагают, что *pit*-оперон был приобретен исходной версией ИКЭ элемента, поскольку, несмотря на повсеместное распространение среди пневмококков, *pit*-оперон, не был найден у других стрептококков [87].

*Tn5253* содержит кластеры генов синтеза лантибиотика [77, 88]. Лантибиотики (лантионин-содержащие антибиотики) – небольшие секретлируемые циклические пептиды, содержат кольца лантионина, образованные стереоспецифическим внутримолекулярным взаимодействием цистеина и дегидратированными остатками серина или треонина [89]. Лантибиотики часто функционируют как бактериоцины. В зависимости от типа, они могут действовать либо посредством ингибирования трансгликозилирования пептидогликана, либо посредством образования пор в клеточных мембранах; также известно, что лантибиотики действуют как биосурфактанты и ингибиторы фосфолипазы A2 [89]. Большинство генных кластеров синтеза лантибиотика у пневмококка находятся рядом с генами, кодирующими регуляторы транскрипции типа *plcR* (идентифицированные у разных видов *Bacillus*) и секретлируемые пептиды [77]. Такие транскрипционные факторы, в сочетании с пептидными аутоиндукторами, могут образовывать минимальную чувствительную к кворуму систему [90]. Подобная межклеточная сигнализация может быть необходима для вариабельного микроорганизма, чтобы отличать клональных родственников от



других родственных штаммов [90]. Кроме того, регуляция кластера генов синтеза лантибиотика, кодируемого ИКЭ, зависит от плотности клеток, что может быть выгодно для самого мобильного элемента, поскольку подавление выработки антимикробного соединения до тех пор, пока ИКЭ не насыщает доступную популяцию потенциальных хозяев, вероятно, будет способствовать его горизонтальному распространению [77].

Очевидно, что ИКЭ сыграли важную роль в эволюции пневмококка, поскольку они вносят значительную долю в пан-геном *S. pneumoniae* и могут способствовать быстрой адаптации бактерии-хозяина к новым селективным воздействиям.

#### 1.4.1.2.2. Трансдукция

Около 76% пневмококков являются носителями профагов [74, 91-93]. Профаг может потерять способность к реактивации, если происходит потеря или необратимая мутация его важных генов. Тем не менее, остатки генома профага могут находиться в хромосоме хозяина и передаваться его потомкам или другим бактериальным клеткам посредством других механизмов горизонтального переноса.

Во многих случаях бактериофаги могут быть ответственны за генетическую изменчивость между штаммами и передачу известных или предполагаемых факторов вирулентности [49, 77, 94, 95]. Одним из примеров фаг-опосредованной адгезии являются два поверхностных белка стрептококков, PblA и PblB, которые ответственны за адгезию к тромбоцитам [96]. Гены этих белков находятся на индуцибельном профаге SM1 [97]. Разрушение *pblA* или *pblB* приводило к снижению связывания *S. mitis* с тромбоцитами [96].

Многие фаги кодируют метилтрансферазы, защищающие фаговый геном от рестриктаз [96-100]. В некоторых случаях геном бактерии может содержать две сходные С5-метилтрансферазы, одна из которых кодируется фагом, а другая транспозоном (мультирезистентный штамм *S. pneumoniae* 8249, хозяин фага MM1). При этом геномы как фага, так и хозяина метилированы [100]. Существование обеих С5-метилтрансфераз позволяет предположить возможность коэволюции путем горизонтальной передачи. Интересно, что ИКЭ *Tn5252* не присутствует в геномах штаммов пневмококка, несущих фаги VO1 или НВ-3 [100].

Содержание профагов высоко вариабельно даже в пределах конкретных генетических линий пневмококка. Считают, что частый горизонтальный перенос неизбежно влечет за собой быструю диверсификацию анти-фаговых механизмов у

бактерий. Эти механизмы также могут препятствовать обмену другими геномными островками.

#### **1.4.2. Роль систем рестрикции-модификации в эволюции *S. pneumoniae***

##### **1.4.2.1. Распространенность систем рестрикции-модификации**

Считают, что благодаря своей способности распознавать «свое и чужое» – свойству, присущему иммунной системе высших организмов – системы рестрикции модификации (СРМ) функционируют как примитивные иммунные системы, распознающие и удаляющие чужеродную ДНК [101-103]. Фаги уклоняются от рестрикции либо модификацией своего генома (метилирование, гликозилирование и др.) [104, 105], либо уменьшением в геноме числа палиндромных последовательностей, распознаваемых рестриктазами [106-109]. Бактерии, в свою очередь, экспрессируют специфичные для модификации СРМ для удаления ДНК адаптированных фагов [110]. Обычно СРМ нечувствительны к переносу однонитевой ДНК, поскольку они могут связываться только с двуспиральной ДНК [111]. Однако импорт новых геномных островков посредством естественной трансформации также может ингибироваться СРМ – это происходит после интеграции последовательности в хромосому в том случае, если синтезированная комплементарная цепь неметелирована.

##### **1.4.2.2. СРМ у *S. pneumoniae***

На основании механизма действия, распознаваемых последовательностей-мишеней и структуры ферментов, было выделено четыре гетерогенных семейства СРМ [112], два из которых, СРМ типа I и СРМ типа II, представлены у всех пневмококков.

##### **СРМ типа II у *S. pneumoniae***

Все штаммы пневмококка содержат одну из трех СРМ типа II: DpnI, DpnII или DpnIII [111]. Компоненты СРМ типа II кодируются двумя генами и представляют собой независимо функционирующие ферменты, которые расщепляют либо метилируют ДНК внутри или рядом с сайтом распознавания. Все три СРМ узнают один и тот же мотив, GATC, но отличаются различным паттерном метилирования и специфичностью. *DpnI* расщепляет метилированные по аденину GATC (GmGATC) мотивы в двунитевой ДНК, *DpnII* расщепляет неметилированные GATC мотивы [111-114]. При этом система *DpnII* включает, помимо метилтрансферазы двунитевой ДНК, редкую метилтрансферазу однонитевой ДНК (*DpnA*), которая экспрессируется только во время состояния компетентности для защиты вновь захваченных локусов от рестрикции [111]. СРМ типа

II представляют собой эффективный способ в поддержании барьера для обратного взаимного переноса генетического материала между генотипами: если системы СРМ узнают различные мотивы, то двунитевая ДНК, которая включает только один из мотивов, будет удалена при переносе от донора к реципиенту.

DpnIII обнаружена в малой части популяции пневмококка [114, 115], преимущественно в генетической линии, ассоциированной с мультирезистентным клоном PMEN1 [114]. DpnIII расщепляет неметилованную и гемиметилованную двухцепочечную ДНК в последовательности 5'-GATC-3', а C5-цитозин-специфическая ДНК-метилаза узнает и метилирует цитозин в последовательности-мишени GATC [114]. DpnIII будет ограничивать инкорпорацию ДНК во время естественной трансформации от штаммов, имеющих DpnI или DpnII систему [114]. Действительно, штаммы генетической линии PMEN1 являются частыми донорами, но не реципиентами ДНК [116].

Кодирующие СРМ генные комплексы типа II, имеют тенденцию размножаться как эгоистичные элементы, тем самым стабилизируя генетическую линию хозяина [103, 104]. В случае их утраты невозможность поддержания статуса метилирования генома приводит к гибели клеток [105, 106, 117, 118]. Эгоистичное поведение не характерно для СРМ типа I, в которых метилтрансфераза и рестриктаза являются субъединицами одного и того же белкового комплекса – внутриклеточные соотношения этих ферментов не меняются при утрате СРМ, т.к. возможно только взаимозависимое их функционирование в качестве модуля [119, 120].

### **СРМ типа I у *S. pneumoniae***

СРМ типа I кодируется у пневмококков локусом *ivr* («inverting variable restriction»), генами *hsdR*, *hsdM* и *hsdS* (гены эндонуклеазы, метилтрансферазы и субъединицы специфичности, соответственно). Ген *hsdS* кодирует ДНК-связывающий белок, содержащий два распознающих мишень домена, каждый из которых узнает одну из двух частей мишени. Эти два домена разделены консервативной пептидной последовательностью, которая определяет число нуклеотидов между двумя элементами последовательности-мишени. Компоненты СРМ типа I могут функционировать в виде двух различных ферментативных комплексов: пентамерный фермент (две R, две M и одна S-субъединицы), обладающий эндонуклеазной активностью, или трехмерный фермент (две M и одна S субъединицы), обладающий только метилтрансферазной

активностью [121-124]. Метилирование происходит на двух цепях, ассиметричным образом (разделенных на две части последовательностей), что отражает симметричное взаимодействие белков рестрикции-модификации с цепями двойной спирали ДНК [124]. Ферменты полного пентамерного комплекса связывают и метилируют ДНК непосредственно в сайте распознавания, при этом инициируется транслокация двунитевых цепей ДНК за счет хеликазной активности R-субъединицы [121]. Эндонуклеазная активность запускается только при наличии препятствия для хеликазной активности СРМ комплекса (при конфликте с другой СРМ или возникновении обструкции – суперспирализованной ДНК) [124]. Событие рестрикции может происходить на расстоянии нескольких тысяч п.н. от сайта распознавания [122, 123]. Эндонуклеазное расщепление запускается неметилированной ДНК, метилаза активируется при наличии гемиметилированной ДНК.

СРМ типа I посредством фазовой вариации могут модифицировать свою специфичность к мишени метилирования за короткий период времени и переключать бактерии между альтернативными паттернами метилирования ДНК, посредством межгеномной рекомбинации генов *hsdS*, приводящей к формированию множества аллелей *hsdS* [121-126]. Лocus *ivr* содержит три независимых инвертированных повтора (330 bp, 85 bp и 15 bp), при этом только рекомбинации в последовательности повтора 15-bp интенсивно контролируются тирозиновой рекомбиназой (у штаммов, нокаутных по рекомбиназе, рекомбинации в двух больших повторах продолжаются) [126].

#### **Другие фазо-вариабельные СРМ типа I у пневмококка**

У *S. pneumoniae* фазовая вариация локуса СРМ типа I может происходить посредством транслокации ДНК между прямыми повторами. Этот процесс происходит в *tvr*-локусе («translocating variable restriction») [124]. Как и другие локусы СРМ типа I, *tvr*-локус содержит гены, кодирующие субъединицы пентамерного комплекса и ген рекомбиназы (*tvrR*), а также гены токсин-антитоксина системы, *tvrAT*. Однако, в отличие от инвертированных локусов, все гены *tvr*-локуса кодируются на одной и той же цепи. Роль токсин-антитоксина системы в данном локусе еще не была изучена детально, однако, считают, что данная система вовлечена в стабилизацию локуса СРМ на уровне популяции, способствуя уничтожению дочерних клеток, у которых часть локуса утрачена [124].

### 1.4.2.3. Фенотипы, ассоциированные с особенностями фазо-вариабельной системы типа I

Помимо рестрикции профага, фазо-вариабельное метилирование может определять множество эпигенетических состояний клеток бактерий, которые потенциально могут приводить к глобальным изменениям фенотипа [124, 126-128]. Таким образом, клетки могут использовать СРМ в различных биологических процессах для повышения своей относительной приспособленности в популяции.

В частности, были показаны различия в степени вирулентности мутантов *S. pneumoniae*, экспрессирующих только одну из возможных S-субъединиц, при моделировании инфекций на мышах [126]. Пневмококки с А-вариантом *hsdS* характеризовались сниженной способностью к колонизации эпителия при моделировании инфекций на фоне носительства, В-мутанты характеризовались сниженной вирулентностью при моделировании бактериемии [126]. Уменьшение вирулентности у В-мутантов коррелировало с низкой экспрессией оперона капсулы, хотя не было идентифицировано очевидного механизма, который мог бы объяснить данный эпигенетический эффект [126]. При мониторинге инвазивной инфекции, вызванной штаммами дикого типа (содержащими большую фракцию Е-варианта *hsdS*), со временем наблюдался сдвиг к увеличению числа А-вариантов, что позволило предположить влияние селекции определенных вариантов хозяином [126]. С другой стороны, среди более 600 изолятов пневмококка, полученных от больных менингитом и сепсисом, не было обнаружено ассоциаций *hsdS*-вариантов и паттернов метилирования с клиническими параметрами [128].

Известно, что прозрачные колонии ассоциируются с колонизацией, а мутные колонии – с инвазивными заболеваниями [129]. В двух независимых исследованиях было показано, что штаммы с вариантом А гена *hsdS* образуют мутные колонии, тогда как штаммы с вариантом В гена *hsdS* – более чем в 90% случаев образуют прозрачные колонии [122, 129]. Отличие в результатах двух исследований заключалось в том, что в одном из них был сделан вывод о неоспоримой ассоциации мутной морфологии колоний с единственным вариантом метилирования [122, 126], тогда как в другом исследовании мутанты с единственным вариантом метилирования давали как мутные, так и прозрачные колонии [124], но частота мутных и прозрачных колоний варьировала [124]. Был сделан вывод о том, что, помимо фазо-вариабельной системы,

дополнительные локусы должны быть вовлечены в регуляцию фенотипа мутности колоний у пневмококка.

Таким образом, способность к быстрому переключению между различными эпигенетическими состояниями может быть преимуществом, обеспечивающим популяции бактерий быстро возвращаться к фенотипическому разнообразию после селективного или неселективного действия сдерживающих рост факторов, которые обычно происходят при инфицировании хозяина. Несмотря на множество фактов о том, что изменения характера метилирования могут приводить к различиям в фенотипе вирулентности, детали эпигенетических механизмов в настоящее время еще не установлены.

#### 1.4.3. Короткие диспергированные повторы *S. pneumoniae*

Короткие диспергированные повторяющиеся элементы – пространственно разделенные геномные области сходной последовательности, обычно длиной менее 200 п.н.. В геноме *S. pneumoniae* были идентифицированы три семейства повторов (BOX, RUP и SPRITE) [130-133].

BOX-элемент – составной повтор, состоящий из последовательностей *boxA* и *boxC*, обычно разделенных переменным числом элементов *boxB*, имеющих тандемное расположение [133]. Некоторые крупные BOX-элементы кодируют открытые рамки считывания на обеих цепях хромосомы (при этом BOX-элементы могут транскрибироваться как продолжение как 5'-, так и 3'-областей оперонов), другие образуют составную структуру РНК с двумя Т-бок «рибосвитчами» («Т box riboswitches») [133]. Считают, что BOX-элементы могут действовать как регуляторные мотивы [133]. Было показано, что элементы *boxA* и *boxC* способны стимулировать экспрессию ниже расположенных генов, хотя элементы *boxB* могут оказывать противоположное ингибирующее действие, в зависимости от их ориентации [132]. Также предполагают, что BOX-элементы увеличивают частоту вариации фаз пневмококка (расщепление фенотипа на варианты), влияя на регуляцию соседних генов [134].

Простой повтор пневмококка («Repeat Unit of Pneumococcus», RUP) длиной 107 п.о., также присутствует в большом числе копий в геноме пневмококка. RUP имеет TIR, сходный с таковым у *IS630-Spn1* (другого транспозона, обычно встречающегося у *S. pneumoniae*) [131]. На основании распределения в геноме штамма *S. pneumoniae* TIGR4,

было предложена гипотеза о том, что RUP интегрируются преимущественно в элементы IS или рядом с ними [133], что позволяет предположить, что эти элементы могут служить для ограничения числа функциональных генов транспозазы в хромосоме.

Третий тип повтора – «*Streptococcus pneumoniae* Rho-Independent Terminator-like Element» (SPRITE) [133]. Структура SPRITE является менее плотной, чем структура BOX или RUP, и состоит из дуплекса длиной 18 п.н., за которым следует относительно богатый уридином (~ 48%) участок, который, вероятно, наделяет его свойствами Rho-независимого терминатора [133]. Однако структура повтора отличается тем, что как дуплекс, так и Т-богатый участок намного длиннее, чем размер ~ 10 п.н. (у типичных Rho-независимых терминаторов стрептококков) [135].

Короткие диспергированные повторы встречаются у пневмококка со стабильной частотой, с небольшими различиями между генетическими линиями [115]. Все три типа повтора присутствуют со сниженной плотностью в геноме близкородственного вида *S. mitis* и почти полностью отсутствуют у всех других видов стрептококков [133]. Поскольку и *S. mitis*, и *S. pneumoniae*, способны обмениваться ДНК, неясно, присутствовали ли эти повторы у их последнего общего предка или были приобретены после видообразования и в последствии распространялись горизонтально. Если следовать принципу максимальной парсимонии, то вывод заключается в том, что данные элементы распространились в геноме пневмококка после расхождения от наиболее удаленных представителей митис-группы [133].

#### **1.4.4. Ассоциация геномных островков с генетическими линиями**

Исследовательской группой Croucher N.J. и соавторами в США была проанализирована динамика популяции пневмококков, циркулирующих у детей-носителей с 2001 по 2007 гг. в штате Массачусетс, на фоне применения антипневмококковых конъюгированных полисахаридных вакцин [136, 137]. Исследователи показали, что генетические линии пневмококка отличались по СРМ типа II [115]. Наблюдался высокий уровень разнообразия в локусах СРМ типа I (*ivr* и *tvr*), даже между близко родственными изолятами [115].

Была выявлена тенденция к потенциальному видообразованию атипичного генотипа: изоляты одной из генетических линий отличались фенотипом (отсутствие капсулы), тропизмом к заболеванию (конъюнктивиты), а также метаболическими особенностями (отсутствие фукозного оперона) [115]. У изолятов данной генетической

линии отсутствовал *pit* оперон [115], ассоциирующийся с островком патогенности пневмококка 1 (PPI-1) [137], функциональный *cps*-локус синтеза капсулы и консервативный набор больших поверхностных белков [115].

Авторы показали, что генетические линии отличаются структурой 3'-вариабельной области локуса PPI-1 [137]. Известно, что гены локуса PPI-1 могут быть ответственны за отличительные свойства генетических линий – по крайней мере, в некоторых случаях вариация аллелей была ранее ассоциирована с различиями в вирулентности в моделях заболеваний у мышей [127].

Мобильные генетические элементы, не ассоциированные с PPI-1, насчитывали почти четверть всех ортологов. Это позволило предположить, что такие элементы не обязательно проявляют высокий уровень мобильности, но могут вносить вклад в стабилизацию различий между SCs [115]. Некоторые типы инсерционных последовательностей проявляли стабильную ассоциацию с генетическими линиями [115].

Авторы идентифицировали около 2260 локусов, ассоциированных с ИКЭ, при этом 16 сайтов инсерций находились в пределах ядерной части генома, в межгенных областях [115]. В частности, все гены, ассоциированные с резистентностью к антибиотикам, кодируемые мобильными генетическими элементами, были ассоциированы с Tn5253-подобными МГЭ [59].

Островки патогенности, ассоциированные с фагами, демонстрировали небольшие вариации в размере и сохраняли один и тот же базовый набор функций [115]. В отличие от ИКЭ, наблюдалось мало случаев, когда профаг устойчиво ассоциировался с генетической линией [115]. В частности, изотипы профага *φOXC141* были стабильно связаны с пневмококками, имеющими генотип, ассоциированный с серотипом 3, преобладающим в Массачусетсе [137]. Самое большое число геномных островков, для которых наблюдалась стабильная связь с сиквенс-кластером хозяина, было найдено у представителей генетической линии PMEN1, характеризующихся мультирезистентностью. У представителей PMEN1 островки представляли собой остатки профага, утратившего механизм мобилизации [59, 137].

Таким образом, не все геномные островки стабильны, разные типы демонстрируют разную динамику в популяции. Ассоциированные с профагами и ИКЭ геномные островки стабильны у представителей отдельных генетических линий [115].



Вероятно, вертикальная передача данных МГЭ способствовала клональному распространению ИКЭ-ассоциированной резистентности к антибиотикам [50]. В отличие от такой «симбиотической» долгосрочной ассоциации между ИКЭ и хозяевами, профаги являются гораздо менее устойчивыми [115]. Вероятно, деструктивный механизм передачи фага обуславливает высокую селективную стоимость клеток-хозяев и изоляты, несущие профаг, могут быть сравнительно быстро удалены из популяции.

### 1.5. *cps*-локус

Каждому серотипу соответствует уникальный оперон (*cps*-локус), кодирующий белки синтеза полисахаридной капсулы, и определяющий химическую структуру полисахарида и его серологические свойства. Размер *cps*-локуса варьирует в пределах от 10337 п.н. (серотип 3) до 30298 п.н. (серотип 38) [18, 138]. У всех пневмококков *cps*-локус находится в одном и том же месте хромосомы, между генами *dexB* (кодирует глюкан-1,6-альфа-глюкозидазу) и *aliA* (кодирует гипотетический внеклеточный олигопептид-связывающий белок), экспрессируется как единая транскрипционная единица [18]. Для всех, кроме двух серотипов пневмококка, 3 и 37, биосинтез капсулы осуществляется по *Wzy*-зависимому (полимеразо-зависимому) механизму [18, 138-140]. Для серотипов 3 и 37 синтез капсулы осуществляется посредством синтазо-зависимого механизма [18].

Все *Wzy*-зависимые варианты *cps*-локусов начинаются с четырех высококонсервативных генов *cpsA-D* (*wzg*, *wzh*, *wzd* и *wze*), кодирующих лигазу и три киназы, а также содержат гены *cpsE/wchA*, *wzy* и *wzx* [18]. Показано, что последовательности регуляторных генов *cpsB*, *cpsC* и *cpsD* образуют два отдельных кластера [141]. При этом один из этих кластеров ассоциируется с носительством, а другой – с инвазивными заболеваниями [141].

Остальные гены серотип-специфичны, кодируют гликозил- и ацетилтрансферазы и ферменты, необходимые для синтеза углеводов-предшественников. Биосинтез пептидогликана начинается со сборки повторяющихся остатков моносахаридов на внутренней стороне мембраны клетки [18, 138, 142]. Для большинства серотипов иницирующим углеводом является глюкозо-1-фосфат (Glc-1-P), и реакция катализируется мембранным белком гликотрансферазой *CpsE/WchA* [18, 138, 142]. Дополнительные гликозилтрансферазы, кодируемые *cps*-локусом, добавляют к растущей олигосахаридной цепи другие моносахариды. Растущая цепь переносится на

внешнюю сторону мембраны с помощью флиппазы Wzx, где Wzy-полимераза полимеризует эти олигосахариды, добавляя соответствующие субъединицы. Считают, что прикрепление к клетке происходит с помощью лигазы CpsA [18, 138, 142].

### 1.5.1. Эволюция генов *cps*-локуса

Изменение типов капсул и *cps*-локусов, наблюдаемое в популяциях *S. pneumoniae*, является результатом сочетания событий горизонтального переноса, перемещения подвижных генетических элементов и мутаций [143-145]. Существует гипотеза, что, по крайней мере, часть *cps*-локуса была приобретена пневмококком от других видов путем горизонтального переноса [138, 141], так как в отличие от первых четырех генов *cps*-локуса и присутствующих в *cps*-локусе генов биосинтеза углеводов, имеющих типичное для пневмококков содержание GC [138], «серотип-специфичные» гены (особенно *wzy* и *wzx*) имеют более АТ-насыщенные последовательности [138]. Кроме того, гены *cps*-локуса и фланкирующие его гены *dexB* и *aliA* могут быть окружены полными или частичными последовательностями мобильных генетических элементов и в ряде случаев находятся рядом с интронами группы II (серотипы 19F, 25F, 25A и 38) [138, 146-149].

Серотипы одной серогруппы, 19F и 19В, имеют сходные структуры капсул, но значительно отличаются структурой *cps*-локуса. *Cps*-локус серотипа 19В содержит область размером 10,5 т.п.н., включающую пять генов, которая не является общей с *cps*-локусом серотипа 19F [150]. Низкое содержание GC (27,2–29,7 мол.%) позволяет предположить, что приобретение этих генов произошло в результате событий горизонтального переноса [151]. Продукты этих генов кодируют трансферазы и транспортные белки, необходимые для добавления специфических моносахаров в структуру капсулы серотипа 19В (дополнительный углеводный остаток в полисахаридной цепи, а также дополнительную дисахаридную боковую цепь) [150]. Серотипы 44 и 46 отличаются от 12F только генами, инсерцированными вместе с транспозазой [138].

В некоторых случаях между штаммами разных серогрупп может быть обнаружена общая гомология типоспецифичных областей. Например, часть *cps*-локуса серотипа 23F демонстрирует гомологию с *cps*-локусом серотипа 19F [149]. Было показано, что гены в этом регионе кодируют белки, связанные с синтезом и процессингом dTDP-рамнозы, которая присутствует в капсулах как серотипа 23F, так и серотипа 19F [149].

Одиночные нуклеотидные замены в *cps*-локусах часто ответственны за серологические различия капсул внутри серогрупп (например, серотипы 6-й, 9-й и 11-й серогруппы) [18]. Нуклеотидные замены могут изменить специфичность гликозилтрансфераз (например, серотипы 11A и 11D) и создать условия для изменения структуры полисахаридной капсулы и последующего уклонения от иммунной системы хозяина [18]. Также отмечена вариация *cps*-локусов среди штаммов, экспрессирующих один и тот же серотип. При сравнении *cps*-локусов >100 изолятов серогруппы 6 была выявлена вариабельность трех генов (*wciP*, *wzy*, *wzx*), при этом *cps*-локусы серотипа 6В кластеризовались на две группы, отличающиеся дивергенцией нуклеотидов >5% [152, 153]. Для *cps*-локусов штаммов серотипов 6А также было выявлено несколько подтипов [154]. Основные различия между подтипами серотипов 6-й серогруппы были обнаружены в генах *wzg* и *rmlA-D*, необходимых для синтеза рамнозы. При аналогичном анализе полных последовательностей *cps*-локусов штаммов 19-й серогруппы были идентифицированы три и четыре подтипа для серотипов 19А и 19F, соответственно [154].

Для нетипируемых и атипичных изолятов характерны изменения следующих типов: *cps*-локусы были частично или полностью удалены (между генами *dexB* и *aliA* присутствовала только транспозаза) [155]; *cps*-локусы могли быть заменены кластером инсерционных последовательностей с геном *nspA/pspK* [155, 156]; *cps*-локусы содержали гены, значительно отличающиеся от генов референтного *cps*-локуса (22F), но кодирующие пептиды с той же функцией [155].

### 1.5.2. Переключение серотипов (факты, механизмы)

Изменения в распределении циркулирующих серотипов наблюдались как до, так и после внедрения ПКВ [17, 138, 157-169]. При этом наблюдение за пневмококковой популяцией после введения вакцин показало снижение частоты инвазивных ПИ, вызванных пневмококками вакцинных серотипов, но рост заболеваемости и носительства, вызванных невакцинными штаммами. Это может быть связано как с событиями замещения (вытеснения) серотипа, так и событиями переключения типа капсулы. Известно, что для пневмококков разных MLST-типов характерны преобладающие серотип-специфические ассоциации [21, 43, 170]. Поэтому считается, что любой изолят, характеризующийся новой или необычной комбинацией генотип-серотип, потенциально ассоциируется с событием переключения типа капсулы.

В различных исследованиях были идентифицированы события генетической рекомбинации, ассоциированные с фактами переключения серотипов [52, 145, 171, 172]. Диапазон размеров фрагментов рекомбинации варьировал от ~21,9 до 56,5 т.п.н. [52, 145, 171, 172]. В одном из исследований была изучена генетически разнообразная коллекция исторических и современных пневмококков (426 изолятов), датированных 1937–2007 гг. [173]. Были идентифицированы 36 независимых событий переключения серотипов, 34 из которых предшествовали введению антипневмококковых конъюгированных вакцин: четыре события произошли до 1950 года; 11, 17, 22 и 28 накопленных независимых событий произошли до 1960, 1970, 1980 и 1990 годов, соответственно [173]. Размер рекомбинационных фрагментов варьировал от ~19,0 т.п.н. до  $\geq 58,2$  т.п.н. В некоторых случаях фрагменты рекомбинации включали часть или полные гены пенициллин-связывающих белков *pbp2x* и *pbp1a*, расположенных на расстоянии ~8 т.п.н. выше 5'-области и ~7 т.п.н. ниже 3'-области *cps*-локуса, соответственно [173]. Не было установлено ассоциации паттерна рекомбинационных фрагментов и времени выделения изолята [173]. В двух случаях, каждый из которых датирован не позднее 1960 года, импортированные фрагменты включали *cps*-локус и близлежащие гены пенициллин-связывающих белков, *pbp2x* и *pbp1a* [173]. Таким образом, переключение серотипа капсулы было регулярным явлением в течение последних 7 десятилетий и предшествовало как повсеместному введению вакцины, так и широкому применению антибиотиков.

### 1.6. Факторы вирулентности *S. pneumoniae*

Главным фактором вирулентности пневмококка является полисахаридная капсула, способствующая уклонению от очищения из носоглотки человека физиологическими жидкостями (слизь и слюна) на ранней стадии колонизации [18]. Колонизация *S. pneumoniae* носоглотки человека также опосредована адгезией бактерий к поверхности клеток эпителия [173]. Адгезивные свойства *S. pneumoniae* зависят от адгезивных белков, ферментов и других поверхностных молекул: протеаза IgA1, фосфорилхолин, нейраминидазы, экзогликозидазы, поверхностный адгезин пневмококка, холинсвязывающий белок А (СbpA), липопротеины и стрептококковые филаменты (пили, фимбрии) и т. д. Считается, что переход от бессимптомного носительства к инвазивному заболеванию зависит от баланса между защитными механизмами хозяина и вирулентностью бактерий. Можно выделить следующие этапы

развития инвазивного заболевания: адгезия к клеткам эпителия носоглотки; инвазия в подлежащие ткани и кровь; уклонение от иммунного ответа хозяина; индукция воспаления и прямое повреждение ткани; размножение. Различные факторы вирулентности вносят вклад в каждую из этих стадий.

Факторы вирулентности пневмококка часто проявляют специфичность в отношении рецепторов клеток хозяина. В частности, одним из рецепторов пневмококка, вовлеченных в адгезию, внутриклеточную инвазию и трансцитоз пневмококков через эндотелиальные клетки, является рецептор фактора активации тромбоцитов (рФАТ) [18]. Фактор активации тромбоцитов (ФАТ) – связанный с G-белком глицерофосфолипид, содержащий остатки фосфорилхолина (ФХ). ФАТ экспрессируется преимущественно активированными клетками хозяина и усиливает адгезию *S. pneumoniae* в присутствии воспаления. Варианты колоний пневмококка, имеющих «мутный» или «прозрачный» фенотип поверхности, отличаются по количеству ФХ-содержащих тейхоевых кислот, формирующих детерминанты для взаимодействия с рФАТ, относительно количества полисахарида капсулы. Поскольку рецептор рФАТ связан с G-белком, взаимодействие с ним приводит к ГТФазо-опосредованной активации фосфолипаз C, D и A2, а также активации протеинкиназы C (ПК-C) и тирозинкиназ [18]. При взаимодействии пневмококка и рФАТ в качестве сигнального паттерна преобладает активация сигнальных киназ ERK1/ERK2 [18].

Другим способом манипуляции пневмококками сигнальных каскадов в клетках хозяина является взаимодействие патогена с гликопротеинами внеклеточного матрикса (ЭКМ), которые выступают в качестве молекулярных мостиков для связывания с рецепторами интегрин [18]. Многие экспрессируемые пневмококком поверхностные белки (PavA, PavB, PfbA, CbpG, RrgA) используют адгезивный гликопротеин фибронектин (ФН) в качестве молекулярного мостика, опосредующего взаимодействие с рецепторами эукариотических клеток [18]. Также пневмококки взаимодействуют с белком внеклеточного матрикса тромбоспондином-1 (TSP-1) и связывают витронектин (ВН) через поверхностный белок пневмококка C (PspC) для облегчения прилипания к клеткам-хозяевам [18].

Поверхностные адгезины пневмококков взаимодействуют с факторами комплемента Н (FN) и C1q [18], а также с каскадом фибринолитической системы хозяина. Пневмококковая эндопептидаза О (PerO) и гликолитические ферменты

енолаза, глицеральдегид-3-фосфогидрогеназа (GAPDH) и фосфолицерат киназа (PGK) разрушают пламиноген и его белки-активаторы для ускорения инвазии в ткани, уклонения от действия комплемента и прикрепления к клеточной поверхности [18].

### **1.6.1. Описание некоторых факторов вирулентности**

#### **Полисахаридная капсула пневмококка**

Полисахаридная капсула ингибирует активность комплемента и фагоцитоз [18]. Степень вирулентности конкретного серотипа зависит от химического состава капсулы, и, возможно, отражает различную способность разных серотипов избегать комплемент-зависимого и независимого опсонофагоцитоза [18]. Капсула ингибирует уничтожение *S. pneumoniae* нейтрофильными внеклеточными ловушками – волокнистыми структурами из остова ДНК, гистонов и гранулированных белков нейтрофилов [18].

На сегодняшний день для *S. pneumoniae* идентифицировано более 100 серотипов [174-181] и выделено 48 серогрупп на основе паттернов агглютинации сывороточных антител [176]. Вариации в циркуляции распротраненных серотипов пневмококка наблюдались в разных возрастных группах носителей [182-186] и в разные периоды времени [157-165]. Ретроспективный анализ результатов 169 исследований показал, что до введения ПКВ наиболее распространенными в мире серотипами, ассоциированными с ИПИ у детей в возрасте до 5 лет были серотипы 14, 6В, 1, 23F, 5, 19F и 6А [182]. У взрослых ИПИ ассоциировались с более широким спектром серотипов: серогруппы / серотипы 6, 9, 14, 1, 3, 8, 12F [187]. Внедрение ПКВ повлияло на циркуляцию серотипов и привело к появлению не идентифицированных ранее серотипов [179].

В ряде исследований было показано, что серотипы 4, 1, 14, 19А и 18С характеризовались более высокой инвазивностью, тогда как серотипы 23F, 6А, 35F и 11А были в значительной степени связаны с носительством [21, 46]. Инвазивность серотипа положительно коррелировала с частотой серотип-специфических атак, то есть числом эпизодов инвазивного заболевания, возникающих в результате числа эпизодов носительства данного серотипа [188].

Причины серотип-специфических различий в инвазивности не до конца понятны, но могут быть обусловлены различиями в структуре капсул. Пневмококки, экспрессирующие сложные капсулы, производят их в меньшем объеме и менее приспособлены для роста в условиях ограничения питательных веществ [189-191]. Пневмококки, экспрессирующие более простые капсулы, производят больший объем

капсульных полисахаридов и демонстрируют большую устойчивость к нейтрофил-опосредованному фагоцитозу [192]. Было показано, что скорость отложения комплемента на клетках пневмококка варьирует, в зависимости от серотипа [193].

### **Регуляция биосинтеза полисахаридной капсулы**

Контроль экспрессии капсулы необходим для успешной колонизации и развития заболевания [142]. При первичной колонизации, экспрессия капсулы позволяет избежать удаления микроба вместе со слизью [18]. Для обеспечения адгезии белков капсулы к поверхности эпителия необходимо снижение экспрессии капсулы [194]. Для развития инвазивного процесса и уклонения от фагоцитоза требуется увеличение экспрессии капсулы [18].

Известны два морфологических типа колоний пневмококка – прозрачные и мутные [129]. Переход между мутным и прозрачным фенотипами, а также регуляция уровня экспрессии полисахарида, опосредованы эпигенетическим механизмом, в основе которого лежат события рекомбинации между генами *CRM* типа I [124]. Прозрачный фенотип колоний, характеризующийся меньшим уровнем экспрессии капсулы, доминирует во время асимптоматической колонизации носоглотки. Мутный фенотип колоний отличается более высоким уровнем экспрессии капсулы и ассоциируется с развитием инвазивных заболеваний.

В регуляции экспрессии капсулы задействовано множество механизмов. В частности, изменение свойств пептидогликана зависит от мутаций белка-регулятора *RegM* (гомолог *CsrA*-белка, контролирующего катаболитную репрессию) [195], белков *PgdA* (деацетилаза *N*-ацетил-глюкозамина пептидогликана) и *Adr* (деацетилаза *N*-ацетилмурамовой кислоты пептидогликана) [196]. Существует связь между мутацией в гене *arcD*, компонентом аргинин-дезаминазной системы, и снижением уровня экспрессии полисахарида капсулы [197, 198]. Условия окружающей среды (доступность кислорода) играют роль в посттранскрипционной регуляции экспрессии капсулы [199].

Анализ транскриптома штаммов, выделенных из разных источников показал, что экспрессия *cpsA*, первого гена *cps*-оперона, не меняется в зависимости от занимаемой пневмококком ниши (носogлотка, легкие и кровь) [197]. При этом прозрачные колонии преобладали преимущественно в носоглотке, в то время как мутные колонии часто ассоциировались с ИПИ [197]. Изменение уровня транскрипции оперона капсулы в условиях *in vitro*, за счет конститутивного промотора, приводило к увеличению

количества транскрипта – но это не влияло на размер капсулы [200]. Таким образом, регуляция фазовой вариации (мутные и прозрачные колонии) является сложным процессом и не зависит от уровня экспрессии *cps*-локуса.

Наиболее хорошо изучен механизм регуляции синтеза *wzu*-зависимых капсул, опосредованный системой фосфорилирования *CpsA-D* [142, 201-203]. Три белка *CpsB*, *CpsC* и *CpsD* формируют активный белковый тирозин-киназный комплекс [142]. *CpsB* – тирозин-киназа, относящаяся к семейству фосфодиэстераз [204], дефосфорилирует *CpsD*. Мутации в *CpsB* и *CpsD* часто критичны для утраты вирулентности штаммами пневмококка [203-206]. Ингибирование фосфатазной активности *CpsB* приводит к снижению уровня капсульного полисахарида и улучшенной способности прикрепления к макрофагам [207]. *CpsC* относится к семейству полисахарид-зависимых ко-полимераз [208]. *CpsC* действует как трансмембранный активатор, т.к. активность цитоплазматического *CpsD* зависит от его взаимодействия с *CpsC* [208]. Мутации в области экстрацеллюлярной петли *CpsC* приводят к уменьшению фосфорилирования *CpsD* и снижению экспрессии полисахарида капсулы [209]. Не все детали регуляторной системы *CpsA-D* полностью изучены. В частности, недопонимание эффектов мутаций в *CpsB* привело к тому, что до сих пор неизвестно, какая форма тирозин-киназы оптимальна для биосинтеза полисахарида капсулы – фосфорилированная или нефосфорилированная [203-206].

Капсулы серотипов 3 и 37 не содержат активных фосфотрансфераз. Регуляция биосинтеза полисахарида капсулы серотипа 3 зависит от метаболизма углеводов. Активность фосфоглюкомутазы, кодируемой вне *cps*-локуса и участвующей в образовании UDP-Glc, а также доступность моносахаридов-предшественников, критичны для регуляции синтеза капсулы [210]. Также было показано, что мутации фермента UDP-глюкозо-дегидрогиназы (*Cps3D*) в штаммах серотипа 3 также влияют на размер капсулы [211]. Размер полимера серотипа 3 может быть напрямую связан с концентрацией уридин-дифосфоглюкуроновой кислоты (UDP-GlcUA) [212]. Поскольку UDP-GlcUA присутствует во многих серотипах пневмококка, регуляция синтеза этого метаболита критична для многих штаммов.



### **Протеаза IgA1**

Протеазы IgA1 связаны с клеточной стенкой пневмококка и специфически разрушают пептидные связи пролин-треонин или пролин-серин в шарнирной области IgA1 человека, разъединяя фрагменты Fab от фрагмента Fc [213].

### **Фосфорилхолин (ФХ)**

ФХ связан с тейхоевыми (ТК) и липотейхоевыми (ЛТК) кислотами клеточной стенки пневмококка, закрепляет холинсвязывающие белки (ХСБ) [214]. Гидролаза клеточной стенки фосфорилхолинэстераза CbpE модулирует количество ФХ на клеточной стенке пневмококка, вызывает изменение морфологии колоний *S. pneumoniae* и снижение колонизации пневмококком носоглотки у крыс [215]. ФХ способствует инвазии, облегчая трансмиграцию пневмококка через клеточные барьеры хозяина [216].

### **Гиалуронидаза**

Гиалуронидаза расщепляет гиалуроновую кислоту – компонент соединительной ткани. Штаммы *S. pneumoniae*, выделенные от пациентов с менингитом и менингоэнцефалитом, проявляли повышенную активность гиалуронидазы [217-219].

### **Экзогликозидазы**

Гликозилирование белков поверхности клеток человека включает присоединение стандартной коровой олигосахаридной единицы с высоким содержанием маннозы с последующим присоединением дополнительных остатков сахаров (N-ацетилглюкозамина, GlcNAc; галактозы; сиаловой кислоты) [220]. Пневмококк экспрессирует на поверхности ферменты экзогликозидазы: NanA,  $\beta$ -галактозидазу (BgaA) и  $\beta$ -N-ацетилглюкозаминидазу (StrH) [220-222]. NanA, BgaA и StrH последовательно расщепляют концевую сиаловую кислоту, концевую галактозу и концевой N-ацетилглюкозамин, что приводит к обнажению остатков маннозы, необходимых для адгезии *S. pneumoniae* [220].

### **Поверхностный адгезин пневмококков A (PsaA)**

PsaA, поверхностный липопротеин клеточной стенки пневмококка, является компонентом АТФ-связывающего кассетного транспортера, импортирующего ионы марганца [223]. PsaA специфически связывается с трансмембранным гликопротеином E-кадгерином, в монослоях культуры D562. Возможно, что PsaA, как и многие другие факторы вирулентности, выполняет две функции во время инфекции: прямого адгезина, а также участвует в поглощении марганца [224]. PspA предотвращает активацию как

классического, так и альтернативного пути комплемента, за счет своего взаимодействия с компонентом C3b [18]. PspA также взаимодействует с лактоферрином человека, ингибируя его бактерицидное действие [18].

### **Пневмококковый поверхностный белок С / Холин-связывающий белок А (PspC / CbpA)**

CbpA выполняет множество функций во время адгезии, как при носительстве, так и при развитии инфекции легких [225]. CbpA связывается с человеческими рецепторами полимерных Ig (hpIgR), экспрессируемыми клеточными линиями эпителия носоглотки человека [226]. Наличие избыточного свободного секреторного компонента IgG и секреторного IgA (SIgA) в полости слизистой оболочки может привести к связыванию CbpA с SIgA или IgG, способствуя тем самым адгезии *S. pneumoniae* к эпителию носоглотки [33, 194, 226-228]. Экспрессия CbpA выше у пневмококков, прикрепленных к клеткам эпителия носоглотки (клеточная линия D562), по сравнению с пневмококками, выделенными из крови и ликвора мышей [229]. Взаимодействие CbpA с FH приводит к деградации C3b, что мешает опсонизации *S. pneumoniae* компонентами альтернативного пути комплемента [18].

### **Фактор адгезии и вирулентности А пневмококков (PavA)**

Поверхностный белок PavA не содержит типичного мотива LPXTG для закрепления белков, связанных с клеточной стенкой грамположительных бактерий, также как и холин-связывающих доменов. PavA связывается с фибронектином. Делеционные мутанты *S. pneumoniae* по *pavA* демонстрируют пониженную адгезию к эпителиальным клеткам линий A549 и Hep-2 в условиях *in vitro* [230].

### **Стрептококковая ротамаза липопротеинов (SlrA)**

SlrA – липопротеин, принадлежащий к семейству пептидил-пролил-изомераз (PPIases). Функция SlrA неизвестна, но было показано, что SlrA важен для колонизации носоглотки *S. pneumoniae* у мышей, поскольку мутантные штаммы по SlrA демонстрируют повышенный клиренс из носоглотки, по сравнению с *S. pneumoniae* дикого типа [231].

### **Фимбрии (ворсинки, пили)**

Серотипы *S. pneumoniae*, ассоциирующиеся с высоким колонизирующим и инвазивным потенциалом (например, серотипы 4 и 19F), имеют на поверхности пили – длинные отростки, выходящие за пределы полисахаридной капсулы [232, 233]. Пили

кодируются островком патогенности *rlr* [234]. Три из семи генов (*rrgA*, *rrgB*, *rrgC*) кодируют белки с мотивами LPXTG, характерными для поверхностных белков. Другие три гена кодируют гомологи сортазы *srtB*, *srtC* и *srtD*, катализирующие ковалентное присоединение RrgA, RrgB, RrgC, имеющих мотивы LPXTG, к клеточной стенке бактерий [235, 236]. RrgA отвечает за повышенную адгезию в условиях *in vitro*, к клеткам респираторного эпителия человека [237].

### **Белки, связывающие плазминоген**

Поверхностные белки  $\alpha$ -енолаза (Epo) и глицеральдегид-3-фосфатдегидрогеназа (GAPDH) являются протеолитическими активаторами плазминогена [238]. Epo проявляет сродство как к плазминогену, так и к плазмину, а GAPDH – плазмину [239, 240]. Плазмин – фермент, участвующий в разрушении внеклеточного матрикса при ангиогенезе. Мутантные по *epo* штаммы пневмококка неспособны растворять фибрин и характеризуются пониженной вирулентностью в моделях инфекции легких у мышей [241].

### **Пневмолизин (Ply)**

Пневмолизин принадлежит семейству тиол-активированных цитолизин, которые представляют собой связывающие холестерин токсины, способные проникать через клеточную мембрану, вызывать агрегацию и лизис клетки [242]. Воздействие пневмолизина приводит к увеличению проницаемости сосудов легких, уменьшению популяции альвеолярных макрофагов и привлечению нейтрофилов и моноцитов в альвеолярное пространство [18]. Пневмолизин активирует фосфолипазу А в эндотелии легочной артерии, расщепляющую фосфолипиды клеточной мембраны с высвобождением свободных жирных кислот и цитотоксичных лизофосфатидов. Пневмолизин стимулирует высвобождение макрофагами оксида азота и воспалительных цитокинов, фактор некроза опухоли- $\alpha$  (ФНО $\alpha$ ) и интерлейкин-6 (ИЛ-6) [18]. Пневмолизин, независимо от Толл-подобных рецепторов, активирует сигнальный путь, опосредуемый ядерным фактором активированных Т-клеток (NFAT), который, в свою очередь, регулирует экспрессию медиаторов воспаления [18]. Пневмолизин активирует C1q комплемента, связываясь с Fc-рецептором IgG [18]. Пневмолизин способен ингибировать антимикробную функцию полиморфноядерных нейтрофилов *in vitro* в чрезвычайно низкой дозе, усиливая вирулентность *S. pneumoniae* [18].

## **1.7. Распространение резистентности к бета-лактамам у *S. pneumoniae***

Первые резистентные к  $\beta$ -лактамам антибиотикам штаммы пневмококка появились в 40-х годах [18]. В конце 1960-х гг. о резистентных к пенициллину пневмококках впервые сообщили в США и Австралии [243, 244], в конце 1970-х гг. - в Южной Африке [245]. В период с 1980 по 1990-е гг. распространенность резистентных к пенициллину штаммов в мире резко возросла, превысив 50% популяции пневмококка в ряде регионов [168, 246-252]. В этот период антибиотики широко использовались во многих странах. После введения ПКВ7 в 2000 году, в США произошло существенное снижение распространенности резистентных к пенициллину штаммов [253]. Однако впоследствии произошло увеличение частоты распространенности пневмококков невакцинных серотипов [254], что сопровождалось вторичным повышением распространенности резистентности к бета-лактамам антибиотикам [255, 256]. На фоне использования вакцины ПКВ7, значительную часть резистентных к пенициллину штаммов пневмококка стали составлять изоляты, принадлежащие серогруппе 19А, большинство которых ассоциировалось с клональной группой СС320 [17, 257], изначально распространенной в странах Азии [258].

## **1.8. Механизмы резистентности к бета-лактамам антибиотикам у *S. pneumoniae***

К  $\beta$ -лактамам антибиотикам относятся пенициллины, пенициллиноподобные антибиотики (метициллин, амоксициллин и ампициллин) и цефалоспорины (цефотаксим, цефуроксим и цефтриаксон). Пенициллины и пенициллиноподобные антибиотики являются бактерицидными препаратами, вызывают лизис бактерий [259]. Цефалоспорины являются бактериостатическими, т.к. влияют на рост чувствительных клеток, не приводя к лизису.

Первичными мишенями для  $\beta$ -лактамовых антибиотиков является семейство мембраносвязанных транспептидаз (пенициллин-связывающих белков, ПСБ). ПСБ катализируют синтез и ремоделирование пептидогликана клеточной стенки. PBP1a, PBP1b и PBP2a являются бифункциональными – полимеризуют гликановую цепь и проявляют активность транспептидазы, структурируя нити пептидогликана посредством пептидных мостиков [18]. PBP2b и PBP2x ассоциируются только с транспептидазной активностью [260]. ПСБ имеют три консервативных домена активного сайта, Ser-X-X-Lys, Ser-X-Asn и Lys-Thr (Ser)-Gly. Большинство мутаций, имеющих значение для

резистентности к  $\beta$ -лактамным антибиотикам, локализованы внутри транспептидазных доменов, связывающих пенициллин, рядом с консервативными мотивами активного сайта [261-270]. Мутации снижают сродство ПСБ к  $\beta$ -лактамным антибиотикам.

Анализ генов *pbp* показал, что нуклеотидные последовательности чувствительных к  $\beta$ -пенициллину штаммов высоко консервативны [143, 244], но последовательности генов *pbp* устойчивых штаммов содержат фрагменты ДНК, отличающиеся от ДНК восприимчивых штаммов до 25% [266]. При этом существует корреляция между минимальной подавляющей концентрацией  $\beta$ -лактамного антибиотика (МПК) штамма пневмококка и уровнем гетерогенности генов *pbp* [294]. Широко распространена точка зрения, что фрагменты *pbp*-генов резистентных к бета-лактамным антибиотикам штаммов получены от других видов стрептококков [271]. В частности, было показано, что устойчивые к пенициллину изоляты пневмококка в Восточной Европе обладают мозаичными блоками *pbp2b*, сходными с таковыми у штамма *S. oralis M3* [271]. Фрагменты генов *pbp* многих резистентных к бета-лактамам штаммов также показывают высокое сходство с аналогичными последовательностями *S. mitis* [264, 272]. Было показано, что детерминанты устойчивости к пенициллину могут передаваться между *S. mitis* и *S. pneumoniae*, а также между *S. pneumoniae* и *S. oralis* или *S. sanguis* в процессе гомологичной рекомбинации [271].

Высокий уровень устойчивости к  $\beta$ -лактамным антибиотикам требует приобретения как минимум двух измененных ПСБ: RBP2x, RBP1a или RBP2b. RBP2x обладает наиболее высоким сродством к  $\beta$ -лактамным антибиотикам [262, 273, 274]. В отличие от RBP2x, изменения RBP2b ассоциируются только с устойчивостью к пенициллину и не влияют на резистентность к цефалоспорином [275], поскольку цефалоспорины не связывают RBP2b [274]. Экспрессия низкоаффинных RBP1a связана с высоким уровнем устойчивости как к пенициллину, так и к цефалоспорином [276].

Формирование резистентности к пенициллину также ассоциируется с адаптивными мутациями в других генах, кодирующих множество других белков, в частности: гены оперона *turMN / fibAB* [277, 278], гены оперона *ciaRH* [279], locus *clpL* [280].

### **1.9. Резистентность к антибиотикам других классов в популяциях *S. pneumoniae***

В соответствии с данными Европейской сети эпидемиологического надзора за устойчивостью к противомикробным препаратам (European Antimicrobial Resistance

Surveillance Network, EARS-Net), резистентность пневмококков к макролидам в 2010 году в Европе варьировала от 3,8% в Норвегии до 30% во Франции [281]. По данным исследования SENTRY, в 2009 году распространенность резистентных к макролидам пневмококков в США составила 39,2% [256], в странах Африки к югу от Сахары – 7% [282]. Согласно данным Азиатской сети по эпидемиологическому надзору за резистентными к антибиотикам патогенами (Asian Network for Surveillance of Resistant Pathogens, ANSORP), распространение резистентности к макролидам в Азии было более высоким: 96,4% изолятов из Китая и > 80% изолятов из Южной Кореи, Тайваня и Вьетнама [283]. Кроме того, за 10-летний период, начинающийся с 1992 года, возросла распространенность мультирезистентных к бета-лактамам и макролидам штаммов: до 32,7%, 17,0% и 15,3% – во Франции, Испании и США, соответственно [249].

Распространенность устойчивых к тетрациклину пневмококков составляла 3,9-6,9% в Великобритании [284, 285], 1-84,8% в Азии [286], 24,5% в США [256], 13,8% в Германии, 35,1% в Испании и 42,3% во Франции [285]. Показатели резистентности к триметоприму-котримоксазолу также возросли, начиная с 1990-х гг.: 72,2% в Таиланде и 86,1% в Китае [283], 34,3% в США [256], 34,1% в Латинской Америке [287], 28,2% в Великобритании, 48,5% в Испании, 49,7% на Ближнем Востоке и 56,2% во Франции [288].

Показатели резистентности пневмококков к фторхинолонам, используемым как препараты резерва, были ниже в большинстве регионов: 5,0% в Африке к югу от Сахары [306], 3,8% в Великобритании, 8,2% в США, 20,3% во Франции, 26,1% в Испании, 27,1% в Сингапуре, 66,8% в Гонконге [288]. Сообщалось о показателях резистентности пневмококков к левофлоксацину: 0,2% для 10 стран Латинской Америки [287] и 0,8% для США [256]. Низкие показатели распространенности (<1%) также были зарегистрированы в Азии, за исключением 3,3% и 9,1% частоты резистентных к левофлоксацину штаммов в Таиланде и Тайване.

#### **1.10. Механизмы и детерминанты резистентности к антибиотикам других классов**

Макролиды нарушают синтез белка, связывая 50S-субъединицу рибосомы [289]. Ген *ermB* кодирует N-метилтрансферазу, которая метилирует остаток аденина в V-домене 23S рРНК, предотвращая таким образом связывание пневмококков с макролидами [290]. Участок связывания макролидов с рибосомами один и тот же для линкозамидов и стрептограмина В, и приобретение *ermB* также придает устойчивость к

этим противомикробным препаратам (фенотип MLSB). Гены *mefA/E* кодируют эффлюксные насосы, специфичные для 14- и 15-членных макролидов (включая эритромицин, кларитромицин и азитромицин) [289, 291]. Активный эффлюкс позволяет резистентной клетке снизить внутриклеточную концентрацию макролидов и устранить вредные эффекты этих препаратов. Наличие *mefA/E* не придает устойчивости к линкозамидам или стрептограмину В (М-фенотип).

Хлорамфеникол нарушает синтез белка, связывая 50S-субъединицу рибосомы. Резистентность к хлорамфениколу обеспечивается за счет приобретения гена *cat* [292], кодирующего хлорамфениколацетилтрансферазу, которая катализирует превращение хлорамфеникола в производные, не способные связываться с рибосомами [289].

Тетрациклины связываются с 30S-субъединицей рибосом и ингибируют синтез белка, блокируя взаимодействие аминоацил-тРНК с акцепторным сайтом рибосомы. Ген *tetM* кодирует белок рибосомальной защиты, который, как было показано, инициирует обратимое связывание тетрациклина с рибосомой GTP-зависимым образом [293].

Все вышеперечисленные детерминанты устойчивости ассоциированы с МГЭ, такими как генная кассета MEGA, Tn1207.1 и мобильными элементами семейства Tn916 (см. Раздел 1.5.1.2.1) [78-81]. Кроме того, сами Tn916-подобные элементы могут быть интегрированы в независимые ИКЭ-элементы. Обычно такие составные элементы образуются путем ассоциации Tn916-подобного элемента с Tn5252-подобным элементом [50, 59, 78, 294-296]. Среди пневмококков было идентифицировано несколько представителей семейства Tn916 и составных элементов (Таблица 1).

Для представителей пневмококков, ассоциированных с рядом клональных комплексов (CC15, CC18, CC66, CC156/CC162, CC81, CC90, CC242 и CC236/CC271), было характерно множественное, независимое приобретение мобильных элементов, в то время как среди других CC не было доказательств приобретения каких-либо мобильных элементов [295]. ИКЭ не были идентифицированы среди изолятов, принадлежащих CC191 и CC218, CC113 и CC124 [295].

Два наиболее ранних известных представителя ИКЭ семейства Tn916, ассоциированные с изолятами, датированными 1967 и 1968 гг., были структурно очень похожи и имели высокий уровень идентичности последовательностей с Tn916-подобной областью ИКЭ ICESp23FST81 референсного штамма PMEN1, датированного 1984 г. Идентификация таких элементов среди пневмококков, выделенных в конце 1960-х

годов, подтверждает их существование с первого десятилетия, в течение которого сообщалось о резистентности пневмококку к тетрациклинам [297]. Помимо генов *tetM*, содержащихся в двух Tn916-подобных элементах, никакие другие детерминанты устойчивости к *tetM*, *ermB*, *mefA/E* или *cat* не были идентифицированы среди пневмококков, выделенных до 1980 года. Ранее не было никаких признаков каких-либо других ИКЭ-элементов в геномах этих изолятов. После 1980 года, у резистентных пневмококков идентифицированы другие Tn916-подобные элементы – с включением кассеты MEGA [50].

**Таблица 1 – Представители семейства Tn916 и составных элементов у *S. pneumoniae*** (из Wyres K. 2012. [253])

ИКЭ	Особенности	Детерминанты резистентности	Ссылка
Tn916	24 ORF, включая <i>tetM</i>	<i>tetM</i>	[298]
Tn1545	Содержит <i>IS1239</i> между <i>orf13</i> и <i>orf12</i> , <i>MAS</i> -кассету между <i>orf20</i> и <i>orf19</i> (с генами <i>ermB</i> , <i>sat4</i> и <i>aphA-3</i> ), обеспечивает резистентность к макролидам, стрептомицину, канамицину	<i>tetM</i> , <i>ermB</i> , <i>sat4</i> , <i>aph-3</i>	[299]
Tn2008	Интегрирован в <i>cat</i> -содержащий Tn5252-подобный элемент. Содержит <i>MAS</i> -кассету вместо гена <i>tetM</i> ,	<i>cat</i> , <i>ermB</i> , <i>sat4</i> , <i>aph-3</i>	[295]
Tn2009	Содержит <i>MEGA</i> -кассету, инсерцированную в <i>orf6</i>	<i>tetM</i> , <i>mefE</i>	[300]
Tn2010	Содержит: ген резистентности к эритромицину между <i>orf20</i> и <i>orf19</i> ; <i>MEGA</i> -кассету, инсерцированную в <i>orf6</i>	<i>tetM</i> , <i>ermB</i> , <i>mefE</i>	[80]
Tn3872	Содержит Tn917-элемент (с <i>ermB</i> ) между <i>tetM</i> и <i>orf6</i>	<i>tetM</i> , <i>ermB</i>	[299]
Tn5251	Отличается от Tn916 нуклеотидными заменами в гене <i>tetM</i>	<i>tetM</i>	[301]
Tn5253	Tn5251, интегрированный в Tn5252	<i>cat</i> , <i>tetM</i>	[296]
Tn6002	Содержит Tn917-элемент между <i>orf20</i> и <i>orf19</i>	<i>tetM</i> , <i>ermB</i>	[299]
Tn6003	Содержит <i>MAS</i> -кассету, инсерцированную между <i>orf20</i> и <i>orf19</i>	<i>tetM</i> , <i>ermB</i>	[299]
Tn6058	Tn6002, интегрированный в Tn5252	<i>cat</i> , <i>tetM</i> , <i>ermB</i>	[296]
ICESp23FST81	Интегрирован в Tn5252-подобный элемент, содержащий кластер генов синтеза / экспорта лантибиотиков	<i>cat</i> , <i>tetM</i>	[59]
ICESpn11876	Интегрирован в Tn5252-подобный элемент (обратно ориентирован), утративший <i>cat</i>	<i>tetM</i> , <i>ermB</i>	[50]
ICESpn11930	Интегрирован в Tn5252-подобный элемент (обратно ориентирован), утративший <i>cat</i> . Содержит ген резистентности к эритромицину и Tn917	<i>cat</i> , <i>tetM</i> , <i>ermB</i>	[50]

### 1.11. Заключение по обзору литературы

Таким образом, направление эволюции пневмококка детерминируется метаболическими особенностями данного микроорганизма, его зависимостью от питательных веществ хозяина, а также селективным давлением вакцин и антибиотиков.



Диверсификация популяции *S. pneumoniae* обусловлена многочисленными генетическими процессами: естественной трансформацией и рекомбинацией, конъюгацией и трансдукцией. Приобретаемые в ходе данных процессов ДНК-фрагменты других генетических линий и других стрептококков, а также ИКЭ и профаги часто содержат геномные детерминанты, дающие пневмококку эволюционные преимущества для выживания: факторы вирулентности и резистентности.

Системы рестрикции-модификации удаляют чужеродные локусы, а также определяют множество эпигенетических состояний клеток бактерий, которые потенциально могут приводить к глобальным изменениям фенотипа (изменение толщины капсулы на разных этапах колонизации и инвазии, изменение экспрессии поверхностных адгезинов и др.).

События генетической рекомбинации могут приводить к переключению серотипов на фоне применения ПКВ, но подобные процессы происходили и до начала массового использования вакцин. Связь серотипа с резистентностью к бета-лактамам антибиотикам может быть, скорее, следствием особенностей соответствующего типа капсулы, определяющих продолжительность носительства штамма в носоглотке, а также вероятность достижения плотности бактерий, способствующей запуску процесса естественной трансформации.

Различные генетические линии пневмококка демонстрируют различную степень стабильности и диверсификации. Результаты предыдущих исследований использованы в качестве основы для изучения популяционной структуры пневмококка в России.

Установление генетических взаимосвязей между бактериальными штаммами и эпидемическими клонами имеет фундаментальное значение для эпидемиологических и эволюционных исследований. Это возможно благодаря определению клональной структуры популяции с помощью MLST-типирования, а также с помощью методов анализа ядерного генома, имеющих большее разрешение, по сравнению с классическим MLST. Информация о сиквенс-типах и клональных комплексах может быть использована для выяснения распределения популяции по разным географическим регионам, что будет способствовать разработке стратегии надзора для предотвращения распространения эпидемически значимых генетических линий. Изучение структуры пангенома позволяет оценивать баланс между отбором и дрейфом, который играет важную роль в формировании бактериального генома [302].

## ГЛАВА 2. МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

### 2.1. Общий план работы

Общий план работы представлен на Рисунке 2.

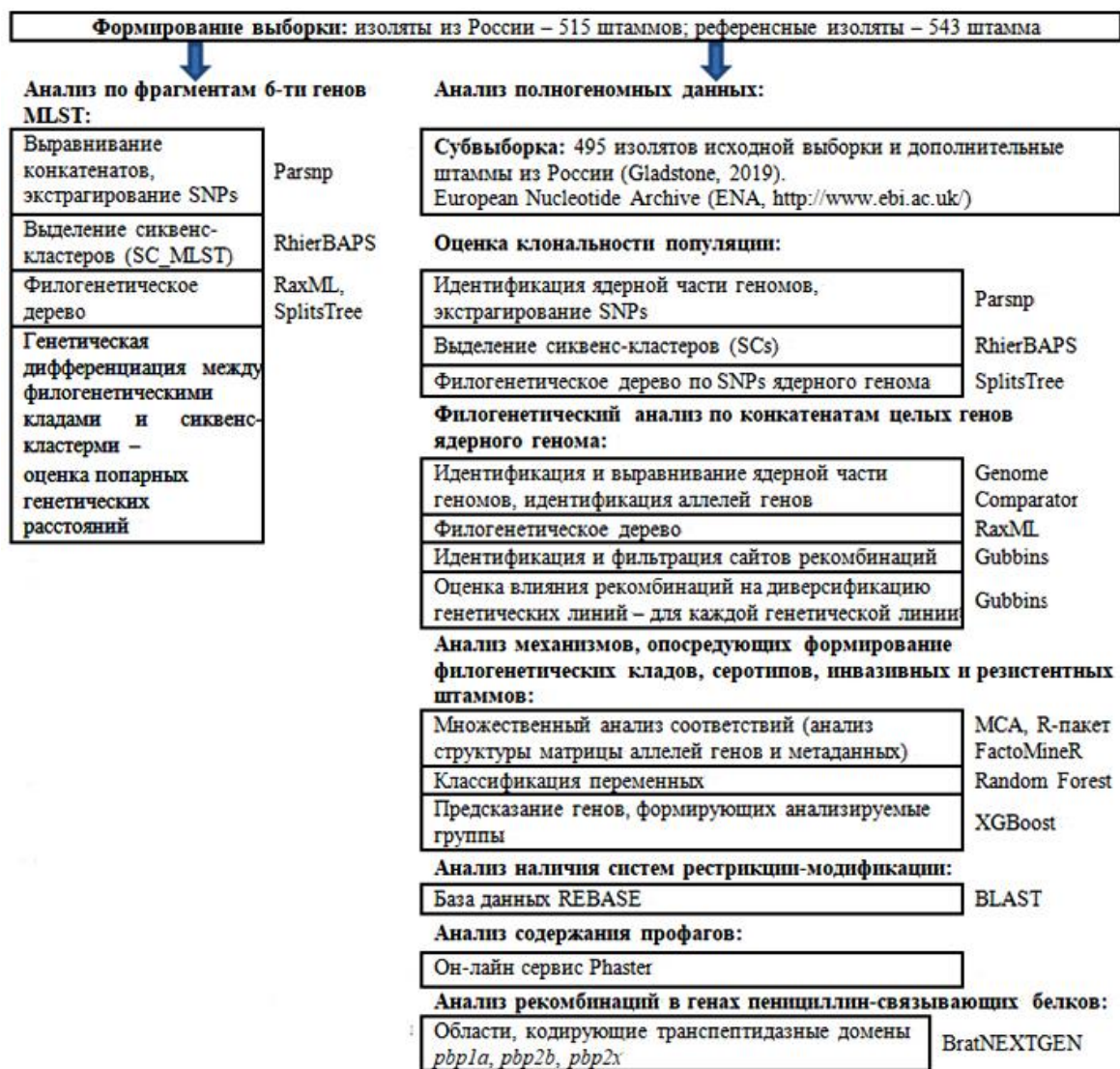


Рисунок 2 – Общая схема исследования.

### 2.2. Формирование выборки

В исследование были включены все штаммы пневмококка из России, доступные в базе данных PubMLST: 515 штаммов, выделенных в период с 1980 по 2017 гг. в различных городах (126 из 515 штаммов были типированы в ФГБУ ДНКЦИБ ФМБА России в период с 2011 по 2017 гг.). Выборка была дополнена референтными штаммами *S. pneumoniae* (n=543), принадлежащими распространенным эпидемически значимым

клонам, либо аналогичным российским редким сиквенс-типам. При отборе штаммов соответствующего сиквенс-типа в выборку включались по два изолята из каждого региона мира с интервалом 1-2 года. Для каждого сиквенс-типа было отобрано эквивалентное число инвазивных изолятов и изолятов от носителей, а также эквивалентное число резистентных и чувствительных к антибиотикам штаммов, если это было возможно. Изоляты распространенных сиквенс-типов имели большую вероятность включения в выборку. Таким образом, данную выборку можно считать стратифицированной по генотипу, распределение штаммов по генотипам соответствует распространенности соответствующих генетических линий в мире и полученные результаты можно распространять на всю популяцию пневмококка. При формировании выборки не учитывались редкие генотипы пневмококков, циркулирующих в других регионах. Однако можно допустить их родственность распространенным в разных географических областях эпидемически значимым клонам. Метаданные, доступные для 932 изолятов (год выделения, источник выделения, географический регион, серотип, сиквенс-тип, данные антибиотикочувствительности), вместе с данными MLST-типирования были загружены из базы PubMLST (за исключением 126 изолятов, идентифицированных и типированных в ДНКЦИБ). Итого, первичная выборка для анализа взаимосвязей между циркулирующими в России изолятами и распространенными в других регионах мира эпидемически-значимыми клонами, включала 1058 изолятов.

При формировании первичной выборки отбирались изоляты, для которых доступны данные полногеномного секвенирования, когда это было возможно. Чтобы увеличить репрезентативность данных, относящихся к циркулирующим в России штаммам, в исследование были включены 80 геномов изолятов *S. pneumoniae* из России, опубликованные в работе Gladstone и соавторов в 2019 году [303]. Итого, вторичная выборка включала 495 изолятов пневмококка: 132 изолята из России и 363 референтных штамма, принадлежащих распространенным клонам.

### **2.3. Характеристика изолятов *S. pneumoniae* из музейной коллекции ФГБУ ДНКЦИБ ФМБА России (Санкт-Петербург)**

В работе были использованы изоляты *S. pneumoniae* из музейной коллекции, полученные от пациентов с инвазивными и неинвазивными пневмококковыми заболеваниями, а также от носителей, из различных регионов России за 2010-2017 годы.

Данные чувствительности к пенициллину (PEN) и эритромицину (ERY) были получены ранее с использованием диско-диффузионного метода и рекомендаций EUCAST. ПЦР-серотипирование проводилось с учетом рекомендаций CDC, MLST-типирование изолятов выполнено по стандартной схеме, [www.mlst.net](http://www.mlst.net). Для MLST-типирования было отобрано 126 изолятов, включающих как PEN-резистентные, так и PEN-чувствительные пневмококки, принадлежащие либо к распространенным в России, либо редким серотипам. Принадлежность сиквенс-типов к клональным комплексам (для однолокусных вариантов) была установлена с помощью eBURST анализа (программа Phylouviz, версия 2.0a). Полногеномное секвенирование 45 изолятов выполнено на платформе MiSeq (Illumina), данные загружены в Генбанк (BioProject PRJNA422133 и BioProject PRJNA738184).

#### **2.4. Филогенетический анализ по конкатенатам последовательностей генов MLST**

Филогенетический анализ был выполнен для 1058 изолятов с метаданными, представленных 343 сиквенс-типами (Том 2 – Приложения. Приложение 1. Таблица 1).

##### **2.4.1. Выравнивание последовательностей**

Конкатенаты последовательностей фрагментов шести из семи генов домашнего хозяйства (за исключением гена *ddl*) были выровнены с помощью программы Parsnp, версия 1.2 [304]. Идентичные последовательности тех же сиквенс-типов, но соответствующие разным штаммам, были добавлены в выравнивание для анализа метаданных.

##### **2.4.2. Анализ структуры популяции**

416 сайтов однонуклеотидного полиморфизма (SNPs) были экстрагированы с помощью пакета Parsnp для построения филогенетического дерева в программе RAxML, версия 8.2.12 [305], с использованием стандартной модели эволюции GTR (General Time-Reversible), с выполнением 100 бутстрэп реплик [305]. Результирующее дерево было обработано по алгоритму укоренения в среднюю точку (midpoint rooting) относительно самой длинной ветви. Аннотация филогенетического дерева с использованием различных метаданных была выполнена в программе iTOL, версия 4.3.2 [306]. То же самое выравнивание SNPs было проанализировано в программном пакете RhierBAPS (R версия 3.5.2, RhierBAPS пакет версии 1.1.0), алгоритм которого основан на реализации байесовского подхода для интерпретации вероятности, для анализа взаимосвязей между генетическими кластерами и их субструктурами [307]. Программа

SplitsTree (версия 4.14.8) была использована для получения филогении по методу ближайшего соседа (neighbor joining, NJ) с выполнением 100 бутстрэп реплик, с применением алгоритма split decomposition (SD) [308]. Последовательности родственного вида *Streptococcus mitis* были использованы в качестве аутгруппы.

### **2.4.3. Попарные оценки генетических расстояний между идентифицированными в популяции группами**

Генетическая дифференциация между филогенетическими кладами была проведена с помощью анализа генетических расстояний с использованием R-пакета и скрипта для попарной оценки расстояний по SNP [309].

## **2.5. Анализ полногеномных данных**

### **2.5.1. Общедоступные базы данных**

Риды 495 изолятов были загружены из базы данных European Nucleotide Archive (ENA, <http://www.ebi.ac.uk/>) (Том 2 – Приложения. Приложение 1. Таблица 1).

### **2.5.2. Сборка геномов *de novo* и аннотация**

Качество сырых ридов было оценено с помощью программы FastQC, версия 0.11.8 [310]; фильтрация ридов по качеству и удаление адаптеров и праймеров ПЦР, используемых при подготовке библиотек, выполнены с помощью программы Trimmomatic, версия 0.33 [311]. Короткие риды (менее 36 bp) также были отфильтрованы. Сборка геномов *de novo* была выполнена с помощью программы SPAdes, версия 3.13.0 [312, 313]. Финальная оценка качества была выполнена с помощью программы Quast, версия 5.0.1 [314]. Образцы, контаминированные другими видами микроорганизмов, были идентифицированы с помощью алгоритма MLST, версии 2.11, и исключены из анализа [315, 316]. Геномы были аннотированы с помощью программного интерфейса myRAST и RAST сервера (Rapid Annotations using Subsystems Technology) [317].

## **2.6. Анализ ядерной части генома**

Для филогенетического анализа популяции на основании полногеномных данных были использованы два альтернативных метода идентификации ядерной части генома.

### **2.6.1. Идентификация ядерной части генома и исключение сайтов рекомбинаций с помощью программного пакета Harvest tools Parsnp**

Множественное выравнивание и идентификация ядерной части геномов изучаемых изолятов было выполнено с помощью Parsnp, версия 1.2 [304, 318], с

применением опций -с и -х (включение всех образцов и исключение рекомбинаций с помощью пакета PhiPack) [319]. Работа алгоритма Parsnp построена на идентификации локальных коллинеарных блоков (local collinear blocks, LCB), и в сравнение включаются только те LCB, которые присутствуют во всех геномах. Выравнивание 49974 переменных SNPs, относящихся к коровой части генома, было экстрагировано для построения филогенетического дерева в программе RaхML, версия 8.2.12. Результирующее дерево было обработано по алгоритму укоренения в среднюю точку (midpoint rooting) относительно самой длинной ветви. То же самое выравнивание SNPs было проанализировано в программном пакете RhierBAPS (R версия 3.5.2, RhierBAPS пакет версии 1.1.0) для анализа взаимосвязей между генетическими кластерами и их субструктурами.

### **2.6.2. Идентификация ядерной части геномов с помощью модуля GenomeComparator в программе BIGSdb (Bacterial Isolate Genome Sequence Database) и исключение рекомбинаций с помощью программы Gubbins**

Идентификация и выравнивание ядерной части генома были выполнены с помощью алгоритма GenomeComparator в программе BIGSdb со следующими параметрами: минимальный процент идентичности – 70%, к ядерной части генома отнесены локусы, присутствующие у 95% изолятов [315]. Выравнивание конкатенатов генов ядерного генома было использовано для построения филогенетического дерева в программе RaхML, с использованием стандартной модели эволюции GTR (1000 бутстреп-реплик). Выравнивание генов ядерного генома и полученное филогенетическое дерево были использованы для анализа рекомбинаций с помощью программы Gubbins, версия 2.3.4 [320]. Алгоритм «sliding window», используемый в программе Gubbins, позволяет идентифицировать области с атипично высокой плотностью SNPs. Результирующее филогенетическое дерево было построено в программе Gubbins после исключения сайтов рекомбинаций. Визуализация сайтов рекомбинаций и аннотация филогенетического дерева были выполнены с помощью программы Phandango [321]. Аннотация также была выполнена в программе iTOL.

## **2.7. Поиск вариабельных генов, вариации в которых участвуют в формировании кладов A/B1/B2, сиквенс-кластеров и серотипов, резистентности к бета-лактамам и макролидам, или ассоциируются с инвазивностью**

### **2.7.1. Идентификация аллелей генов**

Для поиска вариабельных генов, участвующих в формировании изучаемых групп, была использована таблица с вариантами генов корового генома штаммов *S.pneumoniae*, полученная с помощью GenomeComparator и дополненная метаданными. Анализируемые геномы были аннотированы в BIGSdb с помощью автоматизированного алгоритма BLAST (минимальный процент идентичности – 70%, минимальный процент выравнивания локуса – 50%). Сравнение геномов в GenomeComparator проводилось против референсного генома ATCC 700669 (PMEN1). Аллелям, идентичным референсному геному, был присвоен идентификационный номер «1». Аллелям, отличающимся на один или более нуклеотидов, последовательно присваивался следующий идентификационный номер. Результирующая таблица была использована в качестве входных данных для классификации.

### **2.7.2. Множественный анализ соответствий (Multiple correspondens analisys, MCA)**

Предварительный анализ данных для оценки структуры первичной матрицы был выполнен с помощью множественного анализа соответствий (Multiple correspondens analisys, MCA, R-пакет FactoMineR). Вариативность переменных была оценена для 100 измерений (накопительная вариативность данных – 70,93%). Переменные с высоким значением качества ( $\cos^2 > 0,5$ ), одновременно объясняющие вариативность данных в 20 измерениях, были экстрагированы для дальнейшего анализа в качестве вторичной матрицы. Также были проанализированы переменные с высоким качеством  $\cos^2$ , одновременно объясняющие вариативность данных в 27–30 и в 50 измерениях.

### **2.7.3. Классификация с помощью алгоритма машинного обучения, основанного на методе «Случайного леса» (Random Forest)**

Для выполнения алгоритма Random Forest использовался R-пакет bigrf, версия 0.1-4, рассчитанный на анализ большого числа категориальных переменных [322]. В качестве входных данных для классификации использовались первичная матрица, полученная с помощью модуля GenomeComparator и дополненная метаданными, а также вторичная матрица с ограничением входных данных на основании результатов множественного анализа соответствий. Ошибка Out-of-bag error rates и Confusion matrix

оценены по результатам построения 1000 деревьев. Значимые для классификации переменные были идентифицированы по значению z-score. Визуализация результатов была проведена в Rstudio и с помощью Python.

#### **2.7.4. Предсказание генов, формирующих анализируемые группы с помощью алгоритма машинного обучения, основанного на моделях градиентного бустинга для линейных основанных на деревьях моделей XGBoost**

В качестве входных данных для классификации с помощью алгоритма XGBoost (EXtreme Gradient Boosting) [323] на основе приложения Hermes2 (H2O) (версия 3.10.4.2) [324], использовалась первичная матрица, полученная с помощью модуля GenomeComparator и дополненная метаданными. XGBoost – библиотека машинного обучения, реализующая модель градиентного бустинга деревьев решений. Параметры запуска XGBoost были автоматически определены с помощью модуля AutoML того же приложения. Визуализация результатов была проведена в Rstudio.

#### **2.8. Анализ наличия систем рестрикции-модификации**

Для поиска последовательностей, сходных с последовательностями эндонуклеаз рестрикции, метилтрансфераз и субъединиц специфичности, была использована программа blastall с порогом E-value <0.01. Полученные совпадения были отфильтрованы по значениям bit-score и идентичности. Поиск белков систем рестрикции-модификации проводился против базы REBASE [325].

#### **2.9. Анализ содержания профагов в геномах**

Поиск последовательностей профагов в геномах изучаемых изолятов был выполнен с помощью он-лайн сервиса Phaster [326].

#### **2.10. Анализ рекомбинаций в *cps*-локусе изолятов, принадлежащих генетическим линиям CC90, CC236 / CC271 / CC320**

Для изолятов соответствующей генетической линии было выполнено отдельное выравнивание геномов с помощью алгоритма GenomeComparator в программе BIGSdb, анализ рекомбинаций был выполнен в программе Gubbins [171]. Полученные результаты были визуализированы с помощью он-лайн сервиса Phandango. Был проведен поиск рекомбинаций в области *cps*-локуса, ассоциированных с локализованными рядом генами пенициллин-связывающих белков.



### **2.11. Анализ рекомбинаций в генах пенициллин-связывающих белков анализируемых изолятов и других видов стрептококков**

В анализ были включены все изучаемые изоляты пневмококка, а также представители других видов стрептококков (Раздел «3.4. Сравнительный анализ генов пенициллин-связывающих белков изолятов *S. pneumoniae*, относящихся к эпидемическим генетическим линиям»). Фрагменты генов пенициллин-связывающих белков *pbp2x*, *pbp2b* и *pbp1a*, кодирующие области транспептидазного домена, были экстрагированы и выровнены с помощью Geneious (версия 11.0.4) [327]. Выравнивание последовательностей было использовано для построения филогенетического дерева в программе RAxML с использованием стандартной модели эволюции GTR, а также для анализа рекомбинаций в программе BratNextGen. Полученные результаты были визуализированы с помощью он-лайн сервиса Phandango.

### **2.12. Поиск значимых однонуклеотидных полиморфизмов, ассоциированных с резистентностью к бета-лактамам антибиотикам, у изолятов CC320**

Субвыборка для поиска значимых полногеномных однонуклеотидных полиморфизмов, ассоциированных с резистентностью к бета-лактамам антибиотикам, включала 46 изолятов пневмококка CC320, представленных однолокусными и двулокусными вариантами референсного штамма *S. pneumoniae* TW31 (Приложение 6, Таблица 1). В качестве референсного генома использовался геном *S. pneumoniae* TW31. Риды двух изолятов *S. pneumoniae* ST236, чувствительных к пенициллину, а также риды 44 изолятов *S. pneumoniae* CC320, характеризующихся наличием различной степени резистентности к пенициллину по данным МПК, были картированы на референсный геном с помощью пакетов bowtie и samtools. Аннотация значимых однонуклеотидных полиморфизмов была выполнена с помощью программного пакета SnpEff. Были отобраны значимые SNPs, присутствующие не менее чем в 60% ридов в соответствующей позиции. Из полученных значимых SNPs были экстрагированы только те однонуклеотидные полиморфизмы, по которым все резистентные к пенициллину штаммы одинаково отличались от двух чувствительных. Кодирующие последовательности, содержащие отобранные SNPs, были экстрагированы из собранных *de novo* геномов и транслированы в аминокислотные последовательности для дальнейшего анализа полиморфизма индивидуальных белков.

### **2.13. Оценка полиморфизма индивидуальных белков у резистентных к пенициллину изолятов, принадлежащих клональному комплексу CC320**

Для каждого изолята исследуемой популяции кодирующие последовательности были транслированы в протеиновые последовательности и экстрагированы из сборки *de novo*. Сравнение аминокислотных последовательностей, ассоциированных с высокой степенью полиморфизма у резистентных к пенициллину штаммов *S. pneumoniae*, было выполнено путем множественного выравнивания набора последовательностей с помощью алгоритма MUSCLE. В качестве референсных последовательностей использовали протеиновые последовательности, принадлежащие двум диким штаммам из исследуемой популяции *S. aarpneumoniae* (ERR057872 и ERR047995).

### **2.14. Оценка топологии филогенетических деревьев**

Сравнительная оценка топологии и кластеризации филогенетических деревьев, построенных по последовательностям ключевых идентифицированных генов с помощью программного пакета TreeCmp (версия v.1.1-b308) [328-330]. Соответствие кластеризации деревьев оценивалось с помощью параметров «Matching split distance» (соответствие дистанций расщепления) и «MatchingCluster» (соответствие кластеров).

### **2.15. Аннотация метаболических путей**

Реконструкция взаимодействующих метаболических путей была выполнена с помощью базы данных STRING [331].

### **2.16. Анализ наличия детерминант резистентности к макролидным антибиотикам и тетрациклинам**

Был осуществлен поиск известных интегративных конъюгативных элементов [77]. Для поиска последовательностей, сходных с последовательностями интегративных конъюгативных элементов, была использована программа blastall с порогом E-value <0.01. Полученные совпадения были отфильтрованы по значениям bit-score и идентичности. Наличие *mefE/mel*, *ermB* и *tetM* генов, а также ряда сложносоставных мобильных элементов и интегративных конъюгативных элементов было подтверждено с помощью blast-поиска в геномах в программе GenomeComparator.

## ГЛАВА 3. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

### 3.1. Оценка структуры популяции *S. pneumoniae*, циркулирующих в России

#### 3.1.1. Филогенетический анализ на основании сравнения локусов схемы MLST

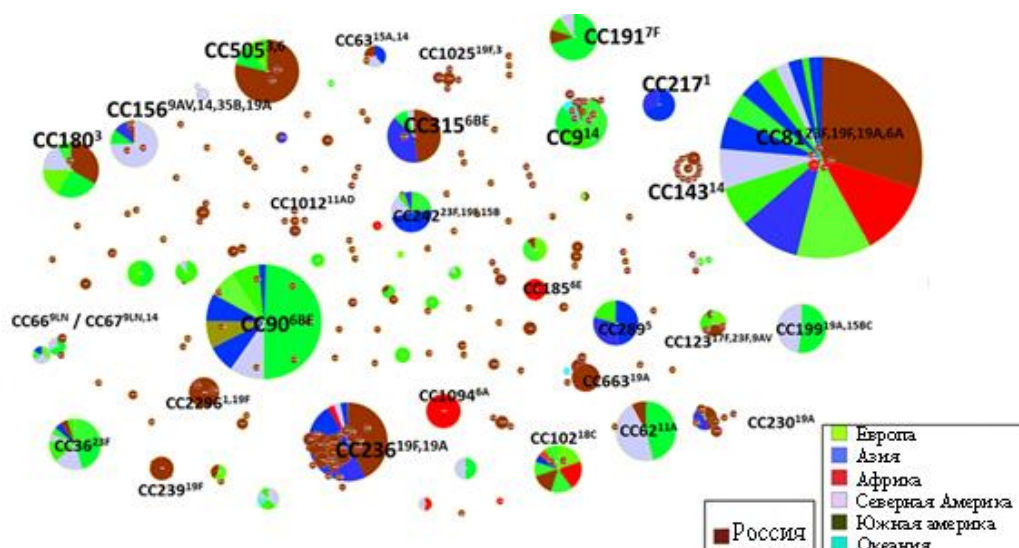
Первичная выборка включала все эпидемически значимые клоны пневмококка, циркулировавшие в различных регионах мира в разное время. Распределение штаммов по сиквенс-типам в выборке отражало тенденцию распространения соответствующего генотипа в мире (Раздел 2.2 «Формирование выборки»), что позволило распространить полученные результаты на мировую популяцию пневмококка.

Для большинства изолятов из России (515 штаммов) доступные данные о генотипе включали только данные MLST-типирования. Анализ с помощью eBURST алгоритма позволил сгруппировать близкие штаммы в клональные комплексы (CC), представленные однолокусными вариантами (SLV) соответствующих распространенных сиквенс-типов.

Первичный eBURST-анализ выборки показал, что в России, в период с 1980 по 2017 гг., циркулировали (в порядке убывания распространенности): CC81<sup>серотипы23F,19F</sup>, CC505<sup>серотип3</sup>, CC236<sup>19F</sup>, CC2296<sup>1,19F</sup>, CC315<sup>6B</sup>, CC663<sup>19A</sup>, CC239<sup>19F,9A/V</sup>, CC180<sup>серотип 3</sup>, CC230<sup>19A</sup>, CC1025<sup>3,19F</sup>, CC143<sup>19A, 14</sup> (Рисунок 3). Однако большая часть пневмококковой популяции в России была представлена редкими сиквенс-типами, эволюционные связи которых с генетическими линиями не могли быть установлены с помощью eBURST-алгоритма.

Для выяснения филогенетических связей и выделения субпопуляций в популяции пневмококка, а также для анализа их ассоциаций с фенотипическими и генетическими особенностями, были применены методы кластеризации, использующие данные множественного выравнивания последовательностей (Раздел 2.4.2 «Анализ структуры популяции»):

- метод минимальной эволюции и метод ближайшего связывания (SplitsTree),
- метод, основанный на расчете генетической модели популяции с помощью метода максимального правдоподобия (RAxML),
- метод байесовской вероятности (hierBAPS).



**Рисунок 3 – eBURST анализ анализируемой первичной выборки *S. pneumoniae*.**

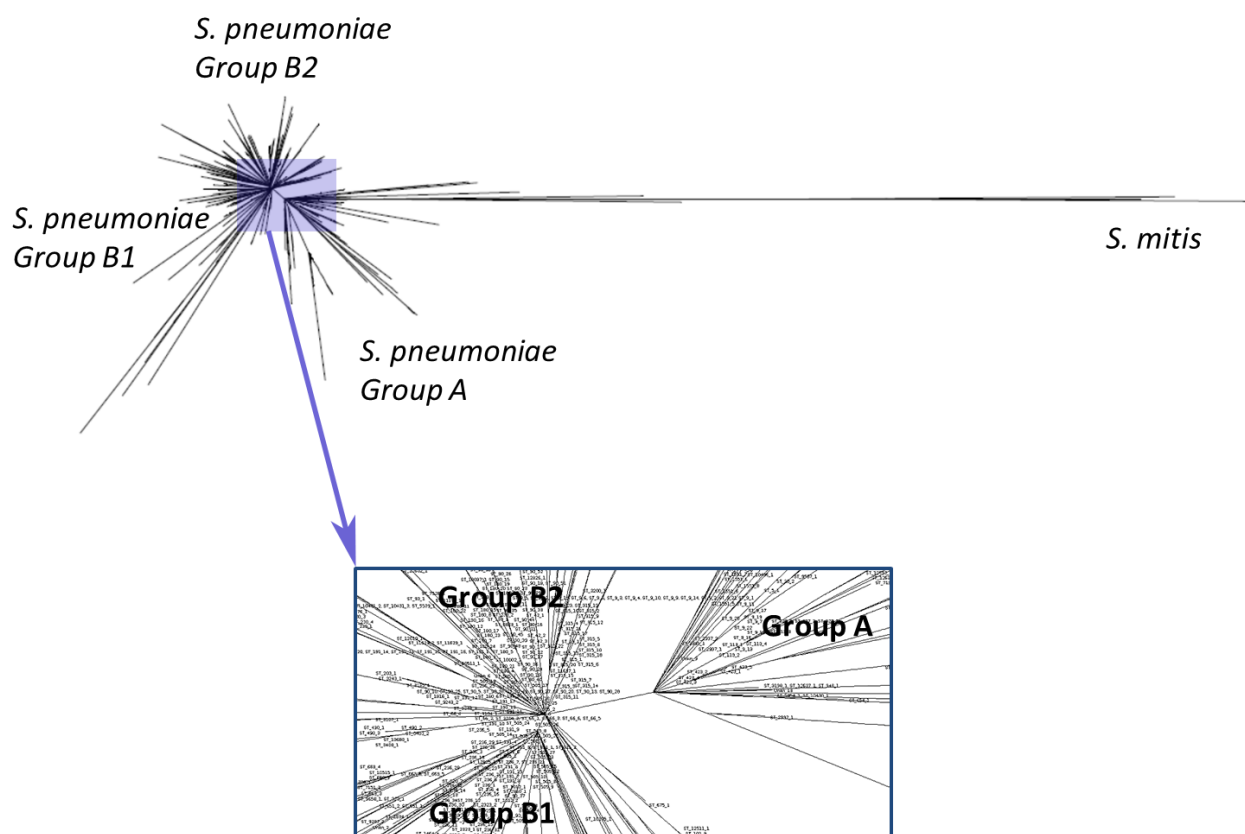
Клональные комплексы представлены однолокусными вариантами. Штаммы *S. pneumoniae* из России отмечены коричневым цветом.

Для методов, основанных на подсчете генетических расстояний (SplitsTree, RAxML), изоляты считаются принадлежащими к одному «кластеру» (филогенетической кладе), если число однонуклеотидных полиморфизмов (SNPs) менее или равно установленного порогового значения.

В программе hierBAPS используется альтернативный подход, при котором во множественном выравнивании последовательностей идентифицируются «сиквенс-кластеры», которые максимизируют вероятность апостериорной Байесовской модели. Данная апостериорная модель основана на мультиномиальном распределении вероятностей Дирихле, учитывающем наблюдаемые данные и сведения об априорной вероятности (предполагаемое число  $K$ -субпопуляций или «сиквенс-кластеров»). Алгоритм hierBAPS применяется для идентификации начального числа «сиквенс-кластеров» и затем итеративно повторяется для последующей кластеризации внутри идентифицированных «сиквенс-кластеров». Для алгоритмов, основанных на байесовской статистике, полученные «сиквенс-кластеры» могут не всегда отражать филогенетические данные. Основная задача подобных алгоритмов классификации – идентифицировать отдельные панмиктичные группы в пределах гетерогенной популяции. Поэтому в начальные «сиквенс-кластеры» могут быть включены изоляты из разных филогенетических кладов, в которых аллели генов распределены случайным образом, все генетические линии одинаково распределены, отсутствует обмен генами с другими «сиквенс-кластерами», не происходит никаких эволюционных изменений.

Здесь и далее под понятием «генетическая линия» подразумевается распространенная монофилетическая группа близкородственных изолятов, принадлежащих к одному или нескольким близкородственным клональным комплексам, входящих в состав одной и той же кледы филогенетического дерева.

Кластеризация изолятов по конкатенатам последовательностей шести генов схемы MLST (кроме *ddl*) позволила разделить популяцию пневмококков на три крупных филогенетических кледы A/B1/B2 на основании ветвей дерева (по программе SplitsTree, Рисунок 4), а также на 11 «сиквенс-кластеров» начального уровня – SC\_MLST (hierVAPS). (Рисунок 5, 6).



**Рисунок 4. – Филогенетическое дерево *S. pneumoniae* по конкатенатам MLST. Методы ближайших соседей и максимальной парсимонии (SplitsTree). *S. mitis* KCOM13 используется в качестве аутгруппы.**

Монофилетические кледы филогенетического дерева соответствовали сиквенс-кластерам начального уровня SC2\_MLST (генетическая линия CC62<sup>11A</sup>), SC3\_MLST (CC236/CC271/CC320<sup>19F,19A</sup>), SC4\_MLST (CC81<sup>23F,19F</sup>), SC7\_MLST (CC90<sup>7F</sup>), SC12\_MLST (CC199<sup>15BC,19A</sup>) (Таблица 2, Рисунок 6). Сиквенс-кластеры, представленные монофилетическими кледями, соответствуют генетическим линиям, дивергировавшим в

процессе эволюции. Остальные сиквенс-кластеры объединяли различные клады филогенетического дерева (Таблица 2, Рисунок 6). Сиквенс-кластеры, включающие неблизкородственные генетические линии, могут отражать последствия конвергентной эволюции (события горизонтального переноса генов, рекомбинации и др.).

Как видно из Таблицы 2, региональные российские генетические линии относились к одному сиквенс-кластеру SC9\_MLST и включали: CC36, CC179, CC239, CC2296, CC801. В то же время в России циркулировали генетические линии, распространенные повсеместно (CC81, CC156, CC289), в странах Азии (CC236/271/320) и Европе (CC63, CC180, CC230, CC315, CC505, CC801). При этом в России не были распространены генетические линии, циркулировавшие преимущественно в странах Северной Америки, Южной Америки и Африки: CC66 и CC191; CC90; CC1094; CC217; CC344 и CC448; CC306; CC199 (Таблица 2).

Таблица 2 – Генетические линии пневмококка, распространенные в России и мире

SC_MLST	Клональные комплексы	PMEN	Серотип	Распространенность (регион)	Период (гг.)
SC1_MLST	<i>S. mitis</i> – 2 штамма*				
SC2_MLST	CC236 / CC271 / CC320	PMEN 14 - Taiwan <sup>19F</sup> -14	19F, 19A	Россия, Азия (Китай, Малайзия, Вьетнам, Тайланд), Сев. Америка	1997 – 2016
SC3_MLST	CC62		11A	Россия, Северная Америка, Европа (Исландия)	2001 – 2016
SC4_MLST	CC81	PMEN 1 - Spain <sup>23F</sup> -1	23F, 19F	Россия, повсеместно	1984 – 2013
SC5_MLST	CC180	PMEN 31 - Netherlands <sup>3</sup> -31	3	Европа (включая Россию), Северная Америка	1988 – 2016
	CC230	PMEN 32 - Denmark <sup>14</sup> -32	19F	Европа (включая Россию), Азия (Тайланд)	2002 – 2012
	CC289	PMEN 19 Colombia <sup>5</sup> -19	5	Россия, повсеместно	2004 - 2010
	CC315	PMEN 20 - Poland <sup>6B</sup> -20	6E	Европа (включая Россию), Азия (Тайланд)	2001 - 2013
	CC505		3	Европа (включая Россию)	1989 – 2017
	CC66		9N	Европа (Исландия), Северная Америка	2001 – 2013
	CC191	PMEN 39 Netherlands <sup>7F</sup> -39	7F	Европа (Исландия, Нидерланды Россия), Северная Америка	1984 – 2014
	CC193	PMEN 30 - Greece <sup>21</sup> -30	21	Европа (Греция, Великобритания)	
CC376	PMEN 23 - N.Carolina <sup>6A</sup> -23	6A	Северная Америка	2001 – 2004	
SC6_MLST	CC9	PMEN 9 - England <sup>14</sup> -9	14	Европа (Россия, Исландия, Великобритания)	2009 - 2013
	CC123		17F	Россия, Европа (повсеместно)	1939 – 2017
	CC113	PMEN 36 - Netherlands <sup>18C</sup> -36	18BC	Европа, Северная Америка	2004 - 2009

Таблица 2 – продолжение таблицы

SC_MLST	Клональные комплексы	PMEN	Серотип	Распространенность (регион)	Период (гг.)
SC6_MLST	CC124	PMEN 35 - Netherlands <sup>14</sup> -35	14	Европа (Нидерланды), Северная Америка	2001 – 2009
	CC185	PMEN 8 - S.Africa <sup>6B</sup> -8	6E	Южная Африка	1984 - 1990
	CC327	PMEN 41 - Portugal <sup>6A</sup> -41	6A	Европа (Нидерланды, Великобритания, Португалия)	
SC7_MLST	CC90	PMEN 2 - Spain <sup>6B</sup> -2	7F	Европа, Азия (Китай, Южная Корея), Северная Америка	1988 – 2013
SC8_MLST	<b>CC156</b>	PMEN 3 - Spain <sup>9V</sup> -3	9A, 35B	<b>Россия, повсеместно</b>	2001 – 2016
	CC1094		6A	Южная Африка	1978 - 1988
SC9_MLST	<b>CC36</b>		23F	<b>Россия</b>	2001 – 2017
	<b>CC179</b>		19F	<b>Россия</b>	2012-2016
	<b>CC239</b>		9V	<b>Россия</b>	1989 - 2011
	<b>CC2296</b>		1	<b>Россия</b>	1981 - 2011
	<b>CC801</b>		4	Европа ( <b>Россия</b> , Чешская Республика)	2008 – 2017
	<b>CC63</b>	PMEN 25 - Sweden <sup>15A</sup> -25	15A	Европа ( <b>Россия</b> , Швеция), Африка, Северная Америка	1988 - 2012
	CC205	PMEN 38 - Sweden <sup>4</sup> -38	4	Европа (Швеция, Нидерланды, Великобритания)	
	CC217	PMEN 27 - Sweden <sup>1</sup> -27	1	Африка, Азия (Тайланд)	2004 – 2010
SC10_MLST	CC344	PMEN 42 - Norway <sup>NT</sup> -42	Бескапсульный	Европа (Швейцария, Норвегия, Польша, Португалия)	1997 – 1999
	CC448	PMEN 43 - USA <sup>NT</sup> -43	Бескапсульный	Северная Америка, Европа (Швейцария, Португалия), Австралия	1995 - 2002
SC11_MLST	<b>CC3544</b>		7F	<b>Россия</b>	2017
	CC218	PMEN 24 - Denmark <sup>12F</sup> -24	12F	Европа (Дания, Нидерланды)	
	CC304	PMEN 40 - Sweden <sup>1</sup> -40	1	Европа (Швеция, Нидерланды, Исландия)	2009
	CC306	PMEN 28 - Sweden <sup>1</sup> -28	1	Европа (Швеция, Нидерланды, Исландия, Великобритания)	2010 – 2012
SC12_MLST	CC199	PMEN 37 - Netherlands <sup>15B</sup> -37	19A, 15BC	Северная Америка, Европа (Нидерланды, Исландия)	2001 – 2014

Примечания: \* 3 штамма *S. mitis*, включенные в качестве аутгруппы, в результате кластеризации hierBAPS отнесены к SC1\_MLST (2 штамма), а штамм *S. mitis* B6 – к SC7\_MLST (вместе с CC90).

Жирным шрифтом и серым цветом фона выделены генетические линии, распространенные в России.



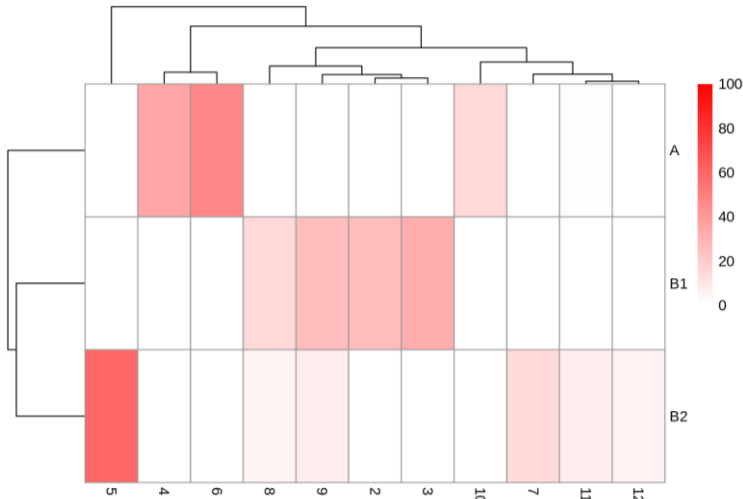


Рисунок 5 – Гистограмма ассоциации между SC-MLST и группами A/B1/B2

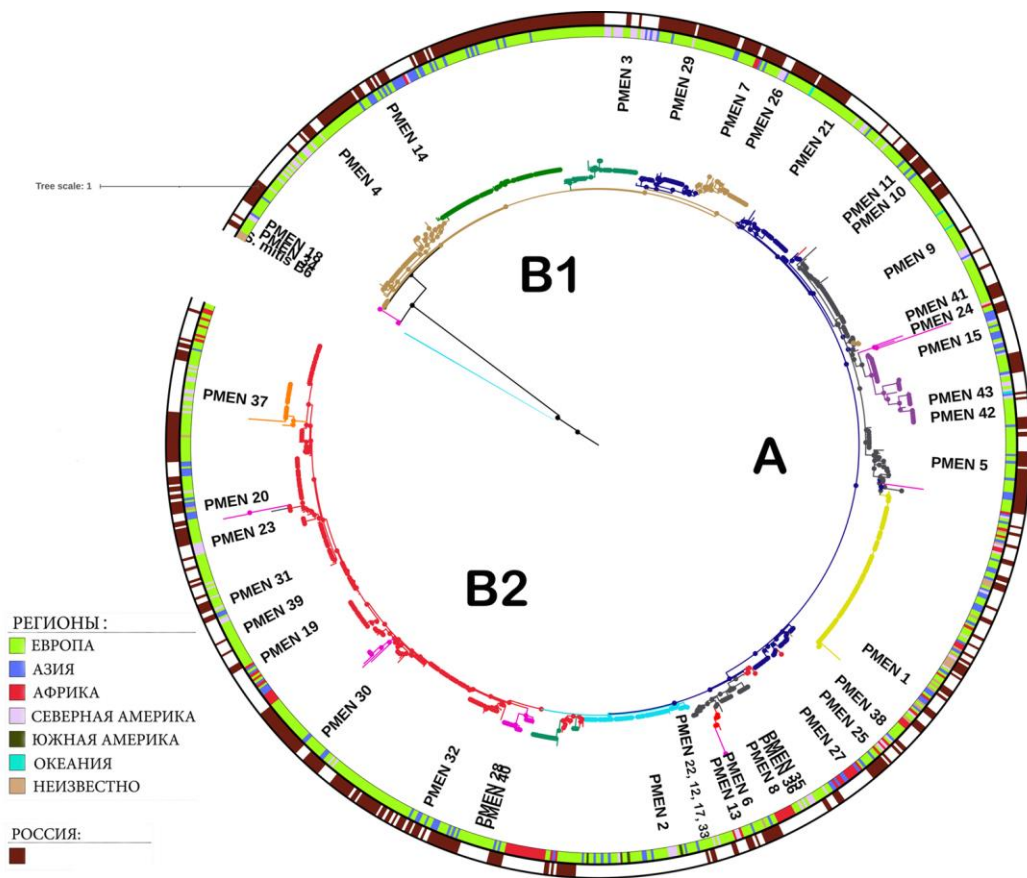


Рисунок 6 – Филогенетическое дерево *S. pneumoniae* по конкатенатам генов MLST.

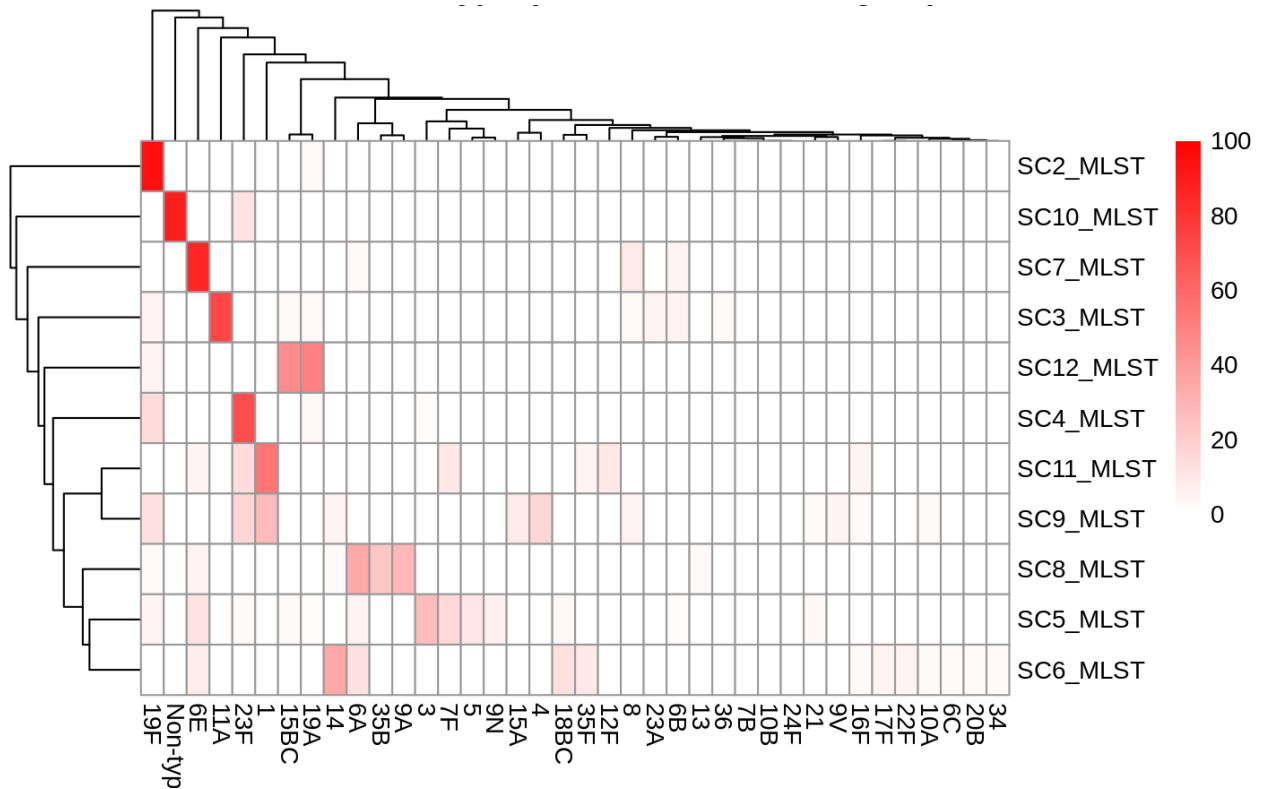
Пронализирована первичная выборка (1058 изолятов), включающая циркулирующие в России изоляты и распространенные в других регионах мира эпидемически-значимые клоны (см. раздел 2.2. «Формирование выборки»). Метод максимального правдоподобия (RaxML). *S. mitis* KCOM13 используется в качестве аутгруппы. География регионов мира отражена цветами во внутреннем кольце области аннотаций. Российские изоляты показаны коричневым цветом во внешнем кольце аннотаций. А, B1 B2 – расположение глобальных групп, выделенных с помощью SplitsTree (Рисунок 4). Кодировка цвета ветвей филогенетического дерева: цвета соответствуют SC\_MLST.



Сиквенс-кластеры SC\_MLST и группы A/B1/B2 ассоциировались с доминирующими серотипами (Рисунок 7 и 8), что можно объяснить способностью серотипа циркулировать в составе с одним или разными сиквенс-типами. Группа А – серотипы 23F, 19F, 14, 23А; группа В1 – серотипы 11А, 19F, 19А, 1, 9N; группа В2 – серотипы 6А/В/Е, 3, 19А, 7F, 5. Интересно, что группа В2 ассоциировалась преимущественно с серотипами, не входящими в состав вакцины ПКВ7.

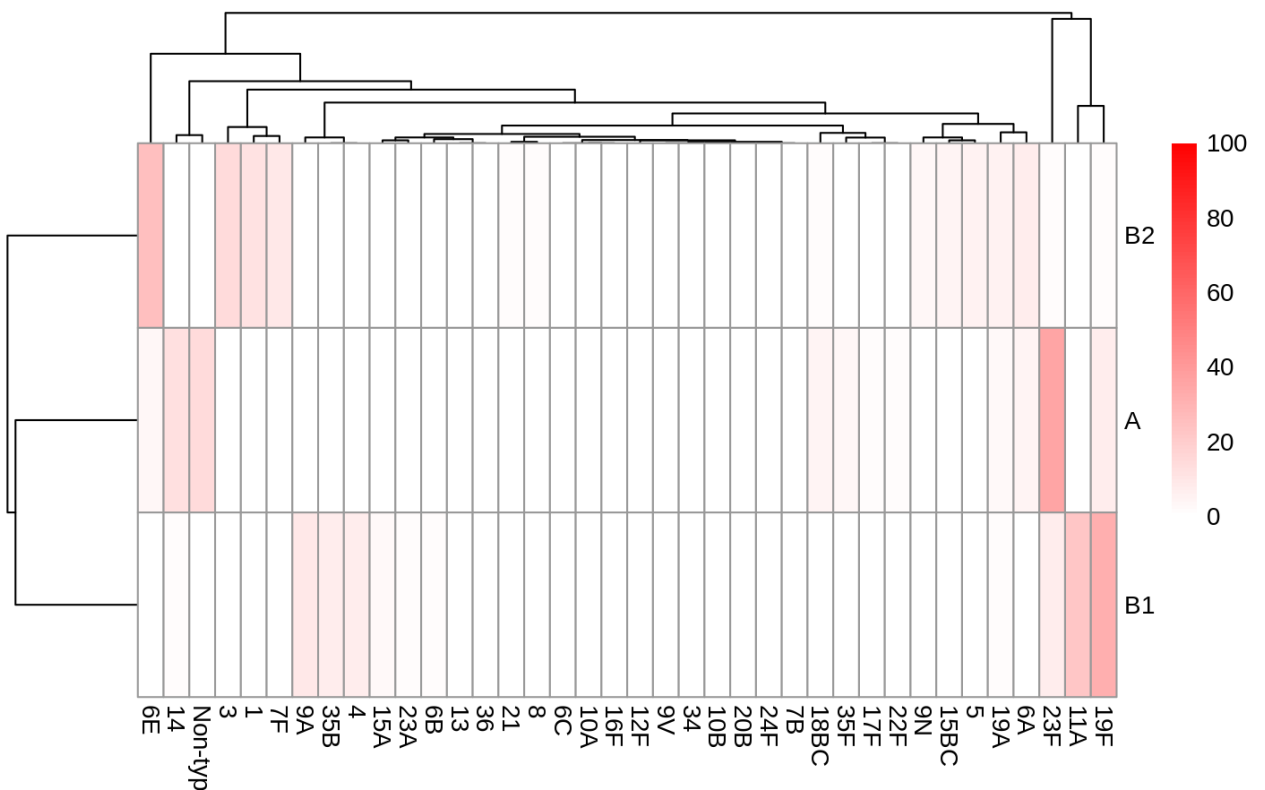
Наблюдалась тенденция ассоциации разных серотипов (и соответственно, разных SC\_MLST и групп А/В1/В2) с источником выделения изолята (Рисунок 9). Наиболее четкой была корреляция инвазивных изолятов (источник выделения – кровь, ликвор) с серотипами: 14, 23F, 3, 6Е, 7F, 18BC, 1, 9N, 10А, 4. Изоляты, выделенные от носителей, были более гетерогенны по серотиповому составу. Редко ассоциировались с инвазивными заболеваниями серотипы 19А, 6А, 11А, 15BC. Серотипы 19А, 19F, 14, 6В, 11А, 15BC часто ассоциировались со средним отитом (Рисунок 9).

К группам А и В1 относились изоляты, выделенные до 1960 г., а также большая часть изолятов (преимущественно серотипы 19F и 23F), выделенных в период 1960-2000 гг. Изоляты, выделенные после 2000 г., распределялись преимущественно по группам В1 и В2. При этом группа В2 ассоциировалась преимущественно с серотипами (6А, 6Е, 3), не входящими в состав вакцины ПКВ7. В группе А также присутствуют изоляты периода после 2000 г., принадлежащие редким сиквенс-типам, часто ассоциирующиеся с невакцированными серотипами (10А, 7С, 6D, 34, 19А, 15В/С, 23А, 23В и др.). Эта картина согласуется с данными о наиболее распространенных серотипах, циркулировавших в мире до начала массовой вакцинации в 2001 г. [248].



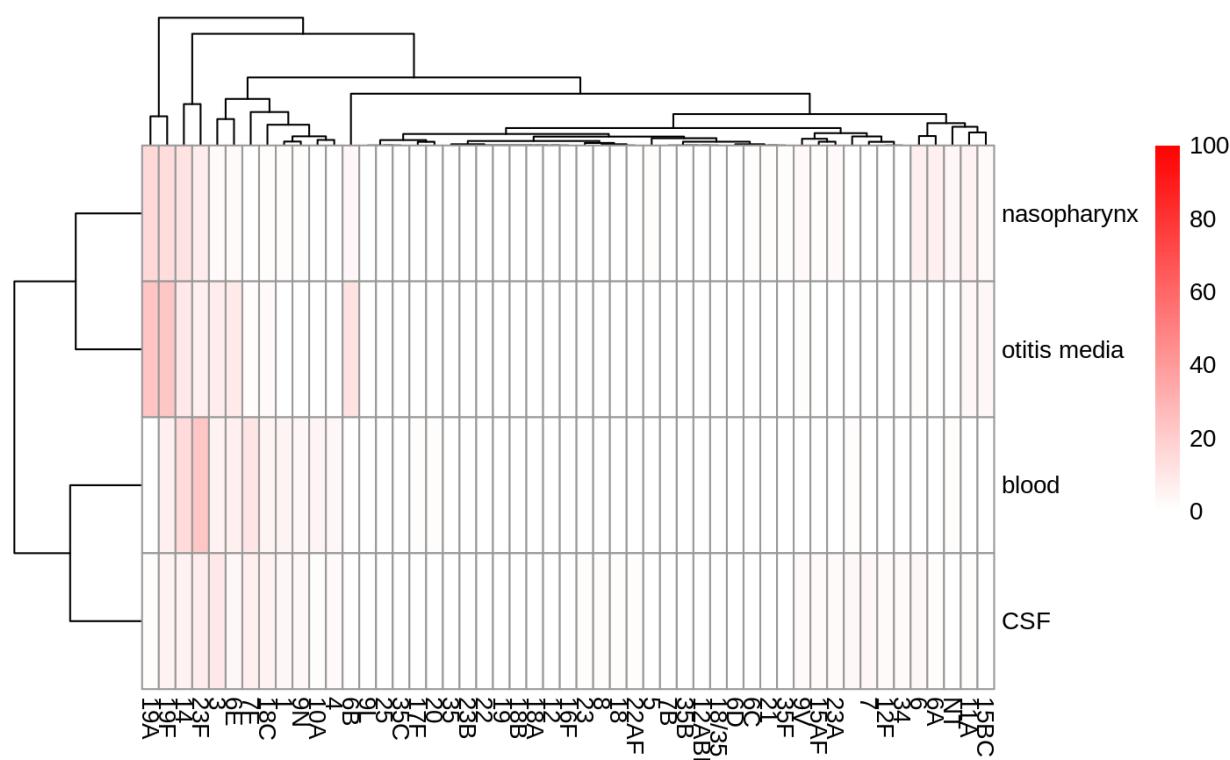
**Рисунок 7 – Гистограмма ассоциации между «сиквенс-кластерами» MLST (SC-MLST) и серотипами.**

Интенсивность цвета коррелирует с % изолятов данного серотипа в данном SC\_MLST



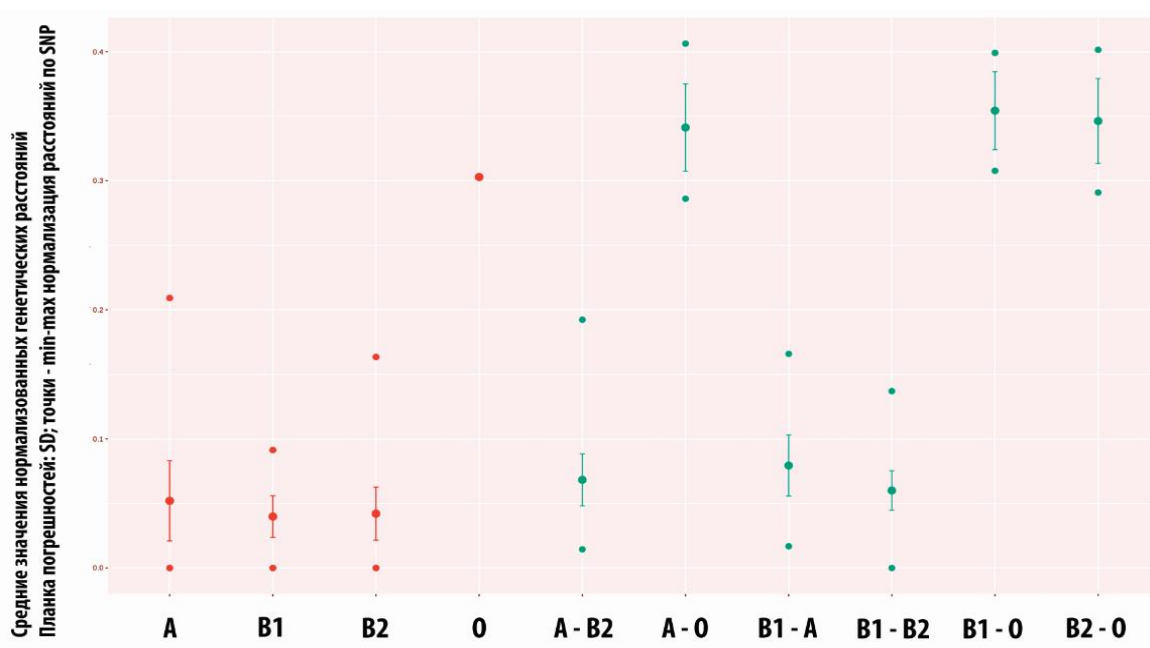
**Рисунок 8 – Гистограмма ассоциации между группами A, B1 и B2 и серотипами**

Интенсивность цвета коррелирует с % изолятов данного серотипа в группах A, B1 и B2



**Рисунок 9 – Гистограмма ассоциации между серотипом и источником выделения изолята**  
 Интенсивность цвета коррелирует с % изолятов данного серотипа в группе «nasopharynx / носоглотка», «otitis media / жидкость среднего уха при среднем отите», «blood / кровь» или «CSF / ликвор»

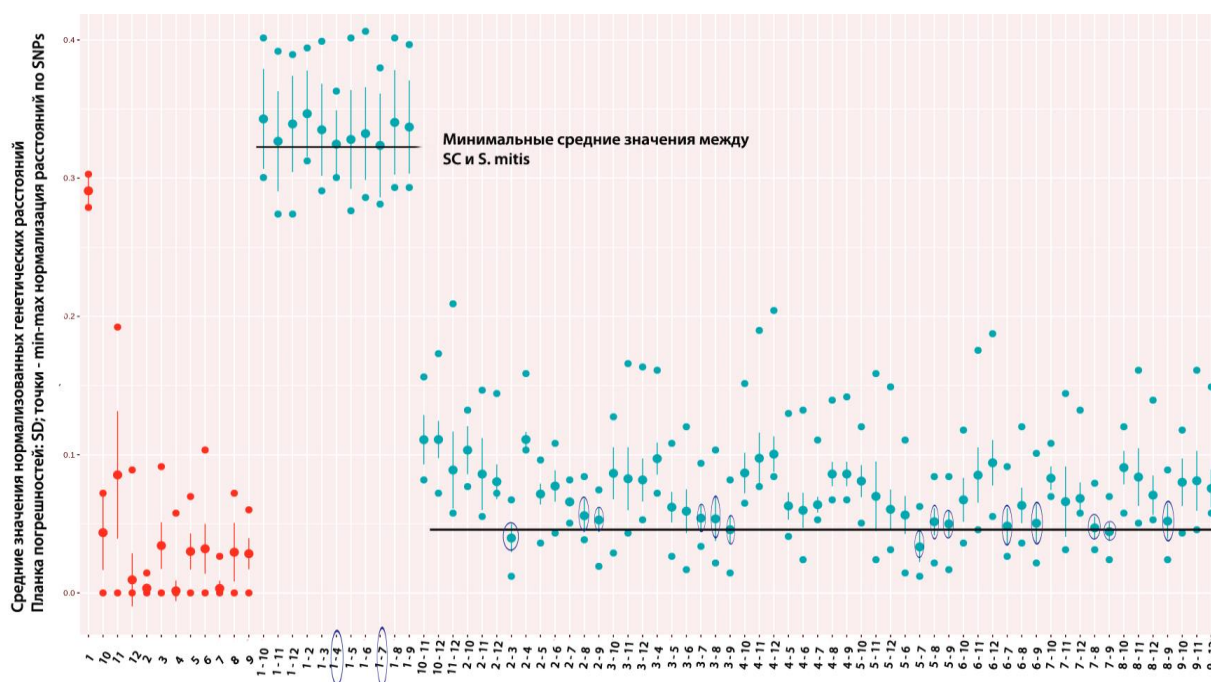
Оценка генетических дистанций между группами A, B1, B2 и *S. mitis* KCOM13 была подтверждена с помощью анализа различий внутри- и межгрупповых средних генетических расстояний, рассчитанных по матрице попарных расстояний, вычисленных по SNP. Внутригрупповые диапазоны стандартизованных значений генетических расстояний для групп B1 и B2 практически не перекрывались с межгрупповыми диапазонами для всех групп (Рисунок 10). Более широкий диапазон генетических расстояний внутри группы A частично перекрывался с диапазонами между группами, что отражало большую степень дивергенции между изолятами группы A. Однако средние значения генетических расстояний между группой A и группами B1 и B2 выше среднего внутригруппового значения для A. Вариабельность изолятов между группами B1 и B2 меньше, чем между представителями группы B и группы A (Рисунок 10). Таким образом, изоляты, входящие в состав группы B1 и B2, генетически ближе друг к другу, чем к изолятам группы A (Рисунок 10). Полученные результаты согласуются с топологией глобальных групп A, B1 и B2 в филогенетическом дереве, построенном с помощью программы SplitsTree.



**Рисунок 10 – Попарные оценки генетических расстояний для групп А, В1 и В2**

Красным и синим цветом показаны внутригрупповые и межгрупповые оценки, соответственно. Приведены стандартизованные значения генетических расстояний после min-max нормализации. А, В1 и В2 – группы *S. pneumoniae*, соответствующие трём глобальным кладам на филогенетическом дереве на Рисунке 4 и Рисунке 6. Группа О – аутгруппа, представленная тремя штаммами *S. mitis*.

Субструктура кладов и топология групп А, В1 и В2 также были подтверждены анализом генетических расстояний между сиквенс-кластерами (Рисунок 11). На Рисунке 11 прямой линией выделены минимальные средние значения генетических расстояний, соответствующие наиболее близким сиквенс-кластерам. При этом группе В1 соответствуют SC2-SC3, SC2-SC9, SC2-SC8, SC3-SC8, SC3-SC9; группе В2 соответствуют SC5-SC7, SC5-SC8, SC5-SC9, SC6-SC7, SC6-SC9, SC7-SC8, SC7-SC9, SC8-SC9. Сиквенс-кластеры клада А (SC4, SC6 и SC10) удалены на большее генетическое расстояние от сиквенс-кластеров, входящих в состав групп В1 и В2 (Рисунок 11). Внутригрупповые и межгрупповые средние значения и 1SD-диапазоны значений попарных генетических расстояний по SNPs для сиквенс-кластеров, как правило, не перекрывались, за исключением SC10 и SC11 (Рисунок 11).



**Рисунок 11 – Попарные оценки генетических расстояний для сиквенс-кластеров SCs 1-12, рассчитанные по SNPs конкатенатов последовательностей генов MLST**

Красным и синим цветом показаны внутригрупповые и межгрупповые оценки, соответственно. Приведены стандартизованные значения генетических расстояний после min-max нормализации. Прямой линией выделены минимальные средние значения нормализованных расстояний между группами: между сиквенс-кластерами пневмококка (снизу) и между сиквенс-кластерами пневмококка и SC1, включающего представителей *S. mitis*.

Вклад каждого гена MLST-схемы в распределение изолятов первичной выборки по группам был оценен путем сравнения индивидуальных выравниваний последовательностей генов *aroE*, *gki*, *gdh*, *recP*, *spi*, *xpt*, экстрагированных из доступных геномов (Том 2 – Приложения. Приложение 4, Рисунки 1–20).

Ген *aroE*, кодирующий шикимат дегидрогеназу, вовлеченную в биосинтез хоризмовой кислоты, участвующей в биосинтезе ароматических соединений, в том числе ароматических аминокислот (например, триптофана) вносил наибольший вклад в дивергенцию CC81 (Том 2 – Приложения. Приложение 4, Рисунки 1–4). Среди представителей CC81 много вирулентных штаммов (выделенных из крови и ликвора). Известно, что особенности метаболизма аминокислот влияют на потенциал вирулентности.

Ген *gki*, кодирующий глюкокиназу, участвующую в фосфорилировании глюкозы с образованием глюкозо-6-фосфата, вносил большой вклад в деление популяции на три группы – B1, B2 и A+B1+B2 (Том 2 – Приложения. Приложение 4, Рисунки 5, 6). При этом значительно выделялись две ветви, дивергировавшие в противоположных

направлениях – CC236/CC271/CC320 (группа В1) и ветвь CC199 и CC304/CC306 (группа В2). Также был идентифицирован промежуточный между двумя группами клад, включающий представителей групп А, В1 и В2.

Кластеризация по гену *gdh* (ген глюкозо-6-фосфат-1-дегидрогеназы) показала, что большая часть популяции была представлена двумя близкими кладами, ассоциированными с многочисленными генетическими линиями и серотипами (Том 2 – Приложения. Приложение 4, Рисунки 7–10), при этом выделялись дивергентные генетические линии:

- CC218 (серотип 12F), CC62 (серотип 11A), CC199 (серотипы 19А и 15BC), CC156 (серотип 9А), CC191 (серотип 7F), CC315 (серотип 6Е);
- CC1094 (серотип 6А), CC185 (серотип 6Е), CC199 (серотипы 19А и 15BC), CC289 (серотип 5), CC236/CC271/CC320 (серотипы 19F и 19А);
- CC327 (серотип 6А), CC344 (бескапсульный), CC448 (бескапсульный);
- CC9 (серотип 14), CC36 (серотип 23F), CC102 (серогруппа 18BC), CC105 (серотип 25F), CC239 (серотип 9V).

Кластеризация по гену *recP* (кодирующему транскетолазу) позволила выявить в популяции пневмококка несколько крупных кладов, каждый из которых объединял различные клональные группы (Том 2 – Приложения. Приложение 4, Рисунки 11–12). Транскетолаза участвует в реакциях неокислительной ветви пентозофосфатного пути, связывающих данный путь с гликолизом, и обеспечивает образование фосфорибозы из метаболитов гликолиза. Распространенные клады при кластеризации по гену *recP* могут отражать наиболее успешные в метаболическом плане сиквенс-типы.

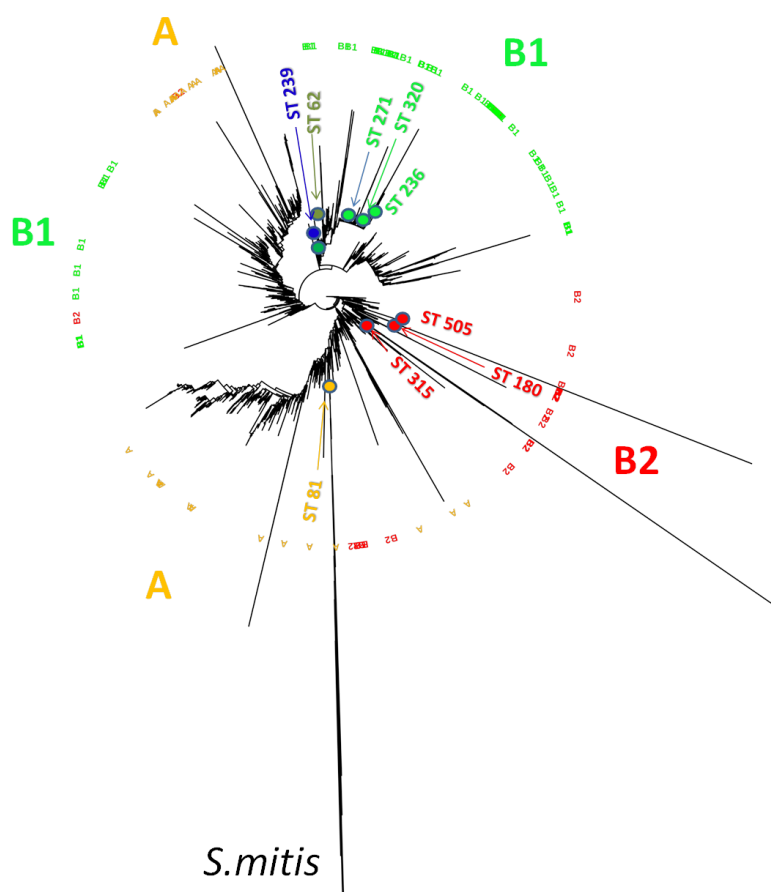
Ген *spi* (кодирующий сигнальную пептидазу I) вносил основной вклад в кластеризацию популяции на группы А и В (Том 2 – Приложения. Приложение 4, Рисунки 13–16). При этом большая часть популяции относилась к гомогенной ветви В, группа А была значительно более гетерогенной. Сигнальная пептидаза I – связанная с мембраной сериновая протеаза, ответственная за процессинг секретлируемых через мембрану белков-предшественников, у большинства представителей группы В должна быть гомологична по структуре и субстратной специфичности.

Кластеризация по гену *xpt*, кодирующему ксантинфосфорибозилтрансферазу, позволила идентифицировать большой клад, представленный генетическими линиями, часто ассоциирующимися с инвазивными изолятами (Том 2 – Приложения. Приложение

4, Рисунки 17–20). Ксантинфосфорибозилтрансфераза превращает ксантин (продукт деградации нуклеиновых кислот) в ксантозин-5'-монофосфат, который может быть повторно использован для синтеза ДНК и РНК, что может быть важно для потенциала вирулентности. Клад, представленный ассоциируемыми с вирулентностью генетическими линиями, преимущественно включал изоляты, принадлежащие группам А и В2. Интересно, что группа В2 ассоциировалась с серотипами, не входящими в состав вакцины ПКВ7. Аналогичные ассоциированные с инвазивными изолятами групп А и В2 клады были идентифицированы при кластеризации по генам *gki* и *gdh*, продукты которых регулируют поток глюкозы в клетку. Таким образом, возможно, что на фоне вакцинации ПКВ7 происходило распространение генетических линий группы В2, ассоциированных с повышенной вирулентностью, имеющих сходные с группой А метаболические типы.

С помощью программы RaхML было построено филогенетическое дерево по данным выравнивания конкатенатов фрагментов 6-ти генов MLST для 14663 сиквенс-типов, доступных на август 2019 г. в базе данных PubMLST. Соответствующие последовательности *S. mitis* KCOM13 были использованы в качестве аутгруппы. На дереве были обозначены сиквенс-типы из анализируемой первичной выборки. Распределение сиквенс-типов нашей первичной выборки на глобальном дереве сиквенс-типов (Рисунок 12) имело сходную тенденцию с кластеризацией, полученной только на основе первичной выборки (Рисунок 4). Популяция, представленная 14663 сиквенс-типами *S. pneumoniae* делилась на три глобальных клада. Наиболее близкой к аутгруппе (*S. mitis*) была генетическая линия СС81 (группа А). Группа А занимала промежуточное положение между группами В1 и В2 и была более гетерогенной по сравнению с ними, при этом промежуточный клад включал сиквенс-типы, относящиеся к группам А, В1 и В2 (выделенным на основании кластеризации изолятов первичной выборки). Генетическая линия СС81 была ближе к группе В2. Данную тенденцию кластеризации сиквенс-типов можно объяснить вкладом гена *gki*, кодирующего глюкокиназу. Именно ген *gki* объясняет топологию дерева, построенного по первичной выборке (Том 2 – Приложение. Приложение 4, Рисунок 5), при этом вариант гена *gki* представителей генетической линии СС81 и СС448 (группа А) и части представителей группы В2 (СС66, СС180, СС191, СС193, СС230, СС505, СС1262) является идентичным (Том 2 – Приложение. Приложение 4, Рисунок 5). С другой стороны, деление первичной

популяции на группы А детерминируется геном *spi*, кодирующим сигнальную пептидазу I (Том 2 – Приложения. Приложение 4, Рисунок 13), при этом группа В является значительно более гомогенной по гену *spi*, чем группа А. Таким образом, по результатам кластеризации сиквенс-типов первичной выборки и сиквенс-типов генеральной совокупности, а также по времени идентификации изолятов, принадлежащих распространенным сиквенс-типам, можно предположить, что после начала применения ПКВ7 в мире эволюционируют и распространяются представители генетических линий групп В1 и В2. При этом распространяющиеся генетические линии имеют общие особенности процессинга секретируемых белков.



**Рисунок 12 – Филогенетическое дерево *S. pneumoniae* (RaXML) по конкатенатам 6-ти генов домашнего хозяйства для 14663 сиквенс-типов, доступных на август 2019 г. в базе данных PubMLST.**

Последовательности *S. mitis* KCOM13 использованы в качестве аутгруппы. Популяция, представленная 14663 сиквенс-типами *S. pneumoniae*, делится на три глобальных клада. На дереве обозначены сиквенс-типы из анализируемой первичной выборки. Распределение сиквенс-типов первичной выборки на глобальном дереве сиквенс-типов имеет сходную тенденцию с кластеризацией, полученной только на основе первичной выборки (Рисунок 4). Группа А занимает промежуточное положение между группами В1 и В2 и более гетерогенна по сравнению с В1 и В2.



### 3.1.2. Анализ популяции *S. pneumoniae* по ядерной части генома

Вторичная выборка изолятов с доступными данными полногеномного секвенирования включала 495 образцов (132 изолята из России; Раздел 2.2 «Формирование выборки»), в том числе 134 инвазивных изолята (источник выделения - кровь, ликвор), 33 – из жидкости среднего уха, 210 – от носителей (носоглотка). Включены изоляты, принадлежащие всем идентифицированным ранее SC\_MLST (Рисунок 13) и распространенным серотипам (Рисунок 14; Том 2 – Приложения. Приложение 2 – Таблица 1).

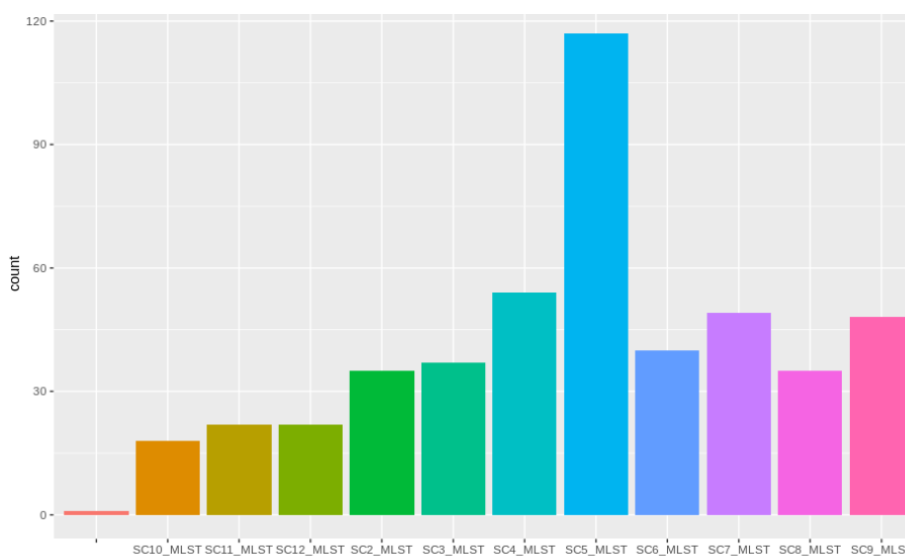


Рисунок 13 – Гистограмма распределения изолятов между группами SC\_MLST в субвыборке с данными полногеномного секвенирования (n=495)

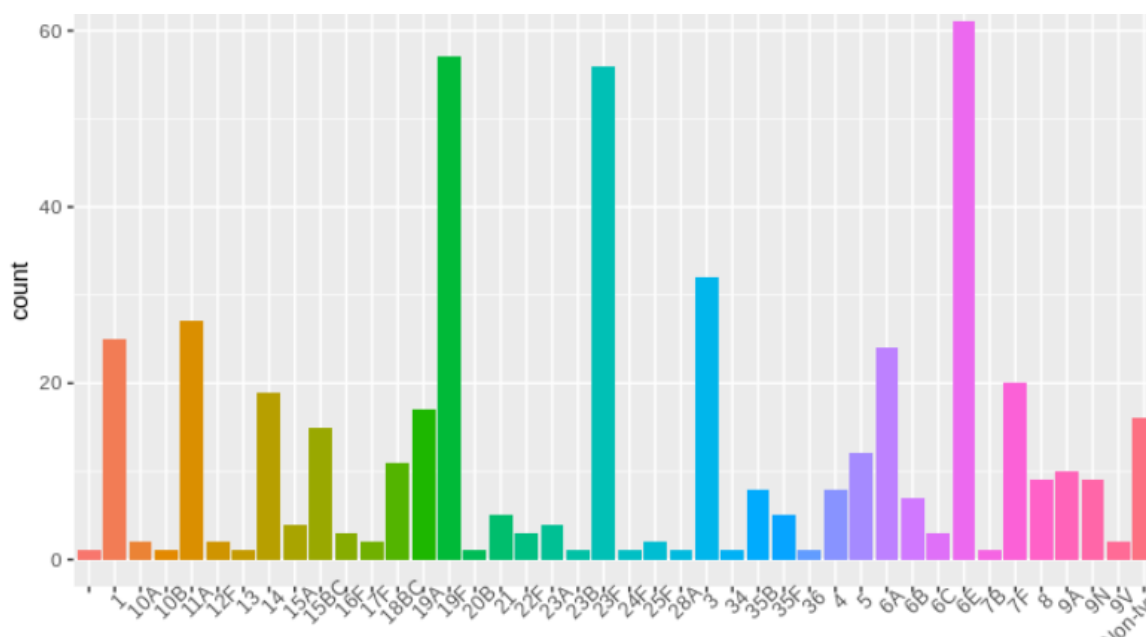
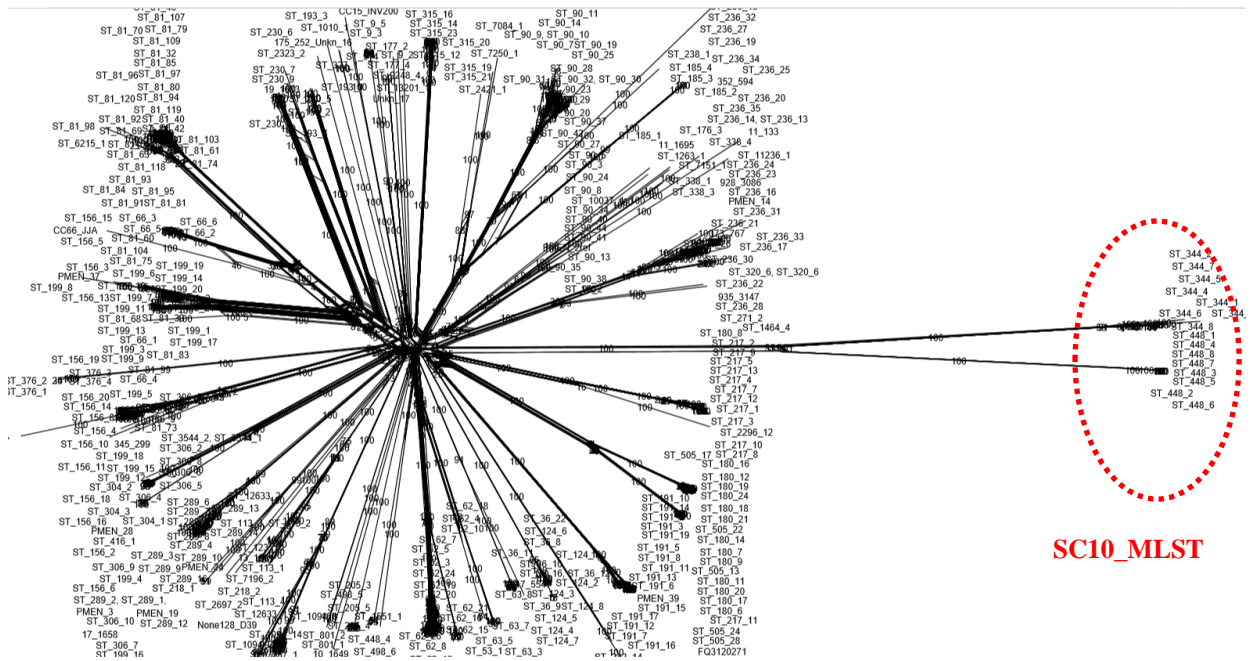


Рисунок 14 – Гистограмма распределения изолятов различных серотипов в субвыборке с данными полногеномного секвенирования (n=495)

Клональность популяции пневмококка по ядерной части генома была оценена с помощью *hierBAPS* с использованием 49974 SNPs, экстрагированных из полногеномного выравнивания с фильтром сайтов рекомбинаций, выполненного программой *Parsnp* (Раздел 2.6. «Анализ ядерной части генома»). Сиквенс-кластеры начального уровня, идентифицированные по SNPs ядерной части генома, преимущественно включали сиквенс-типы, относящиеся к родственным клональным комплексам: CC81/ST6251, CC62/CC53, CC180/CC505, CC66/CC230/CC199, CC156/CC143, CC1094/SLVs, CC90/ST238, CC217/CC304/CC306, CC448, CC236/CC271/CC320/SLVs, CC344.

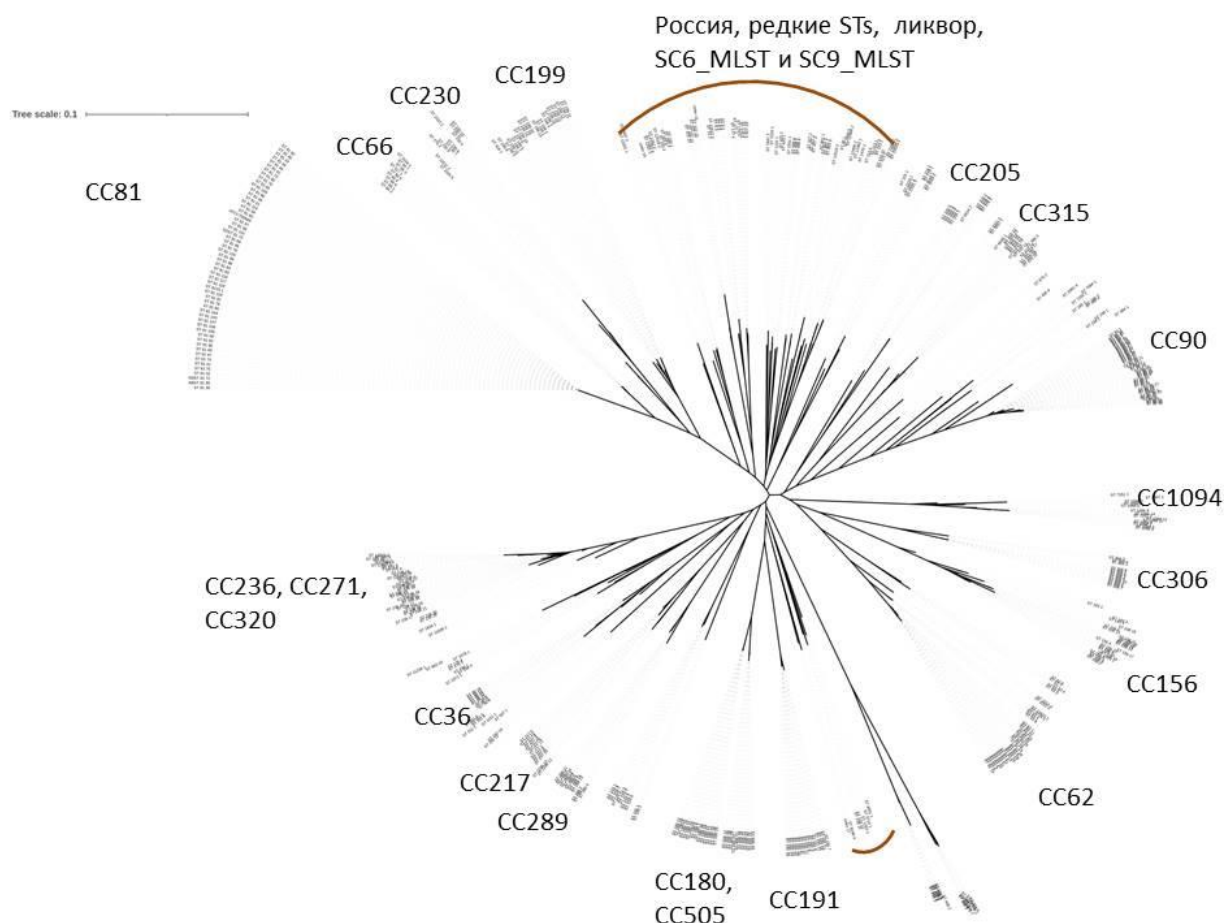
Филогенетические деревья, основанные на ядерной части генома, были построены двумя способами: с помощью пакета *SplitsTree* по выравниванию SNPs ядерной части геномов, экстрагированных из выравнивания геномов в *Parsnp* (фильтр рекомбинаций выполнен в *Parsnp*) (Рисунок 15); а также с помощью пакета *RaxML* по выравниванию ядерной части геномов после фильтра рекомбинаций в программе *Gubbins* (Раздел 2.6. «Филогенетический анализ ядерной части генома») (Рисунок 16). Неукорененные деревья характеризовались звездчато-подобной структурой (Рисунок 15 и 16), почти все узлы которой были равноудалены от гипотетического корня, что отражало панмиктичность популяции пневмококка. Наиболее дивергировавшей от всей популяции пневмококка была группа изолятов, представленная нетипируемыми бескапсульными пневмококками CC344 и CC448 (Рисунок 15). Данные изоляты находились от остальной популяции на генетическом расстоянии, приблизительно соответствующем расстоянию остальных пневмококков от общего предшественника. Некоторые исследователи предлагают считать представителей CC344 и CC448 отдельным видом [128]. Значительно дивергировали в противоположных направлениях группы генетических линий CC81 / CC66 / CC230 / CC199 и CC236/CC271/CC320 (Рисунок 16).



**Рисунок 15 – Филогенетическое дерево (SplitsTree) по SNPs ядерной части генома после выравнивания геномов и фильтра рекомбинаций в Parsnp (n=495).**

Филогенетический анализ ядерной части генома, а также выделение сиквенса-кластеров второго и третьего уровня по ядерной части генома, позволили установить родство ряда редких сиквенса-типов, циркулирующих в России, с распространенными клональными комплексами (Том 2 – Приложения. Приложение 3 – Таблица 1).

Значительная часть редких сиквенса-типов российской популяции пневмококков была представлена инвазивными штаммами, которые кластеризовались в виде отдельных ветвей филогенетического дерева и для них не удалось установить родственных распространенных клональных групп (Рисунок 16, Том 2 – Приложения. Приложение 3 – Таблица 1). Все изоляты данных редких сиквенса-типов были выделены после 2011 года. Таким образом, на фоне вакцинации в мире вакцинами ПКВ7 и ПКВ13 происходило распространение в России инвазивных изолятов редких генетических линий, многие из которых ассоциировались с невакцинными серотипами (6С, 8, 10А, 16F, 17F, 22В, 22F, 25F, 34) (Том 2 – Приложения. Приложение 3 – Таблица 1).



**Рисунок 16 – Филогенетическое дерево (RaxML) по выравниванию ядерной части генома после фильтра рекомбинаций в Gubbins (n=495).**

Значительная часть инвазивных изолятов из России (ликвор) принадлежала редким сиквенс-типам и ассоциировалась с SC6\_MLST и SC9\_MLST.

### **3.1.3. Оценка вклада генов ядерного генома в формирование структуры популяции *S. pneumoniae* (групп A/B1/B2, сиквенс-кластеров, серотипов, инвазивных штаммов резистентных к антибиотикам штаммов)**

Для оценки вклада генов ядерного генома в структуру популяции пневмококка были использованы номера аллелей, идентифицированные при сравнении геномов в GenomeComparator (Раздел 2.7.1. Идентификация аллелей генов). Первоначальный анализ структуры данных был выполнен с помощью множественного анализа соответствий (Раздел 2.7.2. Множественный анализ соответствий (Multiple correspondens analisys, MCA)). Предварительная визуализация данных показала, что варианты генов полного генома описывают группы сиквенс-кластеры ядерного генома (Рисунок 17), SC\_MLST (Рисунок 18) и серотипы. Однако каждое измерение описывало не более 1,3% данных всего генома. При этом кумулятивная вариативность данных, описываемая в измерениях, составила:

- 100 измерений – 70,93%
- 200 измерений – 85,74%
- 300 измерений – 93,22%
- 334 измерений – 95,04%
- 487 измерений – 99,99%

Полученные данные отражают многообразие паттернов взаимодействующих генов, которые вносят вклад в структуру популяции.

С помощью МСА была получена альтернативная вторичная матрица, включающая гены с высоким качеством вариативности ( $\cos^2 > 0,5$ ), объясняющие данные как можно в большем числе измерений одновременно (в первых  $\dim_{20}$ ,  $\dim_{27-30}$ ,  $\dim_{57}$  из 100 измерений). Вторичная матрица переменных с высоким качеством вариативности для 20 измерений включала 228 переменных, в том числе 214 генов (Том 2 – Приложение. Приложение 5 – Таблица 1). Первичная и вторичная матрицы использовались в дальнейшем для оценки ошибки классификации с помощью алгоритма RandomForest.

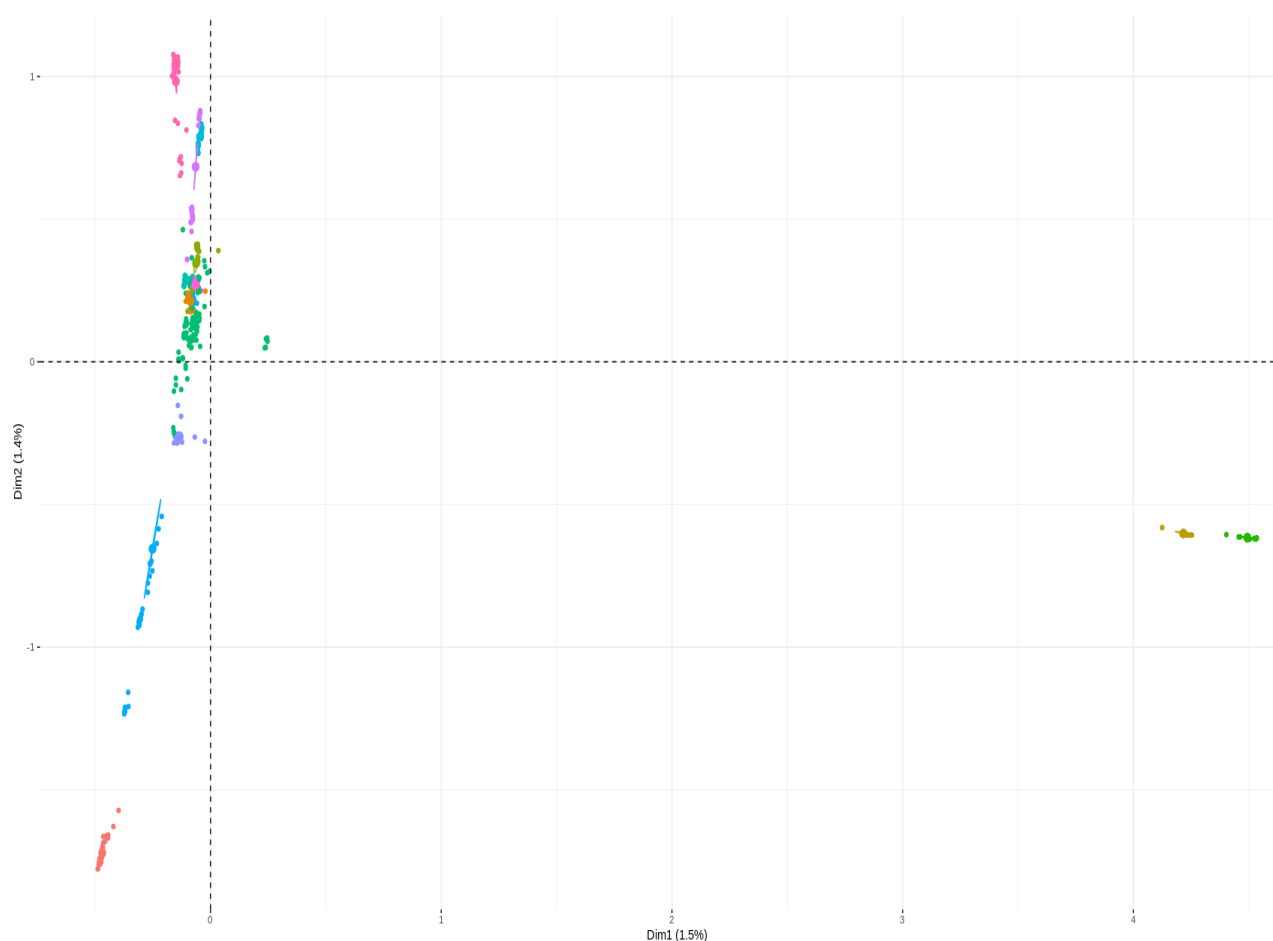
Классификация на группы «SC\_MLST», «A/B1/B2», «Серотип», «Источник выделения» по генам полного генома референса, а также по 214 генам с высоким качеством вариативности, была подтверждена с помощью алгоритма RandomForest. Матрицы несоответствий предсказаний приведены в Таблицах 3–8, значения ошибок предсказания по групповым переменным (out-of-bag оценка) приведены в Таблицах 9–12. Кластеризация изолятов по группам «A/B1/B2», «SC\_MLST», «Источник выделения», «Резистентность к пенициллину –МПК», «Резистентность к эритромицину – МПК», в зависимости от вариантов генов, показана на Рисунках 19-23. Несмотря на наибольшую гетерогенность по составу, группа SC5\_MLST была валидирована с наилучшей оценкой ошибки предсказания (Таблица 9). Группы A/B1/B2 лучше кластеризуются по данным всего генома, чем по 214 генам (Таблица 10, Рисунок 19).

Серотипы предсказывались достаточно точно по данным всего генома. На значение общей ошибки классификации повлияла большая погрешность при предсказании редких серотипов, представленных единичными штаммами (результаты для редких серотипов исключены из Таблиц 6 и 11. Значительная ошибка в предсказании серотипа 6В обусловлена отнесением изолятов к серотипу 6Е (*cps*-локусы данных серотипов отличаются мутациями только в двух генах – *wze*, *wciN*) (Таблица 6).

Ошибки в предсказании серотипов 14, 19F, 23F, 19A, 18BC могут быть связаны с тем, что данные серотипы ассоциируются с разными генетическими линиями.

При анализе инвазивных изолятов наименьшая ошибка предсказания по источнику выделения была получена по данным матрицы 214 генов (Таблицы 7, 12; Рисунок 21).

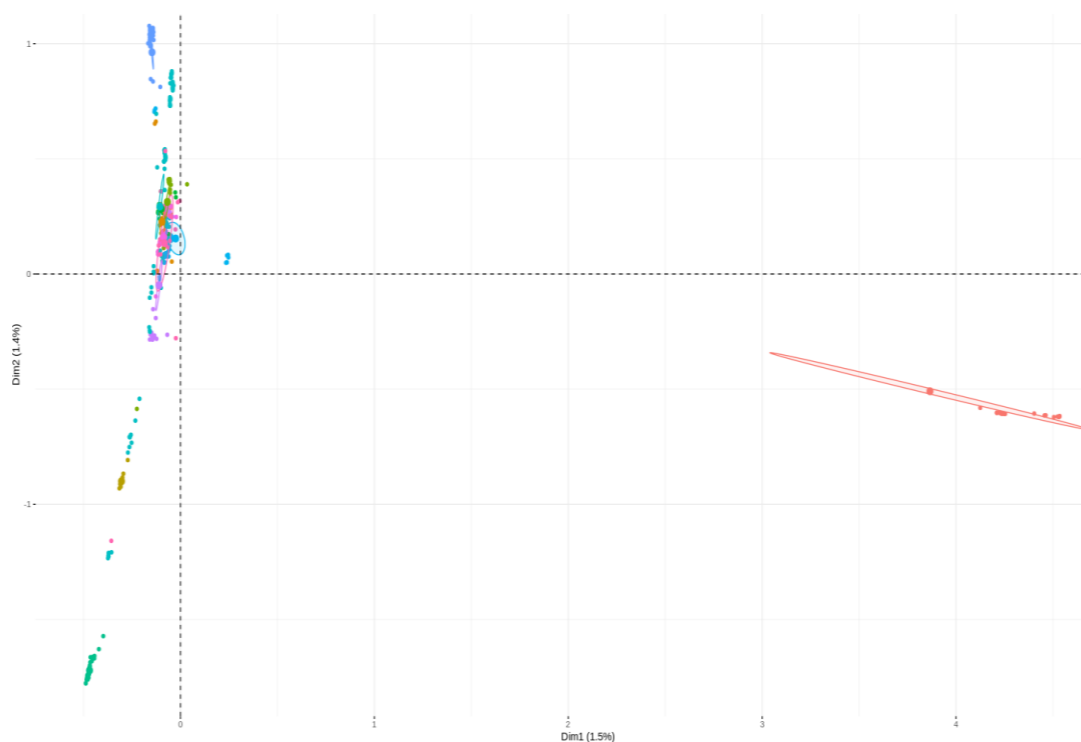
Поиск ассоциаций вариаций аллелей генов с анализируемыми группами изолятов также был выполнен с помощью алгоритма машинного обучения XGBoost (см. «Материалы и методы», раздел 2.7.4). Данный алгоритм позволил идентифицировать варианты генов, по которым происходит расщепление анализируемых групп – определить «границы» между группами.



**Рисунок 17 – Кластеризация изолятов с помощью МСА по вариантам генов полного генома референса ATCC 700669 по групповой переменной «SC-core»**

Соответствие изолятов сиквенс-кластерам по данным ядерного генома (SC-core):





**Рисунок 18 – Кластеризация изолятов с помощью МСА по вариантам генов полного генома референса ATCC 700669 по групповой переменной «SC-MLST»**

Соответствие изолятов сиквенс-кластерам по данным MLST:



**Таблица 3 - Матрица несоответствий результатов классификации (Random Forest) для групповой переменной «SC-MLST»**

		Предсказанные											Недоступны для out-of-bag оценки
		SC10	SC11	SC12	SC2	SC3	SC4	SC5	SC6	SC7	SC8	SC9	
Фактические	SC10	16	0	0	0	0	0	2	0	0	0	0	0
	SC11	0	19	0	0	0	0	3	0	0	0	0	0
	SC12	0	0	22	0	0	0	0	0	0	0	0	0
	SC2	0	0	0	33	0	0	2	0	0	0	0	0
	SC3	0	0		1	31	0	5	0	0	0	0	0
	SC4	0	0	0	0	0	54	0	0	0	0	0	0
	SC5	0	0	0	0	0	0	116	1	0	0	0	0
	SC6	0	0	0	0	0	0	8	32	0	0	0	0
	SC7	0	0	0	0	0	0	0	0	49	0	0	0
	SC8	0	0	0	0	0	0	0	0	0	35	0	0
	SC9	0	0	0	0	0	0	2	0	0	0	46	0

Для классификации использовались аллели всех генов референсного генома ATCC 700669; Результаты предсказания, соответствующие фактическим данным, отмечены зеленым цветом; Результаты предсказания, не соответствующие фактическим данным, отмечены красным цветом.

**Таблица 4 - Матрица несоответствий результатов классификации (Random Forest) для групповой переменной «группы А/В1/В2» по всем генам референсного генома**

	Предсказанные				Недоступны для out-of-bag оценки
		А	В1	В2	
Фактические	А	106	0	6	0
	В1	0	105	5	0
	В2	0	0	220	0

Для классификации использовались аллели всех генов референсного генома ATCC 700669; Результаты предсказания, соответствующие фактическим данным, отмечены зеленым цветом; Результаты предсказания, не соответствующие фактическим данным, отмечены красным цветом.

**Таблица 5 - Матрица несоответствий результатов классификации (Random Forest) для групповой переменной «группы А/В1/В2» по 214 генам**

	Предсказанные				Недоступны для out-of-bag оценки
		А	В1	В2	
Фактические	А	82	0	30	0
	В1	0	105	47	0
	В2	0	0	154	0

Для классификации использовались аллели 214 наиболее вариативных генов; Зеленый цвет – соответствие фактическим данным; красный цвет – несоответствие.

**Таблица 6 - Матрица несоответствий результатов классификации (Random Forest) для групповой переменной «Серотип»**

	Предсказанные																				
	1	11A	14	15BC	18BC	19A	19F	23F	3	35B	4	5	6A	6B	6C	6E	7F	9A	9N	NA	
Факти-ческие	1	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	11A	0	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	14	0	0	17	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0
	15BC	0	0	0	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	18BC	0	0	0	0	9	0	2	0	0	0	0	0	0	0	0	0	0	0	0	
	19A	0	0	0	0	0	13	2	1	0	0	0	0	0	0	0	0	0	0	0	
	19F	0	0	1	2	0	0	54	0	0	0	0	0	0	0	0	0	0	0	0	
	23F	0	0	0	0	0	0	0	56	0	0	0	0	0	0	0	0	0	0	0	
	3	0	0	0	0	0	0	1	0	31	0	0	0	0	0	0	0	0	0	0	
	35B	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	2	0	
	4	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	
	5	0	0	0	0	0	0	0	0	0	0	0	12	0	0	0	0	0	0	0	
	6A	0	0	0	0	0	0	0	0	0	0	0	0	21	0	0	2	0	0	0	
	6B	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	5	0	0	0	
	6C	0	0	0	0	0	0	1	0	0	0	0	0	2	0	1	0	0	0	0	
	6E	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	61	0	0	0	
	7F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	
	9A	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	9	0	
	9N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	

Для классификации использовались аллели всех генов референсного генома ATCC 700669; Зеленый цвет – соответствие фактическим данным; красный цвет – несоответствие; NA – Недоступны для out-of-bag оценки.



**Таблица 7 - Матрица несоответствий результатов классификации для групповой переменной «Источник выделения» по всем генам референсного генома**

Random Forest	Предсказанные					
		кровь	ЦСЖ	ж.с.у.	м.н.г.	Недоступны для out-of-bag оценки
Фактические	кровь	41	2	0	13	0
	ЦСЖ	9	49	0	20	0
	ж.с.у.	2	5	13	13	0
	м.н.г.	7	8	1	194	0

Для классификации использовались аллели всех генов референсного генома ATCC 700669; Зеленый цвет – соответствие фактическим данным; красный цвет – несоответствие.

**Таблица 8 - Матрица несоответствий результатов классификации для групповой переменной «Источник выделения» по 214 генам**

Random Forest	Предсказанные					
		кровь	ЦСЖ	ж.с.у.	м.н.г.	Недоступны для out-of-bag оценки
Фактические	кровь	46	3	0	7	0
	ЦСЖ	8	60	0	10	0
	ж.с.у.	2	6	17	8	0
	м.н.г.	2	7	0	201	0

**Таблица 9 – Ошибка предсказания для групповой переменной SC\_MLST (весь геном)**

Random Forest	Общая ошибка	SC2_MLST	SC3_MLST	SC4_MLST	SC5_MLST	SC6_MLST	SC7_MLST	SC8_MLST	SC9_MLST	SC10_MLST	SC11_MLST	SC12_MLST
Ошибка предсказания	5,03	5,71	16,22	0	0,85	20	0	0	4,17	11,11	13,64	0

**Таблица 10 – Ошибка предсказания для групповой переменной A/B1/B2**

Random Forest	Общая ошибка	A	B1	B2
		Ошибка предсказания (Out-of-bag error)		
весь геном	2,49	5,36	4,54	0
214 генов	17,42	26,78	42,73	0

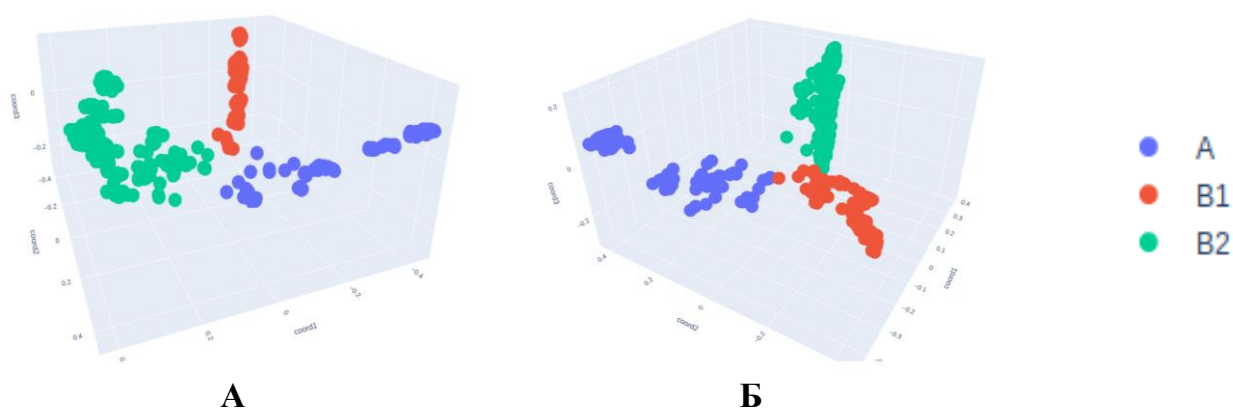
**Таблица 11 – Ошибка предсказания для переменной Серотип (весь геном)**

Random Forest	Общая ошибка – 9,81										
Серотип	1	3	4	5	6A	6B	6E	7F	9A	9N	11A
Ошибка предсказания	0	3,12	0	0	12,5	71,43	0	0	10	0	0
Серотип	14	15BC	18BC	19A	19F	23F	35B				
Ошибка предсказания	10,53	0	18,18	23,53	5,26	0	25				

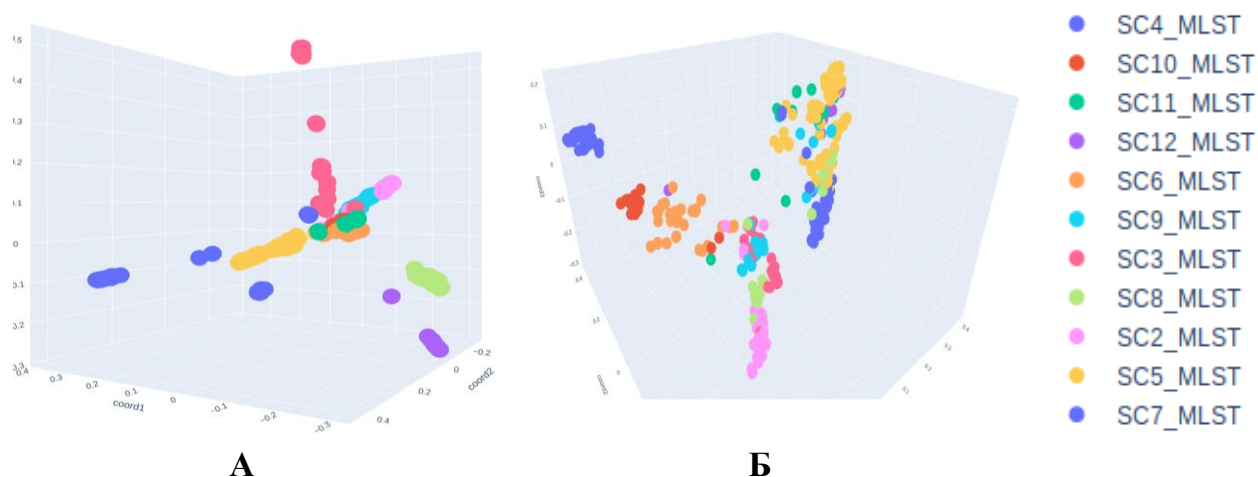
**Таблица 12 – Ошибка предсказания по переменной «Источник выделения»**

Random Forest	Общая ошибка	м.н.г. / носители	ж.с.у.	кровь	ЦСЖ
		Ошибка предсказания (Out-of-bag error)			
весь геном	21,22	7,62	60,6	26,78	37,18
214 генов	14,06	4,28	48,48	17,86	23,08

м.н.г. – мазок из носоглотки/носительство, крв. – кровь, ЦСЖ – цереброспинальная жидкость, ж.с.у. – жидкость среднего уха при среднем отите.

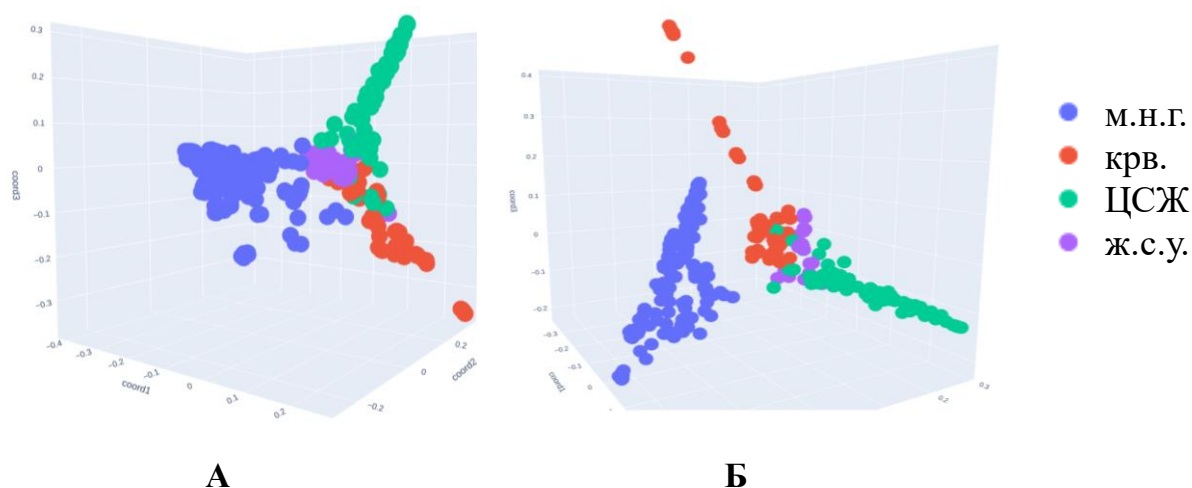
**Рисунок 19 – Кластеризация изолятов по переменной «Группы A/B1/B2» (RandomForest), в зависимости от вариантов генов.**

А – для классификации использовались все гены референсного генома ATCC 700669  
 Б – для классификации использовались 214 генов



**Рисунок 20 – Кластеризация изолятов по переменной «SC\_MLST» (RandomForest), в зависимости от вариантов генов.**

А – для классификации использовались все гены референсного генома ATCC 700669  
 Б – для классификации использовались 214 генов

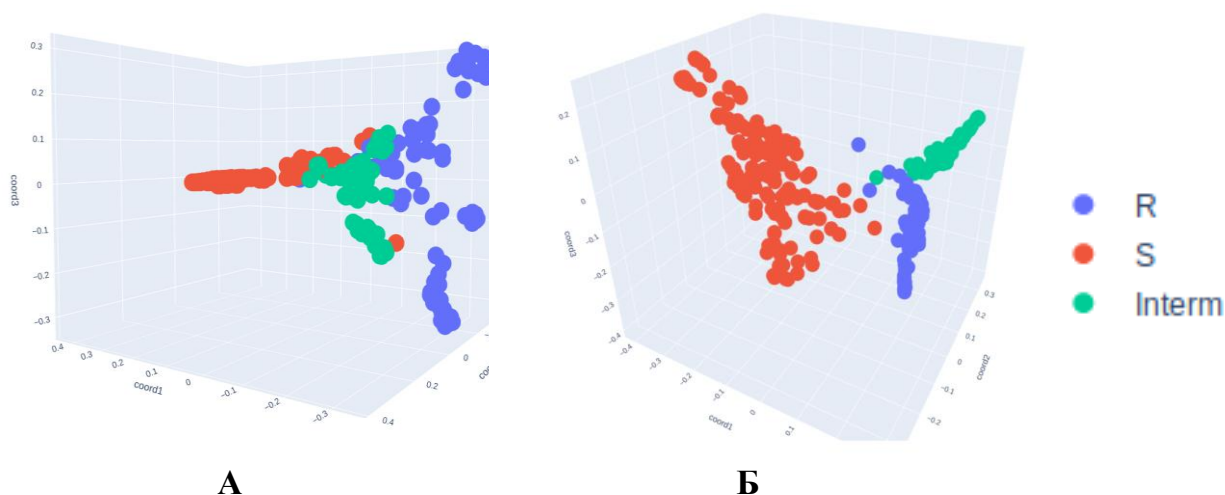


**Рисунок 21 – Кластеризация изолятов по переменной «Источник выделения» (RandomForest), в зависимости от вариантов генов.**

А – для классификации использовались все гены референсного генома ATCC 700669  
 Б – для классификации использовались 214 генов

м.н.г. – мазок из носоглотки/носительство, крв. – кровь, ЦСЖ – цереброспинальная жидкость, ж.с.у. – жидкость среднего уха при среднем отите.

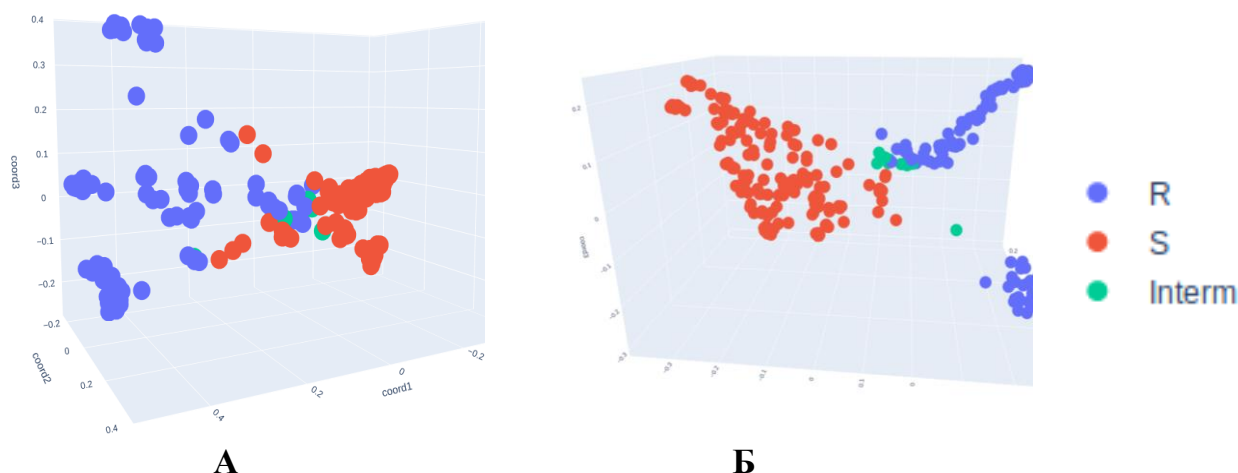
Результаты классификации RandomForest по анализируемым групповым переменным «Резистентность к пенициллину – МПК», «Резистентность к эритромицину – МПК» визуализированы на Рисунках 22, 23.



**Рисунок 22 – Кластеризация изолятов по переменной «Резистентность к пенициллину – МПК» (RandomForest), в зависимости от вариантов генов.**

А – для классификации использовались все гены референсного генома ATCC 700669  
 Б – для классификации использовались 214 генов

МПК пенициллина (EUCAST): R – резистентный ; Interm – промежуточная чувствительность, S - чувствительный



**Рисунок 23 – Кластеризация изолятов по групповой переменной «Резистентность к эритромицину – МПК» (RandomForest), в зависимости от вариантов генов.**

А – для классификации использовались все гены референсного генома ATCC 700669

Б – для классификации использовались 218 генов

МПК эритромицина (EUCAST): R – резистентный; Interm – промежуточная чувствительность, S - чувствительный

### 3.1.3.1. Гены, формирующие группы A/B1/B2A

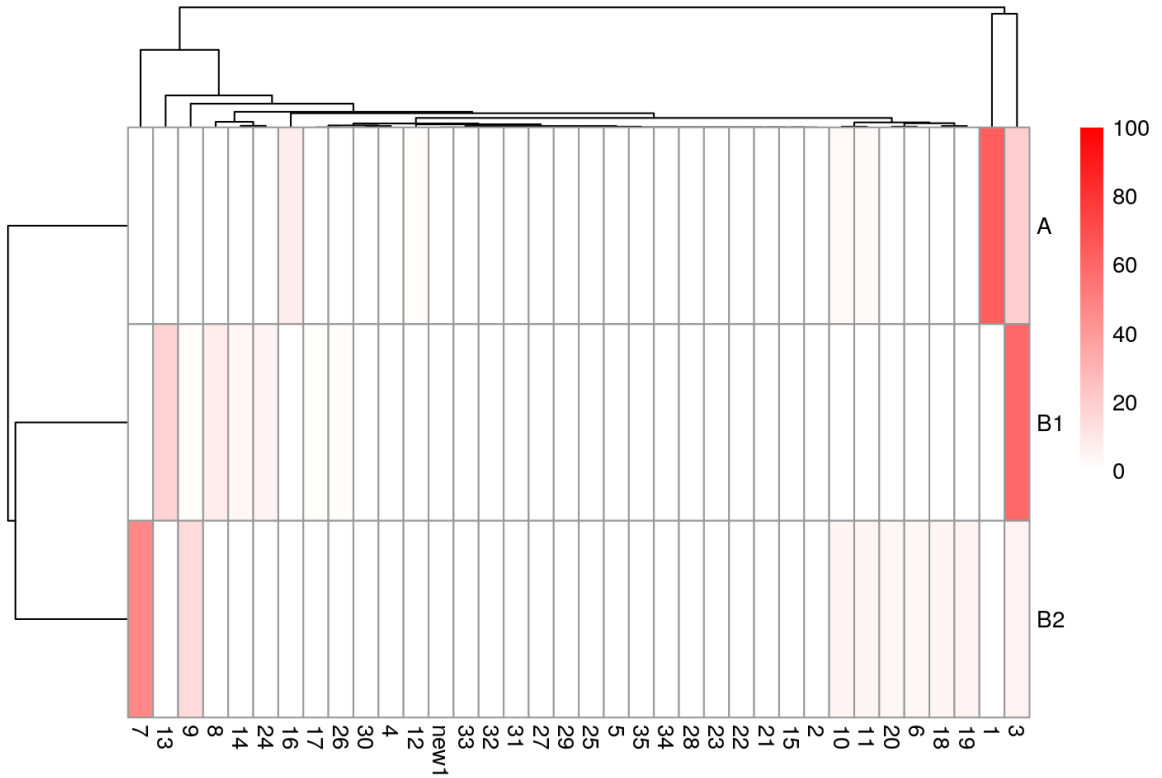
Перечень из 181 значимых генов (по значениям Z-score взаимодействующих переменных в RandomForest, Том 2 – Приложения. Приложение б) включал компоненты различных метаболических путей: биосинтеза жирных кислот, метаболизма галактозы, фруктозы и маннозы, пирувата, пуриновых оснований, биотина, цистеина и метионина, аланина, аспартата и глутамата, фенилаланина, аргинина, а также факторы

вирулентности (ABC-транспортеры, компоненты PTS-системы, металл-транспортирующая АТФаза Р-типа, фибронектин / фибриноген-связывающий белок FpbS, спермидин/путресцин-связывающий белок PotD и другие). Наилучшие маркеры групп A/B1/B2 – ген *accC* (ацетил-СоА-карбоксилазы, участвующей в первых реакциях синтеза жирных кислот); ген *ciaH* (сенсорной гистидин-киназы двухкомпонентной регуляторной системы CiaRH); ген *fabZ* ((3R)-гидроксимиристоил-дегидратазы, участвующей в биосинтезе жирных кислот и биосинтезе липида А) (Рисунки 24–26). Значение нормализованного коэффициента значимости для генов *accC* и *ciaH*, по предсказанию XGBoost – 1 и 0,84, соответственно.

Фермент АссС вовлечен в первичные реакции метаболизма жирных кислот. Биосинтез жирных кислот является первым этапом формирования мембранных липидов. Состав жирных кислот мембраны варьирует, в зависимости от условий окружающей среды, и играет важную роль в адаптации бактериальных клеток к внешним изменениям. В частности, известно, что пневмококк прекращает синтез своих липидов *de novo* в присутствии экзогенного источника жирных кислот. Мутации в гене *accC* могут приводить к снижению каталитической активности фермента. Вариант 1 гена *accC* объединял 68 из 109 представителей генетически гетерогенной группы А – изоляты CC81 (SC4\_MLST), CC124 (SC6\_MLST) и CC448 (SC10\_MLST) и отличал данные генетические линии от остальной популяции пневмококка.

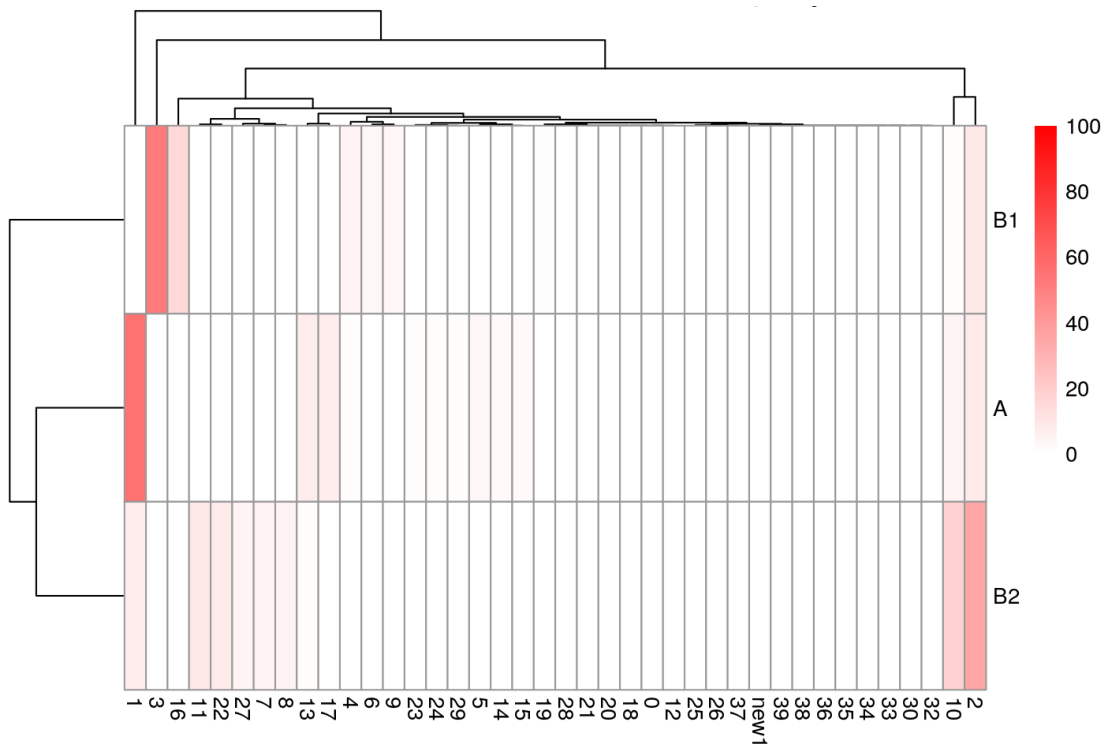
Кластеризация изолятов на основании сравнения последовательностей гена *accC* также позволила идентифицировать группу А (Том 2 – Приложения. Приложение 4, Рисунки 23, 24). Кроме того, несмотря на то, что ген *accC* очень консервативен, кластеризация по его последовательностям ассоциируются с инвазивностью (Том 2 – Приложения. Приложение 4, Рисунок 24).

Двухкомпонентная регуляторная система CiaRH участвует в формировании резистентности к β-лактамам антибиотикам, регуляции естественной компетентности и вирулентности. Вариант 3 гена *ciaH* объединял представителей группы В1 (SC2\_MLST и SC3\_MLST, а также некоторых изолятов SC9\_MLST (CC63)), представленных серотипами 19F и 11A.



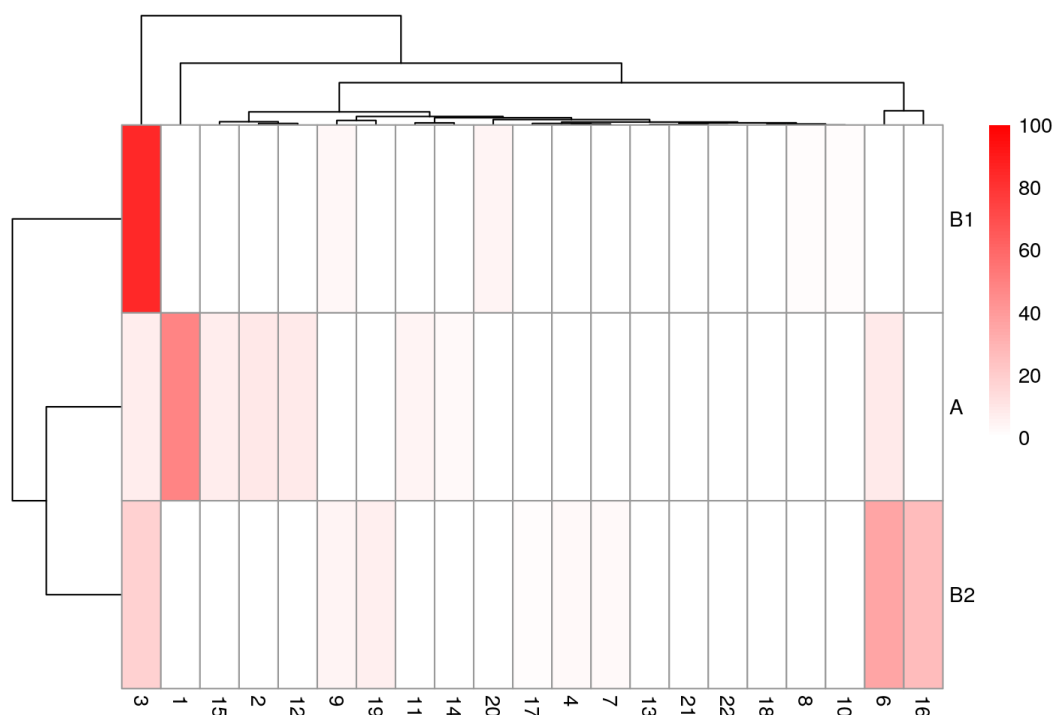
**Рисунок 24 – Гистограмма ассоциации между вариантами гена *accS* и группами A/B1/B2**

По горизонтальной оси отложены варианты исследуемого гена, по вертикальной оси – группы изолятов. Цвет ячейки отражает количество изолятов.



**Рисунок 25 – Гистограмма ассоциации между вариантами гена *ciaH* и группами A/B1/B2**

По горизонтальной оси отложены варианты исследуемого гена, по вертикальной оси – группы изолятов. Цвет ячейки отражает количество изолятов.



**Рисунок 26 – Гистограмма ассоциации между вариантами гена *fabZ* и группами A/B1/B2**

По горизонтальной оси отложены варианты исследуемого гена, по вертикальной оси – группы изолятов. Цвет ячейки отражает количество изолятов.

Филогенетическое дерево на основании сравнения последовательностей гена *ciaH* подтверждало результаты классификации по вариантам генов и было представлено четырьмя распространенными кладами, а также множеством небольших дивергентных кладов (Том 2 – Приложения. Приложение 4, Рисунки 25, 26). Наблюдалась тенденция, отражающая кластеризацию изолятов на группы B1, B2 и A+B2. Клад, представленный CC156, CC236, CC271 и CC320, объединял почти всех представителей группы B1. Также наблюдалась тенденция ассоциации кластеризации с инвазивными штаммами (Том 2 – Приложения. Приложение 4, Рисунки 25, 26). Интересно, что изоляты сиквенс-типа ST199, ассоциирующиеся с разными серотипами (15BC и 19A), кластеризовались с разными кладами дерева *ciaH*. Таким образом, скорее принадлежность к серотипу, чем к сиквенс-типу является определяющим фактором для кластеризации по последовательностям гена *ciaH*.

Филогенетическое дерево на основании сравнения последовательностей гена *fabZ* позволило выделить в популяции пневмококка два распространенных клада, дивергировавших в противоположных направлениях и представленных преимущественно изолятами B1 или B2, соответственно. Также дивергировал клад, представленный CC81 и CC124 (Том 2 – Приложения. Приложение 4, Рисунки 27, 28).

FabZ, (3R)-гидроксимиристоил-дегидратаза, катализирует дегидратацию короткоцепочечных бета-гидроксиацил-АССР[ацил-несущих белков], а также длинноцепочечных насыщенных и ненасыщенных бета-гидроксиацил-АССР белков. FabZ участвует в биосинтезе жирных кислот и биосинтезе липида А.

Таким образом, деление популяции пневмококка на группы А, В1 и В2 детерминировано:

- процессингом секретируемых белков (ген сигнальной пептидазы I, *spi*);
- характером усвоения глюкозы (ген глюкокиназы, *gki*);
- особенностями синтеза жирных кислот, определяющими построение мембраны и клеточной стенки (*accC*, *fabZ*);
- регуляторной системой *ciaRH*, определяющей серотип, вирулентные свойства (*ciaH*).

### 3.1.3.2. Гены, участвующие в формировании сиквенс-кластеров (SC\_MLST)

Перечень 125 значимых генов (по значениям Z-score взаимодействующих переменных в RandomForest, Том 2 – Приложения. Приложение 7) включал: компоненты метаболизма углеводов (глюкокиназа, лактатдегидрогиназа и др.), метаболизма пуринов, транскрипции генов рибосомальной РНК, трансляции, гомологичной рекомбинации, *CiaRH* регуляторной системы, синтеза жирных кислот, синтеза пептидогликана. Наиболее значимые гены, расщепляющие популяцию пневмококка на группы SC\_MLST, приведены в Таблице 13 и в Приложении (Том 2 – Приложения. Приложение 7. Таблица 1).

Кластеризация филогенетического дерева на основании сравнения последовательностей гена *ciaH* отражала тенденцию распределения изолятов по группам SC\_MLST (SC2\_MLST и SC3\_MLST; SC4\_MLST; SC5\_MLST; SC7\_MLST) (Том 2 – Приложения. Приложение 4, Рисунки 25, 26).

Большая часть представителей группы SC5\_MLST (В2) могли обладать общими особенностями сигнальной регуляторной системы *ciaRH*, но при этом они были очень гетерогенны по гену *rpsC*, кодирующему белок S3 малой субъединицы 30S рибосомы (почти вся популяция пневмококка консервативна по гену *rpsC*) (Таблица 13, Том 2 – Приложения. Приложение 4 – Рисунки 29, 30). При этом дивергировали следующие клады изолятов:

- CC180<sup>серотип 3</sup>, CC505<sup>серотип 3</sup>, CC376<sup>серотип 6A</sup> (SC5\_MLST);



- C289<sup>серотип 5</sup> (SC5\_MLST);
- CC315<sup>серотип 6E</sup>, CC1262 (SC5\_MLST);
- CC230<sup>19F,19A,7B,24F</sup> (SC5\_MLST) , CC113<sup>серотип 18BC</sup>, CC123<sup>серотип 17F</sup> (SC6\_MLST – изоляты из России);
- CC239<sup>серотип 9V</sup> (SC9\_MLST - два изолята из России).

Возможно, данный факт отражает распространение генетических линий группы В2 на фоне вакцинации ПКВ7 (прежде всего генетических линий, ассоциированных с серотипами 3, 6А, 6Е).

Белок S3 является одним из универсальных рибосомных белков (которые сохранили свою структуру и выполняемые функции и присутствуют у всех микроорганизмов), доступен на поверхности рибосомы и локализован в области головки 30S-субъединицы. S3 белок, вместе с, по крайней мере, шестью белками малой 30S-субъединицы, осуществляет контроль перемещения матричной РНК по рибосоме. Показано, что боковые цепи положительно заряженных аминокислотных остатков белков S3, S4 и S5 участвуют в формировании так называемого входа для мРНК в 30S-субъединицу.

Также наблюдалась тенденция в дивергенции кладов генетических линий по гену *comX\_1* (ген регулятора транскрипции генных кластеров, кодирующих систему компетентности): CC344+CC448+CC242 (SC10\_MLST) и CC236/CC271/CC320 (SC2\_MLST) (Том 2 - Приложения. Приложение 4 – Рисунки 37, 38).

Также варианты следующих генов были уникальны для некоторых групп SC\_MLST:

- *addB/rexB* (ген АТФ-зависимой экзонуклеазы, субъединицы В); клады филогенетического дерева на основании сравнения последовательностей гена *addB* соответствовали клональным группам (Том 2 – Приложения. Приложение 4, Рисунки 31, 32).

- *pbp1A* (ген пенициллин-связывающего белка, участвующего в синтезе пептидогликана); дивергировали следующие клады филогенетического дерева: CC236/CC271/CC320 (SC2\_MLST); CC81/CC156 (SC4\_MLST+SC8\_MLST); CC90 (SC7\_MLST); CC1094 (SC8\_MLST); CC63 (SC9\_MLST); CC230 (SC5\_MLST); CC185 (SC6\_MLST); CC344/CC448 (SC10\_MLST). Дивергировавшие клады соответствовали резистентным к бета-лактамам антибиотикам генетическим линиям (Том 2 –

Приложения. Приложение 4, Рисунки 33, 34). Известно, что CC156 приобрел детерминанты резистентности к бета-лактамам антибиотикам от CC81. Остальные генетические линии могли приобрести детерминанты резистентности к бета-лактамам антибиотикам независимо, от других видов стрептококков.

- ген *fmt* (ген метионил-тРНК формилтрансферазы, определяющей структуру инициаторной тРНК при трансляции) вносил вклад в формирование группы SC6\_MLST (Таблица 13; Том 2 – Приложения. Приложение 2, Рисунки 35, 36);

- ген *SPN23F21780* (кодирующий  $\alpha$ -1,2-маннозидазу, которая является фактором вирулентности и участвует в расщеплении N-, и возможно, O-гликанов хозяина) вносил вклад в кластеризацию большинства генетических линий пневмококка, дивергентные клады формировали генетические линии CC81 (SC4\_MLST), CC236/CC271/CC320 (SC2\_MLST) и CC90 (SC7\_MLST).

Ген *galE\_1*, кодирующий галактозо-1-фосфат уридинтрансферазу, являющуюся одним из ключевых ферментов метаболизма галактозы (превращает галактозу в глюкозу), отсутствовал почти у половины инвазивных пневмококков анализируемой выборки. Известно, что углеводный метаболизм играет ключевую роль в экспрессии факторов вирулентности у пневмококка (в частности, поверхностных гликозидаз). При этом уровень экспрессии гликозидаз повышается при недостатке питательных веществ. Транскрипционные регуляторы семейства AraC могут регулировать активность галактозидазы у *S. pneumoniae*, которая является фактором вирулентности и играет роль в патогенезе. Интересно, что изоляты SC7\_MLST, несущие оба гена *galE\_1* и *araC*, редко являлись инвазивными (Таблица 13).

Таким образом, деление популяции на группы SC\_MLST обусловлено метаболическими особенностями, в том числе ассоциирующимися с вирулентностью. Некоторые генетически гетерогенные группы SC\_MLST характеризуются общими метаболическими особенностями. Филогенетические деревья по последовательностям большинства из перечисленных генов приведены в Приложении (Том 2 - Приложения. Приложение 4).

Таблица 13 – Ассоциации аллелей генов с группами SC\_MLST (XGBoost)

Ген	Аллель	Отн. значимость	Норм. значимость	Процент ошибки	Продукт	SC_MLST
<i>SPN23F14460</i>	17	379,8818	1,0	0,0458	АТФ-зависимая РНК-хеликаза YfmL	SC8_MLST
<i>comX_1</i>	8	356,9925	0,9397	0,0430	Регулятор транскрипции генов естественной компетентности	SC10_MLST
<i>trpE</i>	1	298,5247	0,7858	0,0360	Компонент I антранилат-синтазы	SC4_MLST
<i>SPN23F04400</i>	7	227,1056	0,5978	0,0274	Гипотетический белок	SC8_MLST
<i>galE_1</i>	3	189,3580	0,4985	0,0228	Галактозо-1-фосфат уридинтрансфераза	SC7_MLST
<i>ciaH</i>	2	174,0420	0,4581	0,0210	Сенсорная гистидин-киназа	SC5_MLST (кроме CC66, CC230, CC199, CC191)
<i>SPN23F13980</i>	11	173,4727	0,4566	0,0209	Регулятор транскрипции AgaC	SC7_MLST
<i>relA</i>	5	168,2187	0,4428	0,0203	ГТФ-пирофосфокиназа	SC3_MLST
<i>addB / rexB</i>	1	153,7222	0,4047	0,0185	АТФ-зависимая экзонуклеаза, субъединица B	SC4_MLST
<i>SPN23F16040</i>	22	139,9578	0,3684	0,0169	Лигаза семейства Mur	SC11_MLST
<i>SPN23F21780</i>	1	138,5466	0,3647	0,0167	Альфа-1,2-маннозидаза	SC4_MLST
<i>aroK</i>	7	136,0699	0,3582	0,0164	Шикимат-киназа I	SC11_MLST
<i>rpsC</i>	1	134,6847	0,3545	0,0162	Белок S3 субъединицы 30S рибосомы	397 изолятов, кроме SC5_MLST
<i>SPN23F07330</i>	8	128,0101	0,3370	0,0154	Белок семейства DJ-1/PfpI (гидролаза, протеаза)	SC3_MLST
<i>SPN23F01160</i>	new102	112,4263	0,2960	0,0136	Гипотетический белок	SC12_MLST
<i>fnt</i>	5	107,2326	0,2823	0,0129	Метионил-тРНК формилтрансфераза	SC6_MLST
<i>SPN23F22520</i>	5	97,0196	0,2554	0,0117	Белок, участвующий в транспорте кобальта	SC5_MLST
<i>pbp1A</i>	3	96,7681	0,2547	0,0117	Пенициллин-связывающий белок Pbp1A	SC2_MLST
<i>SPN23F01390</i>	3	91,6953	0,2414	0,0111	Гликопротеаза	SC2_MLST

«Отн. значимость» – относительная значимость, «Норм. значимость» - нормализованная значимость

### 3.1.3.3. Гены, участвующие в дифференциации на серотипы

Формирование серотипов детерминировалось большей частью генов полного генома (идентифицировано > 1000 генов с высоким значением Z-score по результатам поиска взаимодействующих переменных с помощью RandomForest), особенно с генами, кодирующими компоненты АТФ-синтазного комплекса, генами метаболизма углеводов, пуринов и пиримидинов, синтеза метионина, синтеза пептидогликана и клеточной стенки. Расщепление популяции на серотипы прежде всего зависело от генов *cps*-локуса. Аллели наиболее значимых генов общего метаболизма приведены в Таблицах 14 и 15.

Наличие или отсутствие генов *cps*-локуса зависело от типа и числа синтезируемых остатков моносахаридов, входящих в структуру тейхоевых кислот. Например, гены *rmlA*, *rmlB*, *rmlC*, *rmlD*, кодирующие белки, связанные с синтезом и процессингом dTDP-рамнозы, отсутствовали у 183 изолятов выборки, представленных серотипами: 3, 4, 5, 8, 9A, 9N, 9V, 10A, 10B, 11A, 12F, 13, 14, 15A, 15BC, 20B, 21, 23B, 25F, 34, 35F. Гены, необходимые для синтеза рамнозы, присутствовали в *cps*-локусах серотипов: 1, 6A, 6B, 6C, 6E, 7F, 16F, 17F, 18BC, 19A, 19F, 21, 22F, 23A, 23F.

Вариант *murD-7* гена *murD*, продукт которого (УДФ-N-ацетилмурамоил-L-аланил-D-глутамат синтетаза) осуществляет один из этапов синтеза пептидогликана клеточной стенки (катализ присоединения глутамата предшественнику УДФ-N-ацетилмурамоил-L-аланину), был идентичен у представителей серотипа 1 различных генетических линий. Интересно, что серотип 1 обладает уникальным углеводным составом капсулы [138]. Для остальных изолятов кластеризация по последовательностям гена *murD* скорее отражала распределение по близким клональным группам, чем по серотипам (Том 2 – Приложения. Приложение 4, Рисунки 43, 44).

Вариант *gnd-39* гена *gnd* (кодирующего 6-фосфоглюконат дегидрогеназу) был идентичен у представителей серотипа 6E различных генетических линий (Таблица 15). Кластеризация на основании сравнения последовательностей гена *gnd* подтвердила данную картину (Том 2 – Приложения. Приложение 4, Рисунки 45, 46). Дегидрогиназы пентозофосфатного пути являются основным источником НАДФН.

Вариант *agaD-8* гена *agaD* (кодирующего IID-компонент N-ацетил-галактозамин-специфичной фосфотрансферазной системы) ассоциировался с изолятами серотипа 3.

Вариант *agaD-1* присутствовал у 194 из 495 изолятов, ассоциирующихся с наиболее распространенными серотипами 6А, 9А, 11А, 15ВС, 19А, 19F, 23F, 35В. Кластеризация на основании сравнения последовательностей гена *agaD* подтвердила данную картину: распространенный монофилетический клад был представлен генетическими линиями, ассоциированными с указанными выше серотипами (Том 2 – Приложения. Приложение 4, Рисунки 47, 48). Генетические линии, представленные дивергентными кладами, чаще ассоциировались с инвазивностью изолятов (особенно с источником выделения – кровь) (Том 2 – Приложения. Приложение 4, Рисунок 48). N-ацетил-галактозамин присутствует в структуре капсулы только у одного из серотипов клада *agaD-1* (серотип 35В). Роль данного углевода не ограничена структурой капсулы – он может метаболизироваться в качестве эргетического субстрата. Также известно, что углевод-специфичная фосфотрансферазная система может выполнять регуляторную функцию. Избыточное поступление какого-либо углевода под действием фосфотрансферазной системы подавляет активный транспорт сахаров другими переносчиками (предполагают, что это происходит за счет изменения содержания цАМФ). Таким образом, особенности N-ацетилгалактозамин-специфичной фосфотрансферазной системы критичны для углеводного метаболизма, возможно, также детерминируют углеводный состав полисахаридной капсулы пневмококков и потенциал вирулентности. Можно предположить, что менее энергетически-эффективные типы данной фосфотрансферазной системы могут провоцировать экспрессию факторов вирулентности (в монофилетическом распространенном клада мало изолятов, выделенных из крови, и они ассоциированы только с одной генетической линией).

Вариант *atpG-1* гена *atpG* присутствовал у изолятов серотипа 23F различных генетических линий. Кластеризация на основании сравнения последовательностей гена *atpG* также позволила идентифицировать клад, включающий большинство представителей 23F различных генетических линий (Том 2 – Приложения. Приложение 4, Рисунки 49, 50). Популяция пневмококка делилась по гену *atpG* на две группы – гетерогенную и гомогенную (гомогенная группа была представлена серотипами 1, 3, 6А, 7F, 9А, 9N, 11А, 15ВС, 19А, 19F, 23F распространенных генетических линий) (Том 2 – Приложения. Приложение 4, Рисунки 49, 50). AtpG является компонентом АТФ-синтазного комплекса, выполняющего функцию накопления энергии за счет мембранного потенциала, сформированного в цепи переноса электронов. Возможно,

вариант *AtpG*, ассоциирующийся с гомогенной группой и большим разнообразием серотипов, позволяет пневмококкам преодолеть *fitness-cost*, связанный с возможностью синтеза различных типов капсулы. Интересно, что варианты *agaD-1* и *atpG-1* ассоциированы с похожим гомогенным кладом, представленным распространенными генетическими линиями и разнообразными серотипами. Таким образом, метаболический тип *agaD-1* и *atpG-1* является оптимальным для роста и адаптации большинства пневмококков.

Таким образом, принадлежность изолята к серотипу обусловлена взаимодействием большого числа компонентов глобальных метаболических путей. Серотипы ассоциируются с генетическими линиями, характеризующимися соответствующим метаболическим типом. АТФ-синтазный комплекс и компоненты фосфотрансферазной системы детерминируют формирование типов капсулы.

**Таблица 14 – Гены, вносящие вклад в деление на серотипы (Random Forest)**

Переменная	Z-score	Продукт	Характеристики
<b>Serotype</b>	14,75079701	Серотип	
<i>wzh</i>	11,48184983	Тирозиновая фосфатаза Wzh	<i>cps</i> -локус
<i>wze</i>	11,37441629	Тирозиновая протеинкиназа Wze	<i>cps</i> -локус
<i>wzd</i>	10,91551428	Олигосахарид-зависимая полимераза Wzd	<i>cps</i> -локус
<i>rmlC</i>	10,8723484	dTDP-4-кето-6-дезоксиглюкозо-3,5-эпимераза RmlC	<i>cps</i> -локус
<i>wchA</i>	10,45498815	Ундекаprenилфосфат глюкозофосфотрансфераза WchA (инициирующая углеводтрансфераза)	<i>cps</i> -локус
<i>rmlD</i>	10,20658341	dTDP-4-дегидрорамнозо редуктаза RmlD	<i>cps</i> -локус
<i>rmlA</i>	9,99042969	Глюкозо-1-фосфат тимидил трансфераза RmlA	<i>cps</i> -локус
<i>wchW</i>	8,7869992	Гликозилтрансфераза WchW	<i>cps</i> -локус
<i>wchY</i>	8,57714063	Глицерол-2-фосфат дегидрогиназа WchY	<i>cps</i> -локус
<i>wciA</i>	8,5297719	Фосфотрансфераза	<i>cps</i> -локус
<i>dexB</i>	8,46182266	Глюкан-1,6-альфа-глюкозидаза	
<i>wchV</i>	8,29857963	Гликозилтрансфераза WchA	<i>cps</i> -локус
<i>wchZ</i>	8,01618995	Нуклеотидил трансфераза WchZ	<i>cps</i> -локус
<i>wchX</i>	7,60274542	Глицерол фосфотрансфераза WchX	<i>cps</i> -локус
<i>SPN23F08680</i>	7,55636268	Консервативный гипотетический белок	
<i>SPN23F01130</i>	7,33142835	Гликозил трансфераза	ФВ
<i>SPN23F19780</i>	7,32116549	АВС-транспортер, АТФ-связывающий домен	

Таблица 14 – продолжение таблицы

Переменная	Z-score	Продукт	Характеристики
<i>SPN23F14890</i>	7,2682032	Цистатионин бета-лиаза	Метаб-м цистеина и метионина
<i>wzg</i>	7,24967901	Интегральный мембранный регуляторный белок Wzg	<i>cps</i> -локус
<i>wzx</i>	6,9273775	Флиппаза Wzx	<i>cps</i> -локус
<i>phtE</i>	6,92684792	Белок пневмококковой гистидиновой триады E (Bvh-3)	ФВ
<i>wzy</i>	6,91951622	Олигосахарид-зависимая полимераза Wzy, присоединяющая повторы олигосахаридов	<i>cps</i> -локус
<i>cdd</i>	6,81856641	Цитидиндезаминаза	Метаболизм пиримидина
<i>SPN23F06940</i>	6,7160512	Пептидил-пролил-цис/транс-изомераза типа циклофилина	
<i>SPN23F04920</i>	6,66029392	Гипотетический белок	
<i>cbpJ</i>	6,6000807	Холин-связывающий белок J	
<i>SPN23F19450</i>	6,54463247	Регуляторный белок семейства MarR	
<i>SPN23F21150</i>	6,54402093	ДНК-связывающий белок	
<i>SPN23F20850</i>	6,38745428	Регуляторный белок семейства MarR	
<i>SPN23F18540</i>	6,30027976	Гипотетическая углеводтрансфераза	
<i>SPN23F06280</i>	6,29322465	Гипотетический мембранный белок	
<i>SPN23F11180</i>	6,258688	Белок, участвующий в транспорте ионов	
<i>SPN23F14860</i>	6,2450004	Гипотетический АТФ-связывающий белок	
<i>clpL</i>	6,22027604	АТФ-зависимая протеаза, субъединица ClpL	ФВ
<i>rmlB</i>	6,20690319	dTDP-глюкозо-4,6-дегидратаза RmlB	<i>cps</i> -локус
<i>luxS</i>	6,19411509	S-рибозилгомоцистеиназа	Мет-м цистеина и метионина, ФВ
<i>pcpA</i>	6,12520546	Поверхностный холин-связывающий белок PcpA	ФВ

Таблица 15 – Ассоциации аллелей генов глобального метаболизма с серотипами (XGBoost)

Ген	Аллель	Отн. значимость	Норм. значимость	Ошибка предсказания	Продукт
<i>murD</i>	7	555,4982	1,0	0,0379	UDP-N-ацетилмурамоил-аланин-D-глутамат лигаза
<i>gnd</i>	39	481,3342	0,8665	0,0329	6-фосфоглюконат дегидрогиназа
<i>atpG</i>	1	455,5974	0,8202	0,0311	АТФ-синтаза, гамма цепь
<i>agaD</i>	8	443,6434	0,7986	0,0303	N-ацетилгалактозамин-специфичная фосфотрансферазная система (PTS), компонент IID
<i>mraW</i>	59	421,1080	0,7581	0,0288	S-аденозил-метилтрансфераза MraW

### 3.1.3.4. Гены, ассоциированные с делением популяции *S. pneumoniae* на штаммы, чувствительные и резистентные к бета-лактамам антибиотикам

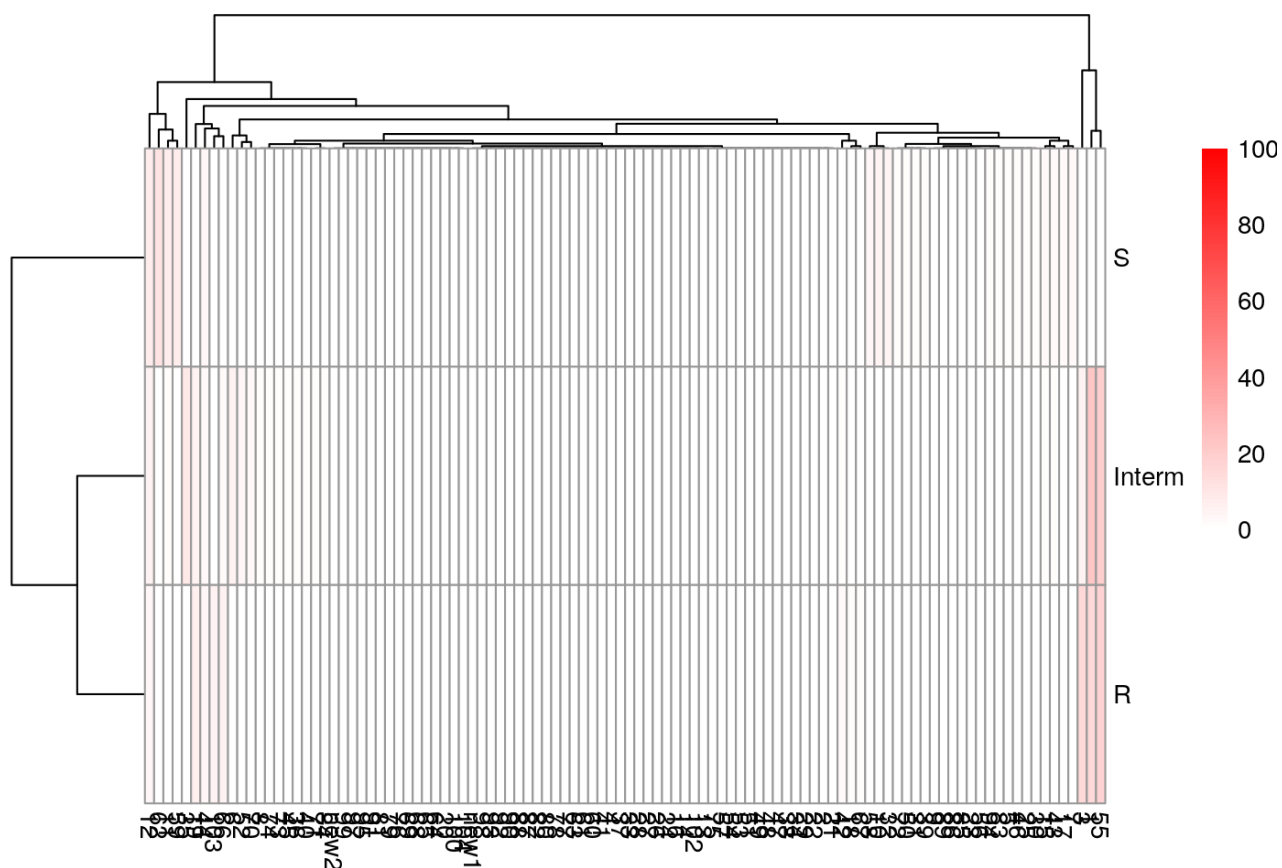
Помимо генов, кодирующих известные мишени бета-лактамов антибиотиков, с помощью алгоритма XGBoost были идентифицированы гены некоторых консервативных гипотетических белков, а также ген *rpoB* (ген бета-субъединицы РНК-полимеразы), которые вносили вклад в деление популяции пневмококков на резистентные и чувствительные штаммы (Таблица 16, Рисунок 27). Ген *rpoB* кодирует бета-субъединицу РНК-полимеразы, которая катализирует транскрипцию ДНК в РНК. Может опосредовать дополнительный механизм регуляции экспрессии генов, поскольку мутации в генах РНК-полимеразы могут приводить к неравномерной транскрипции некоторых оперонов [332, 333].

**Таблица 16 – Ассоциации аллелей генов с резистентными к пенициллину штаммами (XGBoost)**

Ген	Аллель	Отн. значимость	Норм. значимость	Процент ошибки	Продукт
<i>SPN23F09650</i>	20	120,9954	0,4538	0,0373	консервативный гипотетический белок
<i>SPN23F12010</i>	2_15	63,0478	0,2365	0,0195	гипотетический белок
<i>SPN23F03050</i>	4	52,1956	0,1958	0,0161	ДНК-связывающий белок YofE
<i>prfA</i>	25	36,8694	0,1383	0,0114	терминатор трансляции
<i>tmk</i>	23	31,9863	0,1200	0,0099	тимидилаткиназа
<i>rpoB</i>	1	29,7342	0,1115	0,0092	ДНК-зависимая РНК-полимераза, бета-цепь
<i>SPN23F09110</i>	10	27,9274	0,1047	0,0086	гипотетический мембранный белок
<i>SPN23F08690</i>	1	26,0036	0,0975	0,0080	консервативный гипотетический белок

«Отн. значимость» – относительная значимость, «Норм. значимость» - нормализованная значимость





**Рисунок 27 – Гистограмма ассоциации между вариантами гена *groB* и наличием резистентности к пенициллину**

По горизонтальной оси отложены варианты исследуемого гена, по вертикальной оси – группы изолятов. Цвет ячейки отражает количество изолятов.

### **3.1.3.5. Взаимодействие генов, участвующих в делении популяции на группы A/B1/B2, SC\_MLST и серотипы, инвазивные, резистентные и чувствительные к антибиотикам штаммы**

Для идентификации метаболических типов, характеризующих различные генетические линии, была проведена сравнительная оценка топологии и кластеризации филогенетических деревьев, построенных по последовательностям ключевых идентифицированных генов, с помощью программного пакета TreeCmp (Раздел 2.14. «Оценка топологии филогенетических деревьев»). Соответствие кластеризации деревьев оценивалось с помощью параметров «Matching split distance» (соответствие дистанций расщепления) и «MatchingCluster» (соответствие кластеров). Результаты оценки топологии наиболее похожих филогенетических деревьев приведены в Таблице 17.

Таблица 17 – Наиболее похожие топологии филогенетических деревьев

Ген	Ген	Продукт	Соответствие дистанций расщепления	MatchingSplit toYuleAvg	MatchingSplit toUnifAvg	Соответствие кластеров
<i>aroE</i>	<i>gnd</i>	6-фосфоглюконат дегидрогеназа	1.0133	0.4170	1.3119	0.5294
<i>aroE</i>	<i>strH</i>	экзо-β-D-N-ацетилглюкозаминидаза	1.0018	0.4123	0.9373	0.3782
<i>aroE</i>	<i>dexB</i>	Глюкан-1,6-альфа-глюкозидаза	0.9878	0.4065	1.1262	0.4545
<i>aroE</i>	<i>murD</i>	УДФ-N-ацетилмурамоил-аланин-D-глутамат лигаза	0.8703	0.3605	1.4652	0.5949
<i>aroE</i>	<i>pbp1A</i>	пенициллин-связывающий белок 1A	1.0533	0.4347	1.1383	0.4654
<i>aroE</i>	<i>relA</i>	ГТФ-пирофосфокиназа	0.9300	0.3827	1.4159	0.5714
<i>aroE</i>	<i>SPN23F21780</i>	гипотетическая альфа-1,2-маннозидаза	1.0455	0.4305	1.2468	0.5126
<i>aroE</i>	<i>trpE</i>	антранилатсинтаза, компонент I	0.8699	0.3603	1.3206	0.5362
<i>aroE</i>	<i>wzg</i>	Мембранный регуляторный белок Wzg	1.0502	0.4363	1.2796	0.5289
<i>aroE</i>	<i>wzh</i>	Тирозиновая фосфатаза Wzh	1.0975	0.4562	1.3692	0.5648
<i>gdh</i>	<i>addB</i>	АТФ-зависимая экзонуклеаза, субъед. В	1.1688	0.4810	1.4038	0.5665
<i>gdhMLST</i>	<i>addB</i>	АТФ-зависимая экзонуклеаза, субъед. В	1.2127	0.4991	1.1517	0.4648
<i>gnd</i>	<i>addB</i>	АТФ-зависимая экзонуклеаза, субъединица В	0.9733	0.4006	1.4879	0.6004
<i>gnd</i>	<i>dexB</i>	Глюкан-1,6-альфа-глюкозидаза	1.0275	0.4228	1.2682	0.5118
<i>gnd</i>	<i>murD</i>	УДФ-N-ацетилмурамоил-аланин-D-глутамат лигаза	0.8697	0.3603	1.1012	0.4471
<i>gnd</i>	<i>SPN23F21780</i>	гипотетическая альфа-1,2-маннозидаза	1.1164	0.4597	1.0675	0.4389
<i>gnd</i>	<i>wze</i>	Тирозиновая протеинкиназа Wze	0.9842	0.4092	1.4295	0.5919
<i>gnd</i>	<i>wzg</i>	Мембранный регуляторный белок Wzg	0.9879	0.4104	0.9456	0.3909
<i>mraW</i>	<i>addB</i>	АТФ-зависимая экзонуклеаза, субъед. В	1.0301	0.4251	1.3927	0.5694
<i>rpoB</i>	<i>SPN23F01160</i>	гипотетический белок	1.1237	0.4625	1.1514	0.4646
<i>rpoB</i>	<i>SPN23F16040</i>	Лигаза семейства Mur	1.1390	0.4688	1.3808	0.5572
<i>rpoB</i>	<i>trpE</i>	антранилатсинтаза, компонент I	0.8251	0.3418	1.1315	0.4594
<i>rpoB</i>	<i>wze</i>	Тирозиновая протеинкиназа Wze	1.1023	0.4583	1.1940	0.4944
<i>rpoB</i>	<i>wzh</i>	Тирозиновая фосфатаза Wzh	1.0729	0.4460	1.2660	0.5222
<i>strH</i>	<i>dexB</i>	Глюкан-1,6-альфа-глюкозидаза	1.1865	0.4883	1.3904	0.5611

Таблица 17 – продолжение таблицы

Ген	Ген	Продукт	Соответствие дистанций расщепления	MatchingSplit _toYuleAvg	MatchingSplit _toUnifAvg	Соответствие кластеров
<i>xpt</i>	<i>xptMLST</i>	Ксантин фосфорибозилтрансфера за	0.9064	0.3730	1.2486	0.5039
<i>addB</i>	<i>murD</i>	УДФ-N- ацетилмурамоил-аланин- D-глутамат лигаза	0.8184	0.3390	0.9192	0.3732
<i>addB</i>	<i>phtE</i>	белок E пневмококковой гистидиновой триады (Bvh-3)	1.0325	0.4289	1.0412	0.4304
<i>addB</i>	<i>relA</i>	ГТФ-пирофосфокиназа	1.0278	0.4230	1.3283	0.5360
<i>addB</i>	<i>SPN23F21780</i>	гипотетическая альфа- 1,2-маннозидаза	0.9494	0.3909	1.4545	0.5980
<i>dexB</i>	<i>murD</i>	УДФ-N- ацетилмурамоил-аланин- D-глутамат лигаза	0.9541	0.3952	1.4317	0.5813
<i>dexB</i>	<i>pbp1A</i>	пенициллин- связывающий белок 1A	1.1719	0.4836	1.3639	0.5576
<i>dexB</i>	<i>SPN23F21780</i>	гипотетическая альфа- 1,2-маннозидаза	1.0220	0.4208	1.1831	0.4864
<i>dexB</i>	<i>trpE</i>	антранилатсинтаза, компонент I	1.0846	0.4492	1.4410	0.5851
<i>dexB</i>	<i>wzg</i>	Мембранный регуляторный белок Wzg	0.9174	0.3811	1.2009	0.4964
<i>murD</i>	<i>phtE</i>	белок E пневмококковой гистидиновой триады (Bvh-3)	0.9616	0.3997	0.9055	0.3735
<i>murD</i>	<i>relA</i>	ГТФ-пирофосфокиназа	1.0272	0.4255	1.0336	0.4197
<i>murD</i>	<i>SPN23F21780</i>	гипотетическая альфа- 1,2-маннозидаза	0.9396	0.3869	1.1691	0.4807
<i>murD</i>	<i>wzg</i>	Мембранный регуляторный белок Wzg	0.8874	0.3689	1.2719	0.5247
<i>pbp1A</i>	<i>trpE</i>	антранилатсинтаза, компонент I	1.1707	0.4820	1.4535	0.5977
<i>phtE</i>	<i>relA</i>	ГТФ-пирофосфокиназа	1.0826	0.4497	1.1463	0.4738
<i>phtE</i>	<i>SPN23F21780</i>	гипотетическая альфа- 1,2-маннозидаза	1.0541	0.4392	1.3361	0.5545
<i>relA</i>	<i>SPN23F21780</i>	гипотетическая альфа- 1,2-маннозидаза	1.1630	0.4789	1.2272	0.5046
<i>SPN23F21780</i>	<i>wzg</i>	Мембранный регуляторный белок Wzg	1.0928	0.4553	1.0356	0.4298
<i>trpE</i>	<i>wzg</i>	Мембранный регуляторный белок Wzg	1.1137	0.4630	1.4504	0.5983
<i>trpE</i>	<i>wzh</i>	Тирозиновая фосфатаза Wzh	1.1044	0.4591	1.1624	0.4795
<i>wze</i>	<i>wzg</i>	Мембранный регуляторный белок Wzg	0.8554	0.3556	1.2614	0.5203
<i>wze</i>	<i>wzh</i>	Тирозиновая фосфатаза Wzh	0.8833	0.3680	1.2742	0.5233
<i>wzg</i>	<i>wzh</i>	Тирозиновая фосфатаза Wzh	0.8420	0.3508	1.2134	0.5036

Расщепление популяции пневмококка по компонентам углеводного метаболизма (*gnd*, *dexB*) коррелирует с расщеплением популяции по генам синтеза пептидогликана и регуляции экспрессии полисахаридной капсулы (*wzg-wzh-wze*), а также с расщеплением по вариантам гена поверхностной гликозидазы StrH (фактор вирулентности).

Расщепление по гену ДНК-зависимой полимеразы (В-субъединица, *proB*) также коррелирует с расщеплением по регуляторным генам экспрессии полисахаридной капсулы.

Расщепление популяции по гену *aroE* (синтез ароматических соединений, в том числе ароматических аминокислот) коррелирует с расщеплением популяции по генам углеводного обмена, гену регулятора ответа *relA* (ГТФ-пирофосфокиназа), генов синтеза пептидогликана и клеточного деления (*murD*, *pbp1A1*), регуляторным генам экспрессии полисахаридной капсулы (*wzg-wzh-wze*), а также по гену поверхностной гликозидазы StrH.

Расщепление популяции по гену *addB/rexB* (субъединица В АТФ-зависимой хеликазы/дезоксирибонуклеазы) ассоциируется с расщеплением по гену синтеза пептидогликана *murD*, гену регулятора ответа *relA*, а также по гену, кодирующему белок гистидиновой триады PhtE (является фактором вирулентности, может способствовать уклонению пневмококка от иммунной системы человека). Таким образом, особенности метаболизма углеводов и ароматических аминокислот (пневмококки способны синтезировать фенилаланин, тирозин и триптофан) детерминируют регуляцию клеточных процессов пневмококка и фенотипические особенности.

Реконструкция взаимодействующих метаболических путей – мишеней была выполнена с помощью базы данных STRING (Рисунок 28).

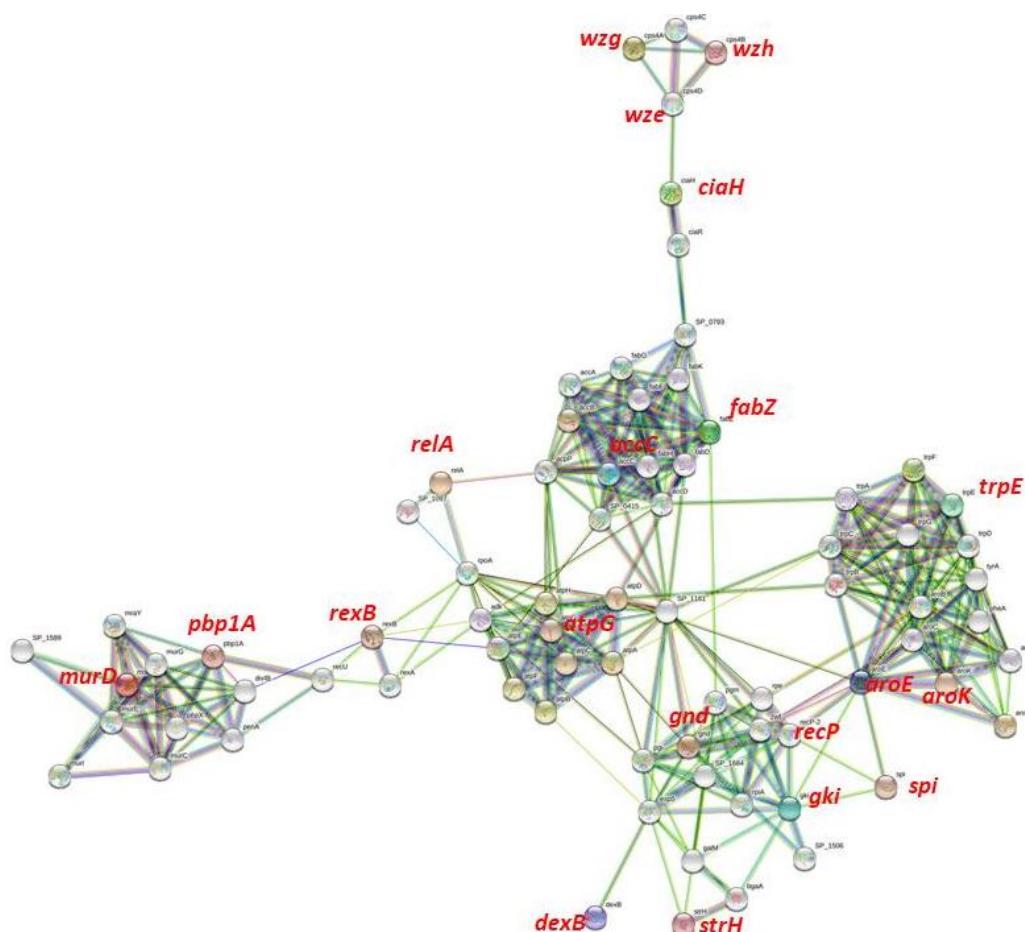
Субстратом для глюкокан-1,6- $\alpha$ -глюкозидазы (DexB) могут быть короткие изомальтолигосахариды. Глюкокиназа (Gki) регулирует глюкозо-специфичную фосфотрансферазную систему и экспрессию поверхностных гликозидаз (StrH). Сигнальная пептидаза I (SpI) взаимодействует с шикимат-дегидрогеназой (AroE), участвующей в биосинтезе ароматических аминокислот. Компоненты АТФ-синтазной системы (AtpG), а также ферменты пентозофосфатного пути (Gnd) являются источником АТФ и НАДФН. АТФ и НАДФН также необходимы в ходе биосинтеза жирных кислот (AccC, FabZ), поэтому варианты *atpG* и, *gnd*, а также компоненты фосфотрансферазной

системы, специфичной к другим источникам углеводов (*agaD*), могут способствовать преодолению fitness-cost, связанного с фенотипической пластичностью генетической линии.

Сенсорная гистидиновая киназа *CiaH* регулирует экспрессию полисахридной капсулы, в зависимости от состояния синтеза жирных кислот.

ГТФ-пирофосфокиназа (*RelA*) регулирует различные клеточные процессы, в том числе координацию синтеза пептидогликана и клеточного деления.

Таким образом, особенности метаболизма пневмококка детерминируют его значительный вирулентный потенциал. Серотип-специфичная вакцинация, приводящая к исключению из циркуляции распространенных генетических линий пневмококка, обладающих большей способностью к адаптации за счет энерго-обеспечивающих систем (АТФ, НАДФН, фосфотрансферазные системы), может приводить к распространению генетических линий, характеризующихся большей вирулентностью.



**Рисунок 28 – Взаимодействие метаболических путей, детерминирующих фенотипы *Streptococcus pneumoniae***

Аннотация метаболических путей выполнена с помощью базы данных STRING.

Красным цветом маркированы гены, расщепляющие популяцию пневмококка на группы.

### 3.1.4. Оценка вклада рекомбинаций в структуру популяции

Рекомбинационные события оказывают значительный вклад в структуру пневмококковой популяции. Рекомбинации были оценены по двум параметрам:

- Относительная скорость рекомбинаций ( $\mu_{r/m}$ ) – отношение числа SNPs, обусловленных событием рекомбинации, к числу SNPs, обусловленных событиями точковых мутаций. Данный параметр позволяет оценить общее разнообразие, накопленное в генетической линии.

- Частота рекомбинаций ( $Av_{bloks}$ ) – среднее число рекомбинационных событий на изолят без учета привнесенной вариации.

При большом числе рекомбинаций у близкородственных штаммов может наблюдаться высокое значение  $Av_{bloks}$ , но низкое  $\mu_{r/m}$ . Высокая частота присутствия предковых рекомбинаций, имеющих клональное происхождение, также может ассоциироваться с высоким значением  $Av_{bloks}$ .

Среднее значение для каждого из этих параметров было подсчитано для распространенных генетических линий. В случае, если генетическая линия включала представителей нескольких серотипов, статистика по каждому из них была посчитана отдельно (Таблица 18).

События рекомбинаций по-разному затрагивали различные генетические линии. Генетические линии отличались размером рекомбинационных блоков. Значение скорости рекомбинаций  $\mu_{r/m}$  варьировало в диапазоне 3,16 – 38,85, со средним значением 14,04 (Таблица 18).

Наиболее высокой рекомбинационной активностью характеризовались генетические линии: СС90 (серотип 6Е), СС217 (серотип 1), СС236/СС271/СС320 (доминантный серотип 19F), СС81 (доминантный серотип 23F), СС289 (серотип 5), СС1094 (серотип 6АЕ), СС156 (серотип 35В), СС304/СС306 (серотип 1), СС124 (серотип 14), СС199 (серотип 19А). Первые 5 перечисленных выше генетических линий часто мутировали не только вследствие рекомбинаций, но и отличались более высокой частотой событий мутаций вне рекомбинаций ( $Av_{mut}$ ). Генетические линии СС236/СС271/СС320 и СС90, СС289, СС217, СС179, СС306 характеризовались наличием большого числа предковых клональных рекомбинаций, которые возможно ассоциировались с успешным распространением данных клонов.

Таким образом, дивергировавшие генетические линии ассоциируются с частотой и особенностями рекомбинационных событий. Необходимо отметить, что резистентные к бета-лактамам антибиотикам изоляты ассоциировались с серотипами, отличающимися высокой эффективностью рекомбинаций (23F, 19F, 6E, 6A, 14, 19A).

**Таблица 18 – Статистика рекомбинационных событий, идентифицированных для каждой генетической линии**

SC_MLST	Серотип	N	$\mu_{re}$	$Av_{blocs}$	$Av_{mut}$	$\mu_{r/m}$	Размер рекомбинационного блока (п.н.)		
							Медиана	25й-75й%%	5-95%
СС 81	<b>23F, 19F, 6AB, 19A, 3</b>	54	0,12	16,72	51,92	17,90	256	159 – 947	5 – 2076
СС 124	<b>14</b>	9	0,26	4,22	1,11	18,84	365,5	59,75 – 6691,5	15 – 12079
СС 344	<b>NT</b>	8	0,27	1,37	5	24,7	502	293,5 – 1837,5	167 – 2944
СС 448	<b>NT</b>	8	6,97	0,25	4,37	6,97		два блока	
СС 236 / 271 / 320	<b>19F</b>	37	0,04	18,45	467,56	3,23	2512	1114,5 – 6027	10 – 13386
СС 62	<b>11A</b>	25	0,12	1,48	11,76	9,79	391	119 – 1123	36 – 1263
СС 156 (серотип 35B)	<b>35B</b>	8	0,37	3,75	10	31,32	3910,5	945 – 6271,5	43 – 11303
СС 156 (серотип 9A)	<b>9A</b>	10	0,13	0,6	4,6	8,17	5470	1252,25 – 8141,25	240 – 15872
СС 36	<b>23F</b>	9	0,13	0,78	5,78	3,83	650	182 – 1118,5	102 – 1132
СС 177 / 179	<b>19F</b>	6	0,25	2,5	9,83	12,81	3050	1711 – 5122	289 – 9302
СС 1094 (серотип 6A)	<b>6A/6E</b>	14	0,41	2,86	7	38,85	2691	305,25 – 6118,75	13 – 13512
СС 217	<b>1</b>	11	0,53	36,09	67,45	19,61	195	106 – 732	16 – 1592
СС 180 / 505	<b>3</b>	31	0,20	2,03	10,32	8,83	497	111 – 1005	5 – 1093
СС 289	<b>5</b>	12	0,32	34	105,83	15,05	262	141,75 – 2643	12 – 6152
СС 315	<b>6E</b>	13	0,11	0,38	3,61	9,17	8934	4421 – 9485	1830 – 15924
СС 191	<b>7F</b>	18	0,14	0,33	2,33	3,16	370	337,75 – 370	327 – 370
СС 66 (ST66/ST517/ST3104)	<b>9N</b>	9	0,67	0,44	6,67	3,67	704,5	691 – 3429,5	691 – 3429,5
СС 230 / 276	<b>19F, 19A, 24F, 7B</b>	7	0,08	3,71	46,71	5,39	5769	1925,75 – 7402,75	109 – 10955
СС 90	<b>6E</b>	43	0,49	40,21	81,16	26,02	413	183 – 1144	5 – 2565
СС 199 (серотип 15BC)	<b>15BC</b>	10	0,07	0,6	8,7	5,49	1618,5	607,7 – 9874,2	89 - 15820
СС 199 (серотип 19A)	<b>19A</b>	12	0,15	1	6,67	16,95	677	302 – 1082	14 – 2222
СС 304 / 306	<b>1</b>	12	0,33	34,17	104,75	19,13	501	281 – 1382,5	9 – 2952

( $\mu_{re}$ ) – среднее число событий рекомбинаций, по отношению к точковым мутациям на изолят;

$Av_{blocs}$  - среднее число блоков рекомбинаций;

( $\mu_{r/m}$ ) - среднее значение рекомбинационных событий, по отношению к точковым мутациям;

$Av_{mut}$  – среднее число событий вне рекомбинаций.

Жирным шрифтом выделены клональные группы и соответствующие им распространенные серотипы; цветом подсвечены значения, соответствующие высокой частоте рекомбинаций.

Согласно данным других авторов, частота рекомбинаций ассоциируется с серотипами, толщиной капсулы и серотип-специфичной продолжительностью носительства [334]. В частности, высокая частота рекомбинаций была отмечена для серотипов 23F, 35B, 6A, низкая частота – для серотипов 1, 16F, 4, 5 [334]. При этом разные генетические линии, ассоциирующиеся с одним и тем же серотипом (например, 6A), отличались высокой частотой рекомбинаций, несмотря на различный генетический фон [334]. Эти данные могут быть объяснены тем, что более длительное носительство может создавать условия для совместной колонизации различными штаммами, при этом в генетической линии, быстрее достигающей критическую концентрацию в носоглотке хозяина, активируется состояние естественной компетентности и она становится реципиентом генетического материала других линий.

В нашем исследовании высокая частота рекомбинаций ассоциировалась с серотипами 6AE, 19F, 23F, 35B, 14, 19A (Таблица 18), а низкая – с серотипами 7F (CC191), 9N (CC66), 15BC (CC199), что частично подтверждало полученные ранее данные. Однако в нашем исследовании генетические линии, ассоциированные с серотипами 1 (CC217; CC304/CC306) и 5 (CC289), характеризовались высокой частотой рекомбинаций. Данные результаты могут быть объяснены генетической гетерогенностью анализируемых штаммов и наличием большого числа клональных рекомбинаций в субкладах, представленных изолятами из разных регионов (Африка, Вьетнам), либо в изолятах родственного сиквенс-типа (ST304 / ST306).

### **3.1.5. Системы рестрикции анализируемых изолятов**

Во всех анализируемых геномах пневмококков были идентифицированы метилтрансферазы CPM I и II типа. Идентифицированные гены имели высокое сходство с уже охарактеризованными последовательностями генов CPM: идентичность анализируемых последовательностей с известными генами в REBASE составила >98,9%. Однако наличие соответствующих генов не означает функциональной активности CPM в анализируемом геноме. Для получения более достоверных сведений об активности CPM в анализируемом геноме. Для получения более достоверных сведений об активности необходимо использовать дополнительные методы секвенирования, позволяющие идентифицировать метилированные основания (Single molecule real time sequencing, SMRT).

Были идентифицированы 54 варианта генов метилтрансфераз CPM I типа, которые ассоциировались с клональными комплексами. Однако CC505 и CC191



(SC5\_MLST), а также CC81 (SC4\_MLST) и CC66 (SC5\_MLST) отличались наличием идентичной пары метилтрансфераз CPM I типа. Для ряда других сиквенс-типов группы B2 также была отмечена 100% идентичность одной из двух метилтрансфераз варианту метелтрансферазы, ассоциирующемуся с CC81: CC230, CC311, ST3104 (Рисунок 29).

Также наблюдалась ассоциация вариантов генов S-субъединиц с источником выделения изолятов (Рисунок 30). Число аллелей генов S-субъединиц среди изолятов, выделенных от носителей, было значительно больше и ассоциировалось с большим разнообразием сиквенс-типов и сиквенс-кластеров среди носителей (Рисунок 30). При этом среди инвазивных изолятов, выделенных из крови и ликвора, преобладали штаммы, имеющие варианты S-субъединиц, специфичные для изолятов групп A и B2 (Рисунок 30). Некоторые варианты S-субъединиц изолятов группы B1 ассоциировались преимущественно с носительством (Рисунок 30). Интересно, что большинство штаммов групп B1 и B2 несут идентичные варианты субъединиц специфичности.

Наши результаты согласуются с данными других исследований, в которых были показаны различия в степени вирулентности у изолятов, имеющих различный уровень экспрессии вариантов S-субъединиц [124]. Вариации метилирования играют важную роль в генерации мутных и прозрачных колоний [124, 126]. Хорошо известно, что прозрачные колонии ассоциируются с колонизацией, а мутные колонии – с инвазивными заболеваниями. При этом частота мутных и прозрачных колоний варьировала, в зависимости от варианта SpnIII, который экспрессировал штамм [124]. Авторы сделали вывод, что способность к быстрому переключению между различными эпигенетическими состояниями может быть преимуществом, обеспечивающим популяции бактерий способность быстро возвращаться к фенотипическому разнообразию после селективного или неселективного действия сдерживающих рост факторов, которые обычно происходят при инфицировании хозяина.

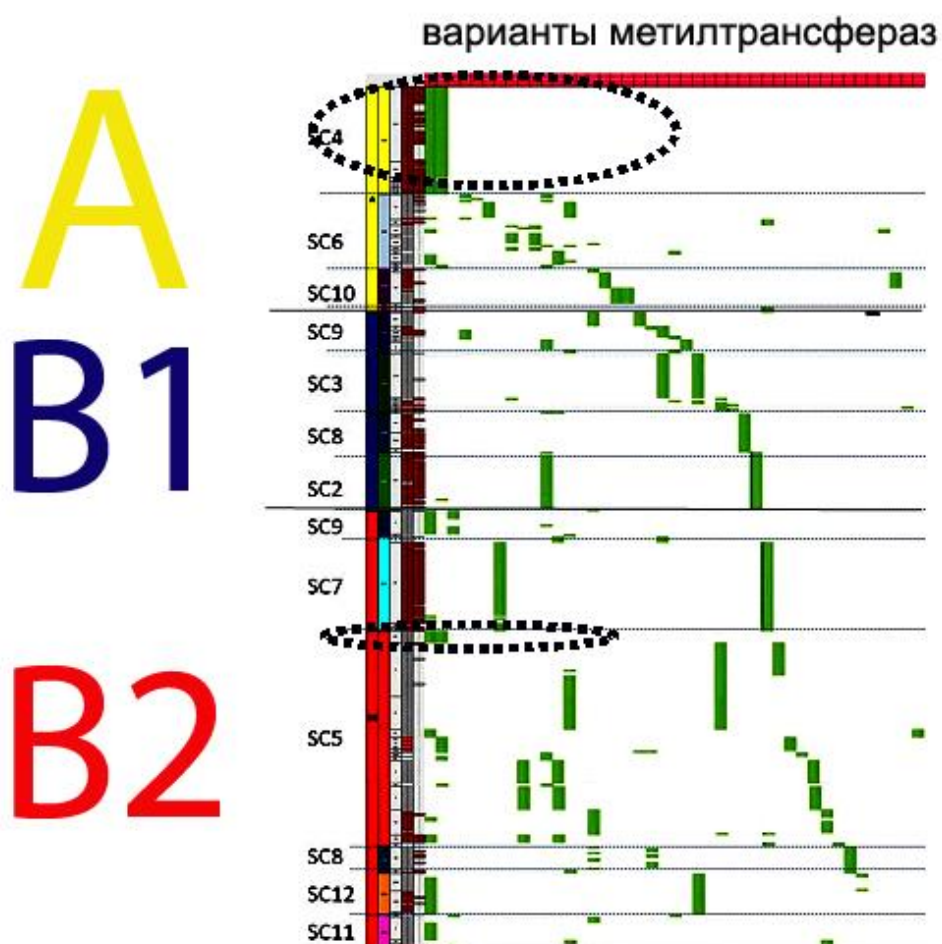
Наши результаты позволяют предположить, что в условиях развития пневмококковых инвазивных заболеваний происходит селекция штаммов групп A и B2, принадлежащих отдельным генетическим линиям, имеющим специфичные варианты генов S-субъединиц, которые могут давать данным штаммам способность к быстрому переключению между различными эпигенетическими состояниями, обеспечивающему популяции бактерий быстрое возвращение к фенотипическому разнообразию после

селективного или неселективного действия сдерживающих рост факторов, возникающих при инфицировании хозяина.

Таким образом, представители нескольких генетических линий группы В2 (СС66, СС230, СС199) имеют метилтрансферазы СРМ I типа, идентичные метилтрансферазам основного представителя группы А (СС81). Возможно, эта особенность способствовала успешности данных клонов, аналогично клону СС81.

Разнообразие вариантов S-субъединиц больше среди изолятов, полученных от носителей. Только некоторые варианты генов S-субъединиц представлены в геномах инвазивных изолятов.

Большинство представителей группы В имеют ряд идентичных субъединиц специфичности, что может быть связано с общими особенностями эпигенетической регуляции.



**Рисунок 29 - Ассоциация вариантов генов метилтрансфераз СРМ I типа с группами А/В1/В2, SC\_MLST и сиквенс-типами**

Горизонтальными линиями показаны границы сиквенс-кластеров. Выделены группы изолятов, принадлежащие ST81 (группа А) и ST66 (группа В2), ассоциирующиеся с одними и теми же вариантами метилтрансфераз.

## варианты гена S-субъединицы

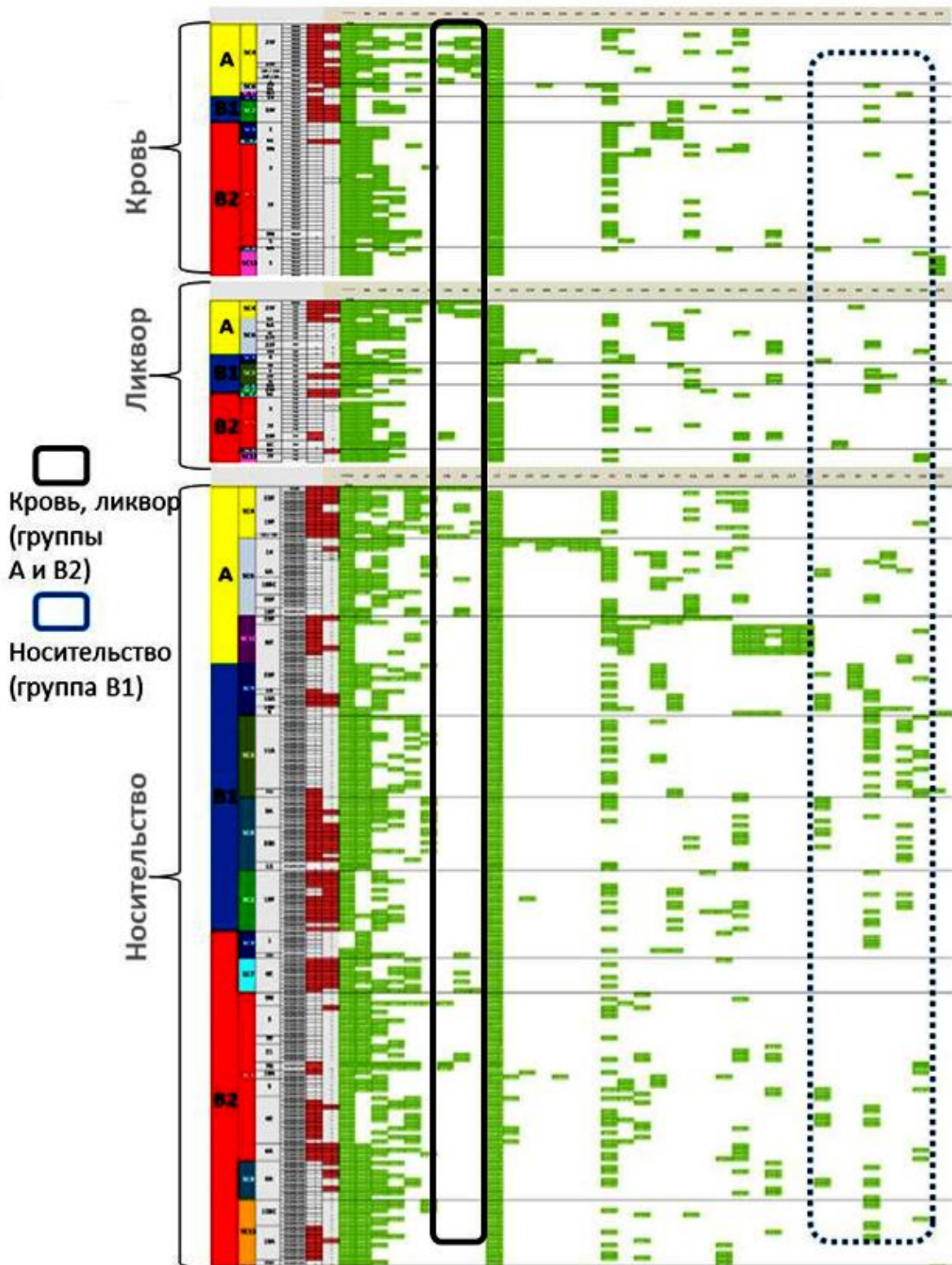


Рисунок 30 – Ассоциация вариантов генов субъединицы специфичности СРМ I типа с источником выделения изолятов (кровь, ликвор, носоглотка / носитель)

### 3.1.6. Анализ присутствия профагов в геномах анализируемых изолятов

Наиболее частый фенотип, ассоциируемый с системами СРМ, – рестрикция профага. Согласно результатам нашего анализа содержания профагов в геномах, все анализируемые изоляты содержат интактные или дефективные профаги, которые можно идентифицировать, либо остатки профагов, локусы которых несут гены факторов вирулентности либо системы рестрикции (Том 2 – Приложения. Приложение 2. Таблица 1). Число профагов в геноме варьировало от 1 до 7 и коррелировало с сиквенс-типом и сиквенс-кластером, вне зависимости от года получения изолята и региона распространения штамма (Том 2 – Приложения. Приложение 2. Таблица 1). Основные идентифицированные интактные профаги: Strept\_phiARI0923, Strept\_MM1, Strept\_phiARI0462, Strept\_phiARI0131, Strept\_phiARI0468, Strept\_MM1, Strept\_phiARI0460.

#### Выводы по разделу:

1. Мировая популяция *S. pneumoniae* делится на глобальные группы А, В1 и В2, которые ассоциируются преимущественно с разными серотипами.
2. Выявлена тенденция к распространению после 2000 г. генетических линий, ассоциированных с инвазивностью (группа В2). На фоне вакцинации ПКВ7 эволюционировала генетически гетерогенная группа В, представители которой ассоциируются с разными серотипами (не входящими в ПКВ7) и повышенной вирулентностью. При этом представители группы В2 имеют характерные общие сходные с группой А метаболические особенности (детерминируются типом глюкокиназы, глюкозо-6-фосфат-1-дегидрогеназы, ксантинфосфорибозил-трансферазы).
3. Согласно кластеризации изолятов по данным ядерного генома, на фоне вакцинации в мире вакцинами ПКВ7 и ПКВ13 происходило распространение в России инвазивных изолятов редких генетических линий, многие из которых ассоциировались с невакцинными серотипами (6С, 8, 10А, 16F, 17F, 22В, 22F, 25F, 34).
4. Генетические линии отличаются частотой и особенностями рекомбинационных событий. Резистентные к бета-лактамам антибиотикам изоляты ассоциировались с серотипами, отличающимися высокой эффективностью рекомбинаций (23F, 19F, 6Е, 6А, 14, 19А). Наши данные подтверждают данные других авторов о серотипах, ассоциирующихся с высокой частотой рекомбинаций (6А, 6Е, 23F, 19F, 14, 19А).

5. Принадлежность к генетическим линиям определяется структурой базовых систем метаболизма (в том числе ассоциирующихся с вирулентностью): системой рестрикции-модификации ДНК, *CiaN* двухкомпонентной регуляторной системой; компонентами синтеза жирных кислот; регуляцией метаболизма углеводов; метаболизмом пуринов.
6. Деление популяции пневмококка на группы А, В1 и В2 детерминировано:
  - белками, определяющими процессинг секретируемых белков (тип сигнальной пептидазы I);
  - характером усвоения глюкозы (типы глюкокиназы и глюкозо-6-фосфат-1-дегидрогеназы, регулирующих поток глюкозы в клетку);
  - особенностями синтеза жирных кислот, определяющими построение мембраны и клеточной стенки (типы ацетил-СоА-карбоксилазы и (3R)-гидроксимиристоил-дегидратазы);
  - регуляторной системой *CiaRH*, определяющей вирулентные свойства и успешность генетических линий.
7. Большинство представителей группы В имеют ряд идентичных субъединиц специфичности системы рестрикции I типа, что может быть обусловлено особенностями эпигенетической регуляции.
8. Особенности метаболизма углеводов и ароматических аминокислот (пневмококки способны синтезировать фенилаланин, тирозин и триптофан) детерминируют регуляцию клеточных процессов пневмококка и фенотипические особенности, в том числе вирулентность. В частности, расщепление популяции пневмококка по компонентам углеводного метаболизма (варианты генов *gnd*, *dexB*) коррелирует с расщеплением популяции по генам синтеза пептидогликана и регуляции экспрессии полисахаридной капсулы (варианты генов *wzg-wzh-wze*), а также с расщеплением по вариантам гена поверхностной гликозидазы *StrH*.
9. Ген *groV* вносит вклад в расщепление популяции пневмококка на резистентные и чувствительные к бета-лактамам штаммы.
10. Особенности метаболизма пневмококка детерминируют его значительный вирулентный потенциал. Серотип-специфичная вакцинация, приводящая к исключению из циркуляции распространенных генетических линий пневмококка, обладающих большей способностью к адаптации за счет энерго-обеспечивающих

систем (АТФ, НАДФН, фосфотрансферазные системы), может приводить к распространению генетических линий, характеризующихся большей вирулентностью.

### **3.2. Анализ детерминант резистентности к антибиотикам изолятов *S. pneumoniae*, относящихся к эпидемическим генетическим линиям**

#### **3.2.1. Анализ детерминант резистентности в геномах российских и референсных штаммов**

Детерминанты резистентности к бета-лактамным антибиотикам были предсказаны с помощью RAST-сервера, и их наличие коррелировало с данными антибиотикочувствительности. В первичной выборке, чувствительные к пенициллину изоляты ассоциировались с генетическими линиями (в порядке убывания процента чувствительных изолятов): CC62<sup>11A</sup>, CC191<sup>7F</sup>, CC113<sup>18BC</sup>, CC123<sup>17F</sup>, CC124<sup>14</sup>, CC327<sup>6A</sup>, CC433<sup>7F</sup>. Резистентные изоляты ассоциировались преимущественно (в порядке убывания процента резистентных изолятов) с: C81<sup>23F,19F</sup>, CC90<sup>6E</sup>, CC236<sup>19F</sup>, CC271<sup>19F</sup>, CC320<sup>19F,19A</sup>, CC448 (бескапсульные пневмококки). При этом небольшой процент умеренно устойчивых к пенициллину изолятов присутствовал практически во всех генетических линиях. Устойчивые к пенициллину изоляты ассоциировались преимущественно с серотипами (в порядке убывания процента устойчивых изолятов): 23F, 19F, 6E, 6A, 19A, 9A, 6B, 15A, 23A (Том 2 – Приложения. Приложение 9. Рисунок 71).

Детерминанты резистентности к макролидным антибиотикам также были предсказаны с помощью RAST-сервера, и их наличие было подтверждено с помощью blast-поиска с использованием базы данных последовательностей (Раздел 2.16. «Анализ наличия детерминант резистентности к макролидным антибиотикам»). Резистентные к макролидам изоляты принадлежали преимущественно CC81<sup>23F,19F</sup>, CC90<sup>6E</sup>, CC236<sup>19F</sup>, CC271<sup>19F</sup>, CC320<sup>19F,19A</sup>, CC448 (бескапсульные пневмококки). Некоторый процент резистентных к макролидам изолятов составлял практически все генетические линии пневмококка, кроме CC53<sup>8</sup>, CC62<sup>11A</sup>, CC1012<sup>11A</sup>, а также редких сиквенс-типов, ассоциированных с серотипом 8. Резистентность к макролидам была обусловлена наличием различных транспозонов семейства Tn916 во всех устойчивых штаммах (Том 2 – Приложения. Приложение 9. Рисунок 72). При этом Tn2010 присутствовал преимущественно у резистентных изолятов из России, а также у изолятов из Китая,

Канады и Малайзии (Том 2 – Приложения. Приложение 9. Рисунок 73). Резистентные к макролидам изоляты из России также содержали другие транспозоны семейства Tn916 (Том 2 – Приложения. Приложение 9. Рисунок 73), при этом вариант транспозона ассоциировался с генетической линией (Том 2 – Приложения. Приложение 9. Рисунок 72). Присутствие транспозонов также коррелировало с серотипом (изоляты серотипов 19F, 23F, 6E, 14, 35B, 24F, 6B, 19A содержали транспозоны семейства Tn916) (Том 2 – Приложения. Приложение 9. Рисунок 74). Высоко устойчивые к макролидам изоляты, принадлежащие преимущественно серотипам 23F и 6E, несли сложносоставной конъюгативный элемент Tn5253, включающий Tn916. Данные изоляты ассоциировались преимущественно с CC81, циркулировавшим ранее повсеместно, а также с CC90, распространенным преимущественно в странах Европы (Том 2 – Приложения. Приложение 9. Рисунок 75).

Ген 23S-рНК-метилтрансферазы *ermB* присутствовал у большинства резистентных к макролидам изолятов в составе кассеты OMEGA (преимущественно у резистентных изолятов из России: CC81, ST11236, ST12825, CC143, CC236, CC271, ST1464, ST2812, ST12822, ST12827), либо в составе Tn917 (изоляты из России CC179) (Том 2 – Приложения. Приложение 9. Таблица 1). Ген *ermB* отсутствовал у резистентных к макролидам изолятов CC236, CC156, CC376, а также у отдельных изолятов CC81. Регистентность к макролидам у изолятов данного сиквенс-типа была обусловлена только наличием MEGA-кассеты с генами *mef(E)/mel* в транспозонах семейства Tn916 (Том 2 – Приложения. Приложение 9. Таблица 1).

Резистентные к тетрациклину штаммы практически полностью составляли генетические линии CC81<sup>23F,19F</sup>, CC90<sup>6E</sup>, CC185<sup>6E</sup>, CC236<sup>19F</sup>, CC271<sup>19F</sup>, CC320<sup>19A</sup>. Некоторое количество резистентных изолятов присутствовало в составе других генетических линий, за исключением CC53<sup>8</sup>, CC62<sup>11A</sup>, CC1012<sup>11A</sup>, CC180<sup>3</sup>, CC505<sup>3</sup>, CC448 (бескапсульные пневмококки). Резистентность к тетрациклину была обусловлена наличием гена *tetM*, кодирующего белок, ответственный за инактивацию тетрациклина и защиту рибосом-мишеней. Ген *tetM* присутствовал в составе конъюгативных транспозонов семейств Tn916 и Tn5253 в устойчивых к тетрациклину изолятах, принадлежащих CC81, CC344, CC242, CC63, CC179, CC 236, CC271, CC320, CC315, CC90, а также резистентных изолятах редких сиквенс-типов (Том 2 – Приложения. Приложение 9. Таблица 1).



Наличие детерминант резистентности к хлорамфениколу и ко-тримоксазолу в структуре генома пневмококков было подтверждено с помощью RAST-сервера (Том 2 – Приложение. Приложение 2. Таблица 1). Резистентные к хлорамфениколу изоляты принадлежали преимущественно клональным комплексам CC81, CC90, а также некоторым представителям CC217. В целом, резистентность к тетрациклинам и хлорамфениколу наблюдалась значительно реже, чем к другим антибиотикам. Изоляты, резистентные к триметоприм/сульфометаксазолу присутствовали в составе многих генетических линий, за исключением CC53, CC62, CC1012, CC180, CC505, CC448 (которые были полностью чувствительны).

Таким образом, наличие резистентности к антибиотикам различных классов ассоциировалось с принадлежностью изолята к генетической линии и серотипу. По данным антибиотикочувствительности, мультирезистентные изоляты (к пенициллину и эритромицину) составляли 75,45% (206 из 273 изолятов). Резистентность к макролидам и тетрациклинам обусловлена наличием конъюгативных транспозонов семейств Tn916 и Tn5253, варианты которых ассоциировались с генетическими линиями. В пневмококковой популяции в России на протяжении длительного времени циркулировали мультирезистентные генетические линии CC236, CC271 и CC320 (и их однолокусные и двулокусные варианты), представители которых несут Tn2010 и были распространены преимущественно в России, странах Азии и в Канаде.

### **3.2.2. Динамика циркуляции генетических линий в России в различные периоды с 1980 по 2017 гг. Ассоциация с резистентностью к антибиотикам**

Для анализа российских изолятов анализируемый период (с 1980 по 2017 гг.) был разбит на четыре временных интервала: с 1980 по 2001 гг. (до начала повсеместного применения ПКВ7); с 2002 по 2010 гг. (период повсеместного внедрения ПКВ7 в мире); с 2011 по 2014 гг. (период повсеместного применения ПКВ10 и ПКВ13 в мире); с 2015 по 2017 гг. (ранний период после начала вакцинации ПКВ13 в России). Наибольший процент изолятов был доступен для второго и третьего периодов. Каждый временной интервал ассоциировался с циркулировавшими в данный период клональными комплексами (Том 2 – Приложение. Приложение 9. Рисунок 76).

В период с 1980-2001 гг., распространенные в России генетические линии пневмококка принадлежали клональным комплексам, преимущественно ассоциирующимся с чувствительностью к пенициллинам и макролидам: CC102<sup>18ABCF</sup>,



СС1527<sup>12F</sup>, СС2296<sup>1</sup>, СС180<sup>3</sup> (Рисунок 31). С устойчивостью к пенициллинам и эритромицину ассоциировалась преимущественно генетическая линия С236 (серотип 19F). Наиболее распространенными серотипами в России, в период с 1980 по 2001 гг., были: 1, 18ABCF, 19F, 3, 12F, 6B, 14 (Рисунок 31).

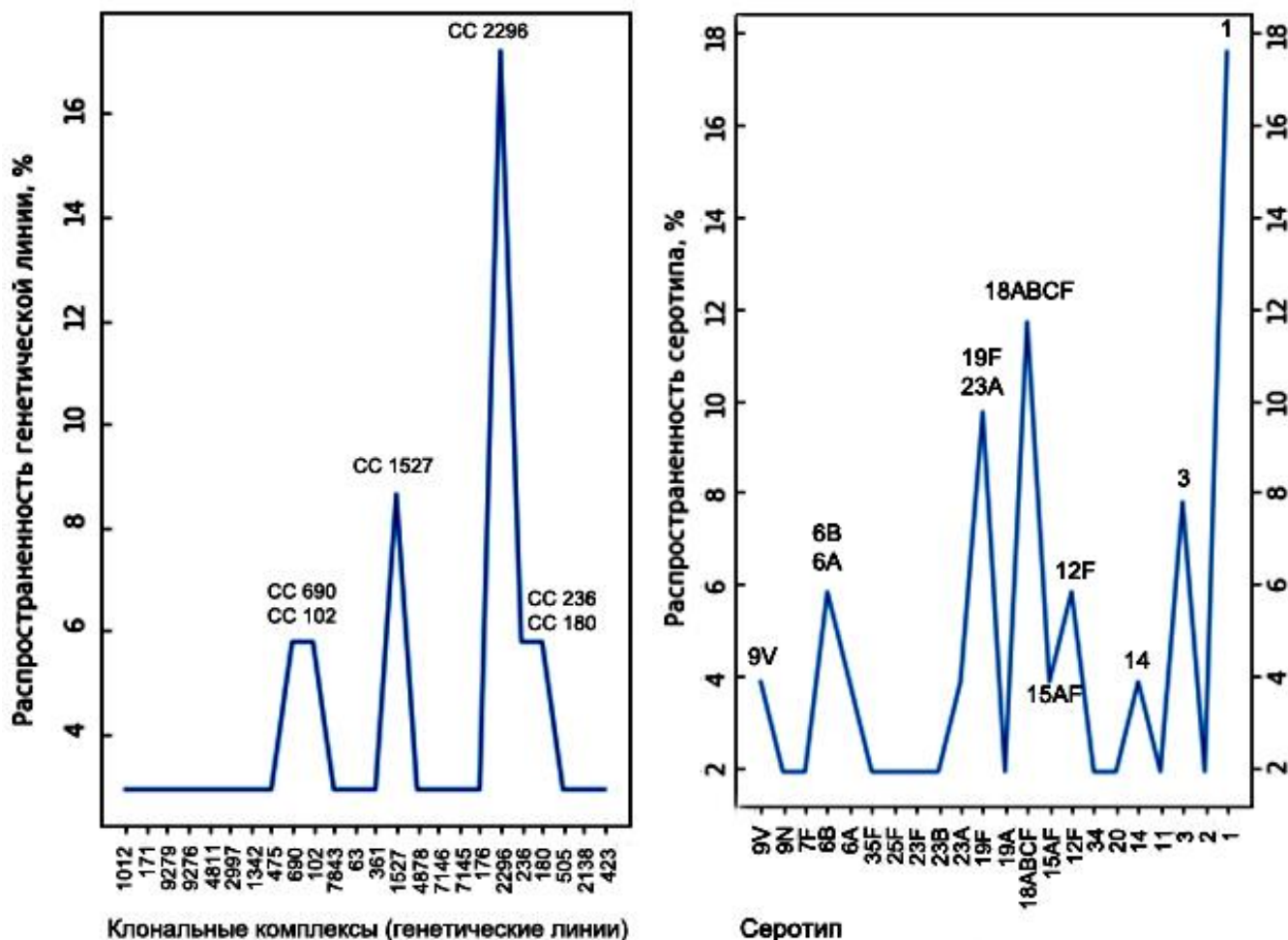


Рисунок 31 – Распространенность генетических линий пневмококков в России в период с 1980 по 2001 гг.

В период 2002-2010 гг. в России происходило преимущественное распространение генетических линий, ассоциированных с резистентностью: СС81<sup>23F</sup>, СС361<sup>19A</sup>, СС236<sup>19F</sup>, СС315<sup>6B</sup>, СС162<sup>19A</sup>, СС30 (ST1500<sup>23F</sup>), СС143<sup>14</sup>, СС230<sup>19F,19A</sup>, СС276<sup>19A</sup> (Рисунок 32). Распространенность СС2296 (серотип 1) снизилась. Также наблюдалась тенденция изменения циркуляции серотипов в этот период – распространение серотипа 19А, вероятно, на фоне повсеместного в мире применения вакцин ПКВ7 и ПКВ10 (Рисунок 32). Резистентные к пенициллину изоляты серотипа 19А ассоциировались с различными генетическими линиями. К другим

распространенным в России серотипам в данный период относились серогруппа 6ABCD, серотипы 3, 19F, 14, 7F, 9N, 11A (Рисунок 32).

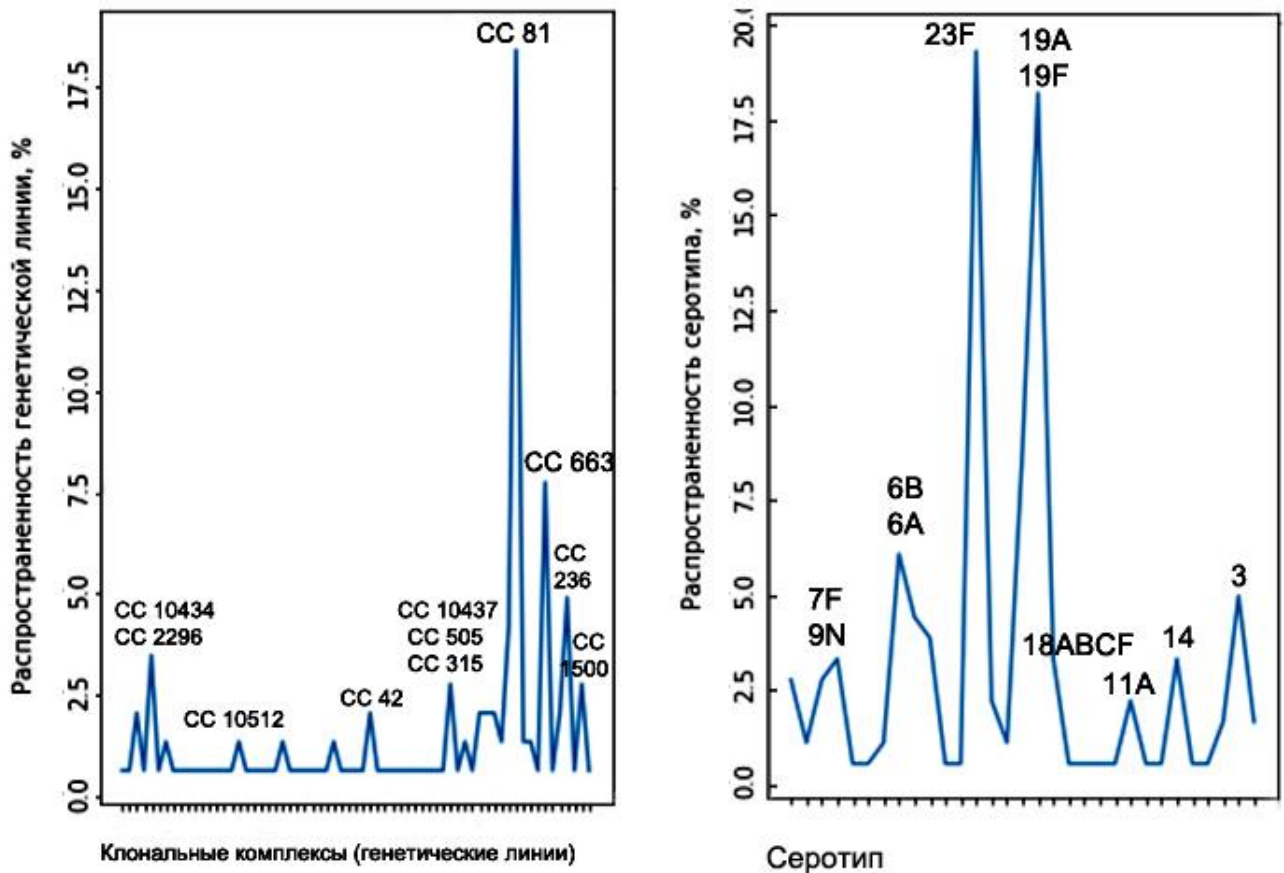
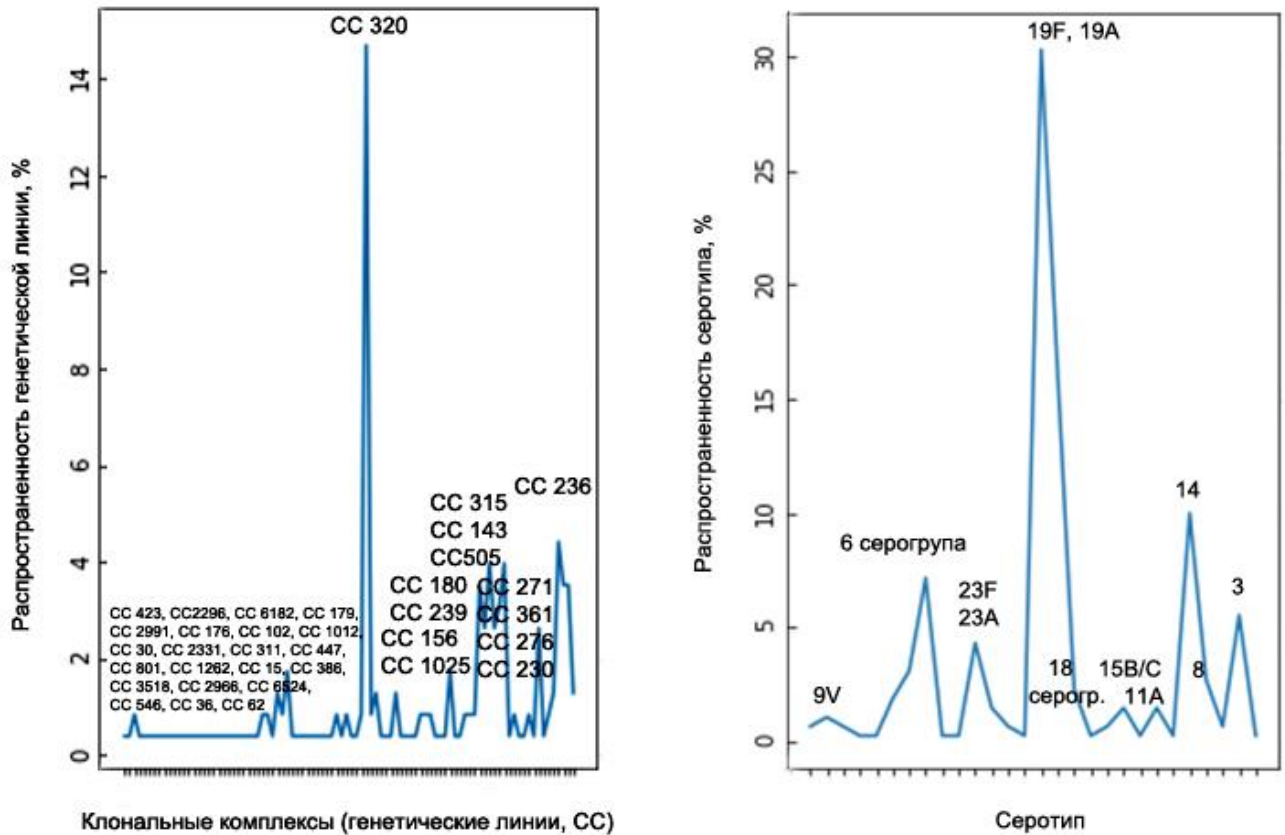


Рисунок 32 – Распространенность генетических линий пневмококковой популяции в России в период с 2002 по 2010 гг.

В период 2011-2014 гг. в России широко распространились генетические линии, большинство из которых ассоциировались с мультирезистентностью к пенициллинам и макролидам: CC320<sup>19A</sup>, а также CC236<sup>19F</sup>, CC315<sup>6B</sup>, CC143<sup>14,19A</sup>, CC180<sup>3,19A</sup>, CC239<sup>19F</sup>, CC505<sup>3,6ABCD,19A</sup>, CC271<sup>19F,19A</sup>, CC230<sup>19A</sup>, CC276 (ST276, ST1611, ST10431 - серотипы 19F/19A), CC361 (ST663, ST9658, ST10434 - серотипы 19F/19A), CC156 (ST156, ST875 - серотип 14), CC1025<sup>3,19F</sup> (Рисунок 33). На фоне повсеместного применения вакцин ПКВ7 и ПКВ13 в мире в России произошло значительное снижение распространенности серотипа 23F (ассоциированная с мультирезистентностью генетическая CC81), при этом значительно распространился серотип 19A (Рисунок 32). При этом детерминанты резистентности приобрели ранее чувствительные к антибиотикам CC180, CC505, по-видимому, в связи с переключением ряда их представителей с серотипа 3 на серотип 19A (Рисунок 33). Возможно, распространение резистентных штаммов связано с

влиянием вакцинации (большая часть резистентных штаммов была ассоциирована с серотипом 19А).



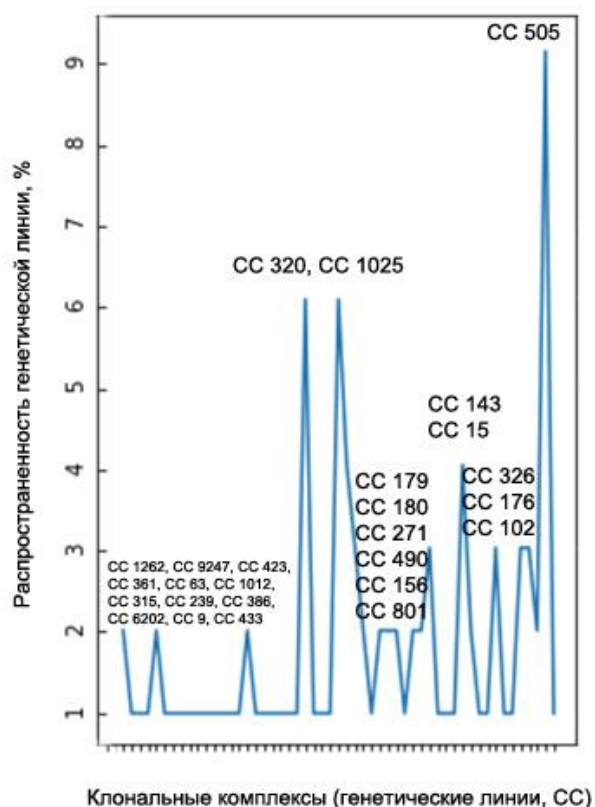
**Рисунок 33 – Распространенность генетических линий пневмококковой популяции в России в период с 2011 по 2014 гг.**

В ранний период после внедрения вакцинации ПКВ13 в России (с 2014 по 2017 гг.) сохранилась циркуляция на высоком уровне CC505 (серотип 3, преимущественно чувствительные к антибиотикам изоляты). При этом эволюционировала генетическая линия CC320 (появление редких сиквенс-типов ST1464, ST2812, ST7178, Unkn\_2, Unkn\_3, Unkn\_5 – серотип 19F), большая часть была представлена мультирезистентными изолятами (Рисунок 34). Среди других мультирезистентных к антибиотикам генетических линий были распространены: CC15 (ST15, ST13658, ST13665, ST13844 – серотип 14), CC143 (ST12618, ST13659, ST13661, ST13664 – серотип 14), CC179 (серотип 19F), CC236 (серотип 19F), CC311 (серотипы 19А, 23F).

Генетическая линия CC1025 включала преимущественно редкие сиквенс-типы (ST1025, ST13662, ST13688 и ST13810 - серотипы 19F, 14 и 15BC, соответственно) и была представлена чувствительными к антибиотикам изолятами. Другие чувствительные к антибиотикам изоляты были представлены генетическими линиями

CC102 (ST102, Unkn\_10 - серогруппа 18ABCF), CC176 (ST176, Unkn\_12 – серотипы 6В и 23F, соответственно).

Таким образом, в ранний период после начала вакцинации в России снизилась распространенность циркулировавших ранее генетических линий серотипов 19А и 19F, ассоциированных с мультирезистентностью. При этом наблюдалась эволюция генетических линий – появились ассоциации редких сиквенс-типов, принадлежащих циркулировавшим ранее клональным комплексам, с вакцинными серотипами 19F и 14. Также распространилась генетическая линия CC505, ассоциированная с серотипом 3. За счет этого не наблюдалось значительных тенденций к изменению серотипового состава популяции (кроме повышения распространенности невакцинных серогрупп 11AD и 15BC, данные не показаны).



**Рисунок 34 - Распространенность генетических линий пневмококковой популяции в России в период с 2015 по 2017 гг.**

### **3.2.3. Значимые однонуклеотидные полиморфизмы, ассоциированные с резистентностью к бета-лактамам антибиотикам у изолятов CC320**

Распространенная в России генетическая линия *S. pneumoniae* CC320 представлена изолятами серотипов 19F и 19А, характеризующимися мультирезистентностью к антибиотикам разных классов. Субвыборка для полногеномного анализа CC320 включала 46 изолятов, представленных однолокусными

и двулокусными вариантами референсного штамма *S.pneumoniae* TW31 (Том 2 – Приложение. Приложение 8, Таблица 1).

В результате оценки миссенс-SNP профиля у трех чувствительных к пенициллину штаммов было обнаружено 139 миссенс-мутаций в 45 генах. В Таблице 19 представлен перечень генов и кодируемых ими продуктов, полиморфизм которых мог быть обусловлен развитием резистентности к пенициллину.

**Таблица 19 – Гены пенициллин-резистентных штаммов *S.pneumoniae*, мутации в которых могут быть ассоциированы с резистентностью к пенициллину**

Ген	Продукт	Функция	
<i>mraY</i>	Ундекапринилфосфат фосфо-N-ацетилмурамоил-пентапептид-трансфераза	Синтез липида I	<b>Синтез пептидогликана</b>
<i>pbp2x</i>	Pbp2X	Регуляция поперечного роста клетки	
<i>pbp2b</i>	Pbp2B	Регуляция периферического роста клетки	
<i>ftsL</i>	FtsL	Регуляция поздней фазы клеточного деления	<b>Клеточное деление</b>
<i>gpsB</i>	GpsB	Белок клеточного деления	
<i>mapZ</i>	Срединно расположенный белок Z	Регуляция ранней фазы клеточного деления	
<i>engB</i>	YsxC/EngB	ГТФазная активность	
<i>recU</i>	Резолваза RecU	Эндонуклеаза. Репарация ДНК, рекомбинация, сегрегация хромосом	
<i>rlmL</i>	Гипотетическая N6-аденин-специфичная метилаза ДНК	Метилирование ДНК	<b>Метилирование ДНК</b>
<i>rpiR</i>	Регулятор утилизации сиаловой кислоты, семейство RpiR	Регулятор транскрипции	<b>Метаболизм фруктозы и маннозы</b>
	Белок семейства MutT/Nudix	Фосфогидролаза	<b>Репарация ДНК</b>
	Гепариназа-II/III-подобный белок	Лиаза	<b>Метаболизм углеводов, одноуглеродный метаболизм, модуляция экспрессии факторов вирулентности</b>
<i>regR</i>	Репрессор утилизации гиалуроната и дезоксикетоглюконат киназы RegR	Регулятор транскрипции	
	Метилтрансфераза H малой субъединицы рибосомальной РНК	Метилирует N4 позицию цитидина в положении 1402 (C1402) 16S рРНК	<b>Процессинг РНК</b>
	Нуклеотид-связывающий белок с гистидиновой триадой тРНК и рРНК цитозин-С5-метилаза	Относится к суперсемейству нуклеотидных гидролаз и трансфераз	
	Гипотетическая тРНК-m1A22 метилаза	Метилирует С5-позицию цитозина в тРНК и рРНК	
	Гипотетическая тРНК-m1A22 метилаза	Модификации нуклеозидов тРНК для стабилизации их третичной структуры и улучшения их способности к процессу декодирования	
	Гипотетический белок (FIG01114714)	Ассоциированный с мембраной гипотетический белок	
<i>upsA</i>	Гипотетический белок (FIG005686)	Консервативный цитоплазматический белок, UPF0398 семейство	

### Идентификация консервативных аминокислотных замен в идентифицированных белках (Таблица 19) у резистентных к пенициллину штаммов *S. pneumoniae*

В сравниваемых последовательностях белка Pbp2X были идентифицированы 54 АМК-замены у всех резистентных к пенициллину штаммов относительно чувствительных штаммов. В области транспептидазного домена (позиции 266-616) обнаружена 31 АМК-замена. В С-терминальной области (616-750) идентифицировано 20 АМК-замен. В пенициллин-связывающей области (49-266) найдены 3 АМК-замены. Из указанных в Таблице 20 АМК-замен ранее были охарактеризованы мутации: T338A (связана с изменением активности серина активного центра); R384G (приводит к конформационному изменению петли, прилегающей к поверхности активного центра); замены R384G, D567N и N605T – связаны с изменением кинетики взаимодействия Pbp2X с  $\beta$ -лактамами антибиотиками [335].

АМК-последовательности белка Pbp2B всех резистентных штаммов отличались от чувствительных штаммов наличием 7 мутаций в области транспептидазного домена (313-680) (Таблица 20). Транспептидазная область белка Pbp2B содержит три консервативных мотива: S386VVK, S443SN, and K615TG. Все обнаруженные 7 мутаций у резистентных штаммов были локализованы между высоко консервативными мотивами S386VVK и S443SN. Мутация T446A была охарактеризована ранее. У точечных T446A R6-мутантов в белке Pbp2B было описано снижение аффинности к пенициллину на 60% по сравнению со штаммом дикого типа [260]. Мутации вблизи консервативного мотива K615TG могут модифицировать гибкость  $\beta$ 3-спирали, граничащей с активным центром, и изменять таким образом кинетику связывания Pbp2B белка с бета-лактамами антибиотиками [260]. Аминокислоты Asn659, Gly660 и Ser664 находятся в С-терминальной области спирали  $\alpha$ 11. Возможно, мутации в данных областях могут индуцировать конформационные изменения в области активного центра, которые создают препятствия для связывания с антибиотиками [260].

Белок клеточного деления FtsL регулирует поперечный рост клетки и участвует в формировании комплекса белков клеточного деления FtsQ и FtsB/FtsL: его периплазматический домен взаимодействует с белком FtsQ, трансмембранный домен взаимодействует с белком FtsB, цитоплазматический домен взаимодействует с трансмембранным белком FtsW [336, 337]. У большинства резистентных к пенициллину

штаммов *S. pneumoniae* обнаружена мутация в коротком N-концевом цитоплазматическом домене (K4R).

Белок GpsB консервативен у всех грам-положительных бактерий. Благодаря взаимодействию молекул в суперспирализованной области N-концевого домена (позиции 36-63), белок GpsB способен к полимеризации. Белок GpsB взаимодействует с белками клеточного деления Pbp1, MreC и EzrA [338]. Предполагают, что белок GpsB играет роль в переключении между синтезом клеточной стенки в периферическом направлении и синтезом перегородки, посредством релокализации белка Pbp1. У всех резистентных к пенициллину штаммов *S. pneumoniae* мы обнаружили мутации в области N-концевого домена, гомологичного белку DivIVA (A2T; N24S; K65Q), а также мутации в области C-концевого домена (Таблица 20).

ГТФ-связывающий белок EngB (YsxC/EngB) контролирует различные биологические процессы, в том числе клеточное деление, клеточный цикл, сигнальную трансдукцию, сборку рибосом и синтез белка, способствуя чередованию неактивной ГДФ-связанной и активной ГТФ-связанной формы белков [338]. ГТФазы содержат консервативный гуанин-нуклеотид связывающий домен, который отвечает за связывание и гидролиз гуаниновых нуклеотидов и представлен центральным  $\beta$ -слоем, фланкированным  $\alpha$ -спиралями. Консервативные участки включают: два металл-связывающих сайта ( $Mg^{2+}$ ; позиции 39 и 61), а также пять областей, связывающих гуанидиновый нуклеотид (32-39 – Р-петля, захватывает фосфатную группу гуанидинового нуклеотида; 59-63 – связывает ионы магния; 77-80, 144-147 и 176-178 – отвечают за связывание второй части гуанидинового нуклеотида). У ряда штаммов *S. pneumoniae* ST236 серотипа 19F, отличающихся высокой резистентностью, была обнаружена мутация C96R.

Срединно-расположенный белок MarZ формирует кольцо на экваторе вновь поделившихся клеток. После начала клеточного деления, во время периферического роста, сформированное белком MarZ кольцо расщепляется на два кольца, которые перемещаются к полюсам клетки [338, 339]. Белок имеет три основных домена: N-концевой топологический домен, локализованный в цитоплазме (ограничен АМК-остатками 1-158), трансмембранный домен (159-179), C-концевой топологический внеклеточный домен (180-464). N-терминальная область взаимодействует с тубулиноподобным белком FtsZ, C-терминальный домен связывает пептидогликан и участвует в

периферическом делении клетки. У всех резистентных к пенициллину изолятов *S. pneumoniae* были идентифицированы мутации как в цитоплазматическом (A13G; F75L; K94E; L110P; D127E; D148E), так и во внеклеточном домене (I161V; V164A; A184N; V257A; L273P; Q327E; K343E; P344T). Мутации в С-концевом домене белка MarZ локализуются между многочисленными областями формирования спиралей,  $\beta$ -цепей и серин-обогащенного участка (311-365).

Резолвала RecU участвует в сегрегации хромосом во время клеточного деления, рекомбинации ДНК и репарации ДНК. Фермент содержит четыре металл-связывающих сайта (позиции 81, 83, 96, 115), а также сайт, стабилизирующий состояние транзиции (98). Аллостерическая транзиция обеспечивает гибкую систему регуляции активности ферментов. У всех резистентных к пенициллину штаммов *S. pneumoniae* были обнаружены мутации V8L; T14P; Q60R; H89N; V97A; V122I; F147Y; E165G. При этом аминокислотная замена V97A локализуется рядом с сайтом, стабилизирующим состояние аллостерической транзиции.

Фермент ундекапринилфосфат фосфо-N-ацетилмурамоил-пентапептид-трансфераза (MraY) вовлечен в первый (цитоплазматический) этап биосинтеза клеточной стенки пептидогликана. Белок MraY является компонентом клеточной мембраны, он содержит 10 трансмембранных областей (позиции: 3-23, 51-71, 79-99, 115-135, 138-158, 169-189, 195-215, 221-243, 247-269, 306-326). У всех резистентных к пенициллину штаммов *S. pneumoniae* идентифицированы мутации T60A; S73K(N); F75L; V170I; I213V.

Структура гипотетической N6-аденин-специфичной метилазы ДНК представлена двумя доменами: N-концевой THUMP-домен (АМК-остатки 45-159) и С-концевой UPF0020 домен (собственно, метилаза). Метилирование ДНК является ключевым механизмом, контролирующим экспрессию генов, поддержание целостности генома клетки, защиту от встраивания ДНК фагов и транспозонов. У всех резистентных к пенициллину штаммов *S. pneumoniae* мы обнаружили мутацию I235V в С-концевом домене метилтрансферазы.

Охарактеризованные белки суперсемейства MutT/Nudix гидролизуют нуклеозид дифосфат, связанный с каким-либо другим фрагментом (X). Гидролазы семейства Nudix могут быть необходимы для защиты клеток от потенциально опасных метаболитов и для снижения накопления промежуточных продуктов ключевых биохимических процессов.



Белки суперсемейства MutT/Nudix содержат симметричный Nudix-мотив (спираль-петля-спираль), играющий роль в связывании белка с ДНК. Мы обнаружили две аминокислотные замены в С-терминальной области (E124Q; D125N) у всех резистентных к пенициллину штаммов *S.pneumoniae*. Идентифицированные АМК-замены локализованы вне области Nudix-мотива, образующего короткую петлю и содержащего каталитические аминокислоты. Считается, что АМК-остатки, находящиеся за пределами Nudix-мотива, определяют субстратную специфичность фосфогидролазы.

Сравнимые аминокислотные последовательности Гепариназа-II/III-подобного белка отличались у всех резистентных штаммов наличием 3-х АМК-замен в области N-терминального домена (G172S; D206E; D385G). Известно, что  $\alpha$ -спирали N-терминального домена гепариназы III у *Bacteroides thetaiotaomicron* участвуют в формировании туннеля, содержащего каталитический активный центр [340]. Гепариназа является фактором инвазии. Ген, кодирующий гепариназа-II/III подобный белок является частью RegR-регулона, который включает три граничащих друг с другом транскрипционные единицы.

Белок RegR относится к регуляторам транскрипции, принадлежащим к семейству LacI/GalR. Белки данного семейства характеризуются наличием мотива спираль-петля-спираль в N-концевом домене, вовлеченного в связывание этим белком ДНК, а также мотива в С-концевом домене, связывающего белки и углеводы и выполняющего регуляторные функции [341]. У всех резистентных к пенициллину изолятов *S.pneumoniae*, принадлежащих ST236, ST271, а также некоторым другим однолокусным вариантам ST236, мутации в последовательностях белка RegR были локализованы в С-концевом домене (N79T; Q151R; S216R; D238N; I247V; L254P; S295A; R313L; N327S). Интересно, что большинство изолятов *S.pneumoniae* ST320 имели только одну замену в С-концевом домене (E181V) белка RegR, однако, у трех штаммов *S.pneumoniae* ST320, отличающихся высоким уровнем резистентности к пенициллину (МПК: 6, 6 и >4 мг/л), был обнаружен почти тот же профиль мутаций, что и у всех резистентных штаммов ST236 (Таблица 20), а также дополнительная мутация в N-концевом домене (E41K).

**Таблица 20 – Консервативные аминокислотные замены у пенициллин-резистентных штаммов *S. pneumoniae* CC320**

Код белка	Наименование	АМК-замены у большинства резистентных штаммов
EngB	GTF-связывающий белок YsxC/EngB	C96R у некоторых высокорезистентных изолятов серотипа 19F
WP_00019363 4.1	Белок семейства MutT/nudix	E124Q; D125N
Pbp2B	Pbp2B Пептидогликан синтетаза	S412P; N422Y; T426K; Q427L; Q438E; T446A; L455I
WP_00065703 3.1	Гепариназа-II/III-подобный белок	G172S; D206E; D385G
RegR	Репрессор утилизации гиалуроната и дезоксикетоглюконат киназы RegR	N79T; Q151R; S216R; D238N; I247V; L254P; S295A; R313L; N327S (у всех <i>S. pneumoniae</i> ST236, ST271 и других однолокусных вариантов ST236)
WP_00071233 2.1	Гипотетический белок (FIG01114714)	S78A; D133G
FtsL	Белок клеточного деления FtsL	K4R
Pbp2X	Pbp2X	A172T/A172D; R254Q; M256V; D301N; T338A; A346S; A347S; G354S; V358Y; L364F; I371T; G382T; R384G; S389L; T401S; N417K; N444S; I462L; P486T; D488N; T490S; A491V; V516I; V523L; S530Y; T536I; V537I; L546V; V565S; D567N; S574T; S576N; N605T; D616E; T622S; T623P; A626N; E628D; S631T; Q632T; Q633E; P635S; P637A; Y641I; E652A; L657I; N671E; S672T; A674V; I688L; A693V; V696I; L710F; Q721E
MraY	Ундекапринилфосфат фосфо-N-ацетилмурамоил-пентапептид-трансфераза	T60A; S73K(N); F75L; V170I; I213V
RecU	Резолваза RecU	V8L; T14P; Q60R; H89N; V97A; V122I; F147Y; E165G
YpsA	Гипотетический белок (FIG005686)	S51T; E164Q
GpsB	Белок клеточного деления GpsB	A2T; N24S; K65Q; K65_P66_insVSSA; V69S; Q70H; A71P; E72D; L74I; E75D; A76V; I78A; T79S; S107T; F109L
RlmL	Гипотетическая N6-аденин-специфичная метилаза ДНК	I235V
MapZ	Срединно расположенный белок Z	A13G; F75L; K94E; L110P; D127E; D148E; I161V; V164A; A184N; V257A; L273P; Q327E; K343E; P344T
WP_001278301 1	Нуклеотид-связывающий белок с гистидиновой триадой	T108A

**Выводы по разделу:**

1. Наличие резистентности к антибиотикам различных классов ассоциируется с принадлежностью изолята к генетической линии и серотипу. Резистентность к макролидам и тетрациклинам обусловлена наличием конъюгативных транспозонов

семейств Tn916 и Tn5253, варианты которых ассоциируются с генетическими линиями.

2. В пневмококковой популяции в России на протяжении длительного времени циркулируют мультирезистентные генетические линии CC236, CC271 и CC320 (и их однолокусные и двулокусные варианты), представители которых несут Tn2010, распространенный преимущественно в России, странах Азии и в Канаде.
3. Возможно, что повсеместная вакцинация ПКВ7 и ПКВ10, проводимая с начала 2000-х гг. до 2014 г. в мире, повлияла на распространение резистентных штаммов в России, так как большая часть резистентных штаммов в России в период с 2011 г. по 2014 г. была ассоциирована с серотипом 19А, что совпадало с мировой тенденцией.
4. В ранний период после начала вакцинации в России снизилась распространенность циркулировавших ранее генетических линий серотипов 19А и 19F, ассоциированных с мультирезистентностью. Однако наблюдалась эволюция генетических линий – появились ассоциации редких сиквенс-типов, принадлежащих циркулировавшим ранее клональным комплексам, с вакцинными серотипами 19F и 14. Также распространилась генетическая линия CC505, ассоциированная с серотипом 3. За счет этого не наблюдалось значительных тенденций к изменению серотипового состава популяции (кроме повышения распространенности невакцинных серогрупп 11AD и 15BC).
5. Первичные мишени антибиотиков, особенно ПСБ-белки, функционируют на стыке регулируемых метаболических путей, поэтому развитие резистентности к антибиотикам часто связано со множеством механизмов адаптации. Помимо генов, кодирующих белки Pbp1A, Pbp2B и Pbp2X, для развития резистентности критичны мутации в генах, отвечающих за синтез пептидогликана (*mraW* и *mraY*), клеточное деление (*ftsL*, *gpsB*), кодирующих шапероны (*clpL*, *clpX*), а также обеспечивающих процесс генетической рекомбинации (*recU*).
6. Гены, отвечающие за синхронизацию процессов синтеза пептидогликана и клеточного деления, входят в состав *dcw*-кластера (division and cell wall) [338, 339, 342, 343]. Наши результаты показывают, что у резистентных к пенициллину штаммов *S.pneumoniae* наблюдается высокая частота полиморфизма в генах, входящих в состав данного кластера (Таблица 20).

7. У всех резистентных к пенициллину штаммов *S.pneumoniae*, являющихся однолокусными вариантами ST236, мы обнаружили мутации в белке RegR, который относится к регуляторам транскрипции, принадлежащим к семейству LacI/GalR. Ген *regR* включен в 15.5-kb область, содержащую открытые рамки считывания, кодирующие гипотетическую глутатион пероксидазу (*gpx*), гиалуронидазу (*hyl*), белки, участвующие в метаболизме 2-кето-3-дезоксиглюконата (*kdg* оперон), белки углевод-фосфотрансферазной транспортной системы (PTS оперон), а также гипотетические белки.

### 3.3. Характеристика геномных локусов изолятов *S. pneumoniae* эпидемических генетических линий, детерминирующих серотиповую принадлежность и устойчивость к бета-лактамам антибиотикам

Согласно описанным выше данным, наличие резистентности к антибиотикам различных классов ассоциируется с принадлежностью изолята к генетической линии и серотипу (см. Раздел 3.2.1.). Эти данные согласуются с результатами других исследований [334]. Предполагают, что детерминанты резистентности приобретаются генетическими линиями, представители которых имеют толстые капсулы, ассоциированные со специфическими серотипами [334]. Для подтверждения данного факта был проведен анализ рекомбинаций для изолятов генетических линий, характеризующихся мультирезистентностью к антибиотикам, – для представителей CC320, распространенных в России, а также для представителей CC90, распространенных в странах Европы. Схема анализа рекомбинаций для изолятов отдельных генетических линий приведена на Рисунке 35.

Поиск сайтов рекомбинаций для каждой генетической линии отдельно	
Выравнивание геномов	Genome Comparator
Построение филогенетического дерева	RaxML
Идентификация сайтов рекомбинаций	Gubbins
Визуализация результатов	Phandango

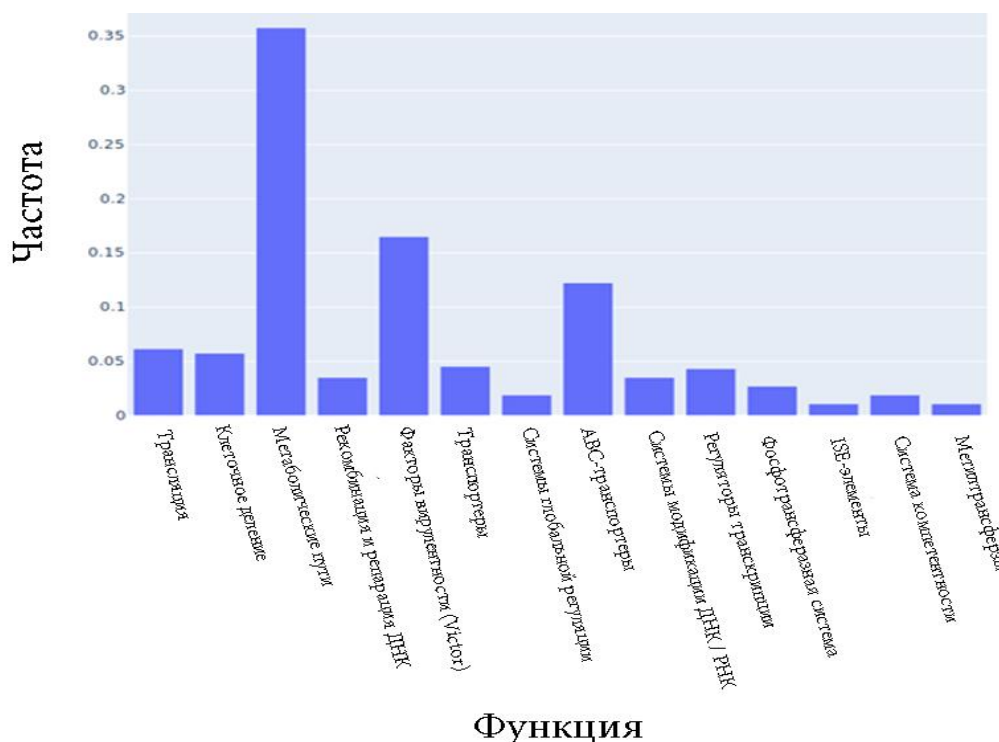
  

Аннотация рекомбинаций и анализ ассоциаций рекомбинаций в <i>cps</i> -локусе и генах пенициллин-связывающих белков <i>pbp2x</i> , <i>pbp2b</i> , <i>pbp1a</i>
---

**Рисунок 35 – Схема анализа рекомбинаций в отдельных генетических линиях**

События рекомбинаций были идентифицированы с помощью программы Gubbins отдельно для основных генетических линий, ассоциированных с резистентностью к

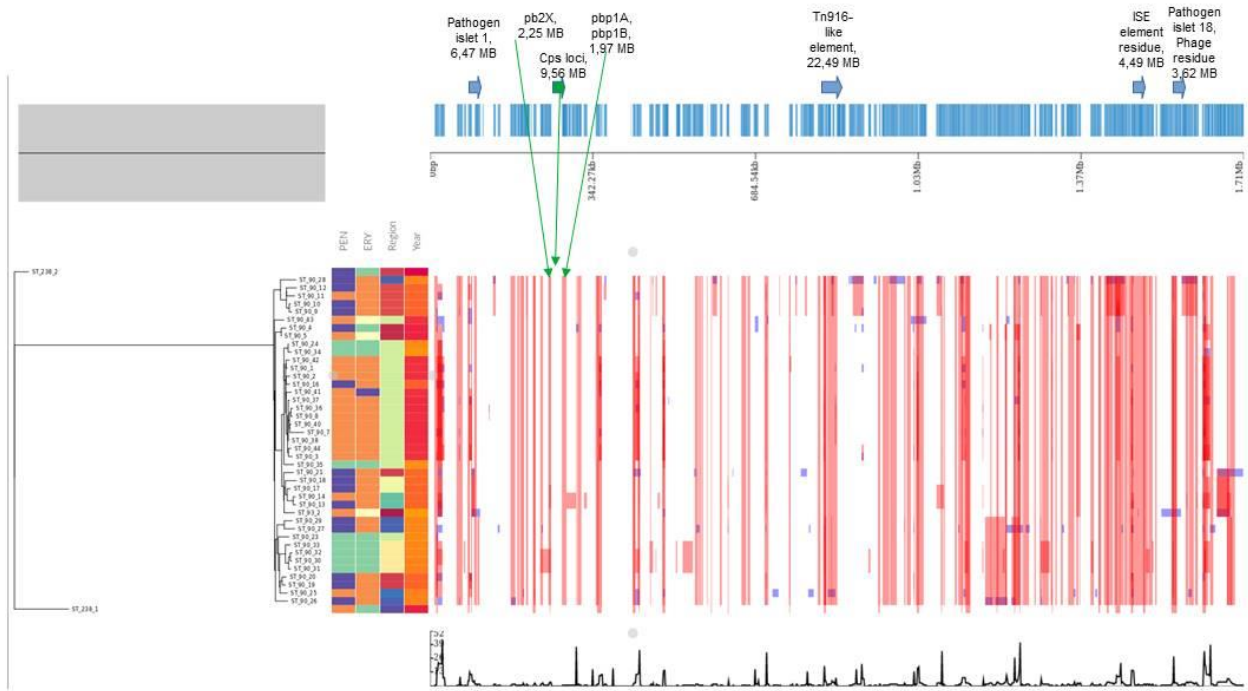
бета-лактамам (CC90, CC236/CC271/CC320). Мультирезистентные генетические линии характеризовались наличием большого числа предковых клонально-ассоциированных рекомбинаций в генах, ассоциированных с различными метаболическими путями, а также кодирующих факторы вирулентности, белки клеточного деления, компоненты фосфотрансферазной системы, транспортеры, ферменты модификации нуклеиновых кислот, регуляторы транскрипции и трансляции, участвующие в репарации и гомологичной рекомбинации ДНК белки (Рисунок 36).



**Рисунок 36 – Гистограмма ассоциации частоты рекомбинационных событий и затронутых локусов в генетической линии CC90**

В генетической линии CC90, ассоциированной с мультирезистентностью, рекомбинации ассоциировались с генами-мишенями бета-лактамных антибиотиков, при этом гены *crs*-локуса не участвовали в рекомбинациях (Рисунок 37).

Большинство сайтов рекомбинаций в изолятах генетической линии CC236/CC271/CC320 также имели клональное происхождение и включали сайты, соответствующие генам, идентифицированным в Разделе 3.2.3 («Значимые однонуклеотидные полиморфизмы, ассоциированные с резистентностью к бета-лактамным антибиотикам у изолятов CC320»). При этом гены *crs*-локуса не участвовали в рекомбинациях (Рисунок 38).

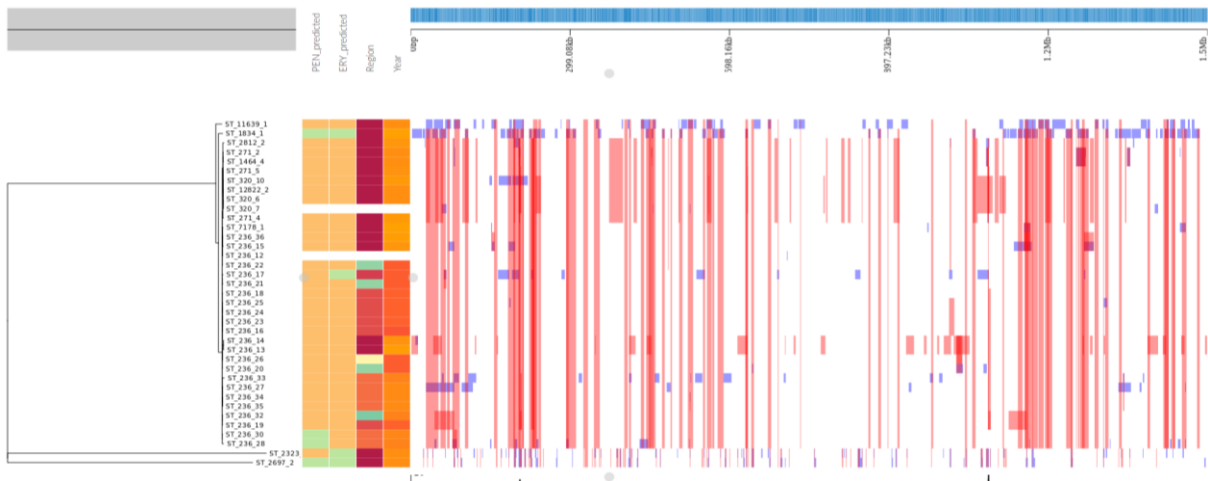


**Рисунок 37 – Карта рекомбинационных событий для генетической линии CC90**

Рекомбинации были идентифицированы с помощью программы Gubbins.

Красными блоками показаны предковые клональные рекомбинации, ассоциированные с внутренними узлами. Синими блоками показаны рекомбинационные события, ассоциированные с конкретными штаммами.

Слева показано дерево, построенное с помощью RaхML для изолятов соответствующей генетической линии (по выравниванию коровой части геномов в GenomeComparator). Цветными кодами аннотированы серотипы, резистентность к бета-лактамам (PEN\_predicted – все штаммы резистентные), резистентность к макролидам (ERY\_predicted), регион циркуляции штамма и год выделения изолята.



**Рисунок 38. Карта рекомбинационных событий для генетической линии CC236 / CC271 / CC320**

Рекомбинации были идентифицированы с помощью программы Gubbins.

Красными блоками показаны предковые клональные рекомбинации, ассоциированные с внутренними узлами. Синими блоками показаны рекомбинационные события, ассоциированные с конкретными штаммами.

Слева показано дерево, построенное с помощью RaхML для изолятов соответствующей генетической линии (по выравниванию коровой части геномов в GenomeComparator). Цветными кодами аннотированы серотипы, резистентность к бета-лактамам (PEN\_predicted – все штаммы резистентные), резистентность к макролидам (ERY\_predicted), регион циркуляции штамма и год выделения изолята.

**Выводы по разделу:**

1. Резистентные к бета-лактамам антибиотикам генетические линии ассоциируются с высокой частотой рекомбинаций, имеющих клональное предковое происхождение.
2. Рекомбинации в генах, кодирующие первичные мишени бета-лактамов антибиотиков, *Pbp1A*, *Pbp2B* и *Pbp2X*, ассоциируются с множеством адаптивных рекомбинаций, в частности, в генах, кодирующих белки клеточного деления и других. При этом не наблюдается ассоциированных рекомбинаций в *cps*-локусе.
3. Ассоциация резистентности к бета-лактамам антибиотикам с серотипами может быть связана с высокой частотой рекомбинаций, отмеченной для серотипов 23F, 19F, 19A, 35B, 6A, 6E.

### 3.4. Сравнительный анализ генов пенициллин-связывающих белков изолятов *S. pneumoniae*, относящихся к эпидемическим генетическим линиям

Был проведен филогенетический анализ генов пенициллин-связывающих белков, *pbp2x*, *pbp2b* и *pbp1a*, являющихся первичными мишенями для бета-лактамов антибиотиков, с целью идентификации в них рекомбинационных событий и установления возможного донора детерминант резистентности в анализируемой популяции пневмококков.

Схема анализа приведена на Рисунке 39.

Для генов <i>pbp2x</i> , <i>pbp2b</i> , <i>pbp1a</i> экстрагированы участки, кодирующие области транспептидазного домена		
Выборка дополнена соответствующими фрагментами других видов стрептококков (из Генбанка)		
	Выравнивание участков, кодирующих транспептидазные домены	Muscle
	Филогенетическое дерево по выравниваниям	RaxML
	Идентификация сайтов рекомбинаций	BratNextGen
	Визуализация результатов	Phandango
Анализ ассоциаций рекомбинаций в генах пенициллин-связывающих белков <i>pbp2x</i> , <i>pbp2b</i> , <i>pbp1a</i>		

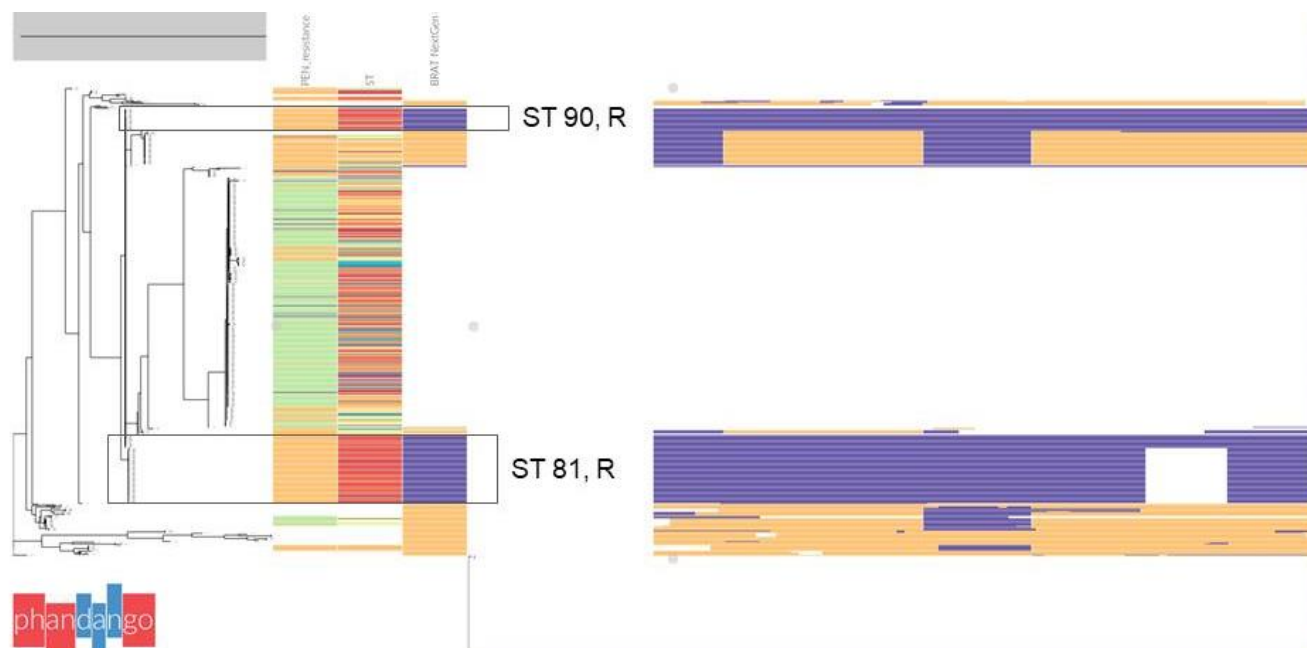
**Рисунок 39 – Схема анализа рекомбинаций в генах пенициллин-связывающих белков**

В качестве референсных последовательностей были использованы последовательности областей транспептидазных доменов соответствующих генов, принадлежащих стрептококкам других видов. Референсные последовательности были

загружены из Генбанка по следующим ID-номерам: KP732106, KP732107, KP732108, KP732109, KP732110, KP732111, KP732112, KP732113, KP732114, KP732115, KP732116, KP732117, KP732118, KP732119, KP732120, KP732121, KP732122, KP732123, KP732124, KP732125, KP732126, KP732127, KP732128, KP732129, KP732130, KP732131, KP732132, KP732133, KP732134, KP732135, KP732136, KP732137, KP732138, KP732139, KP732140, KP732141, KP732142, KP732143, KP732144, KP732145, KP732146, KP732147, KP732148, KP732149, KP732150, KP732151, KP732152, KP732153, KP732154, KP732155, KP732156, KP732157, KP732158, KP732159, KP732160, KP732161, KP732162, KP732163, KP732164, KP732165, KP732166, KP732167, KP732168, KP984507, KP984508, KP984509, KP984510, KP984511, KP984512, KP984513, KP984514, KP984515, KP984516.

Анализ результатов кластеризации изолятов по последовательностям гена *pbp1a* позволил выявить два больших кластера последовательностей, рекомбинирующих между собой и включающих в себя в основном сиквенс-типы ST81 и ST90 с одной стороны, а также ST236, ST230 и ST1094 с другой (Рисунок 40).

У штаммов ST81 идентифицировано несколько рекомбинационных событий в области транспептидазного домена гена *pbp1a*. При первом событии рекомбинации донором был *S. mitis*. Помимо общего с PMEN1 (ST81) блока, изоляты ST236 и их родственные штаммы имеют следы событий независимых рекомбинационных событий.

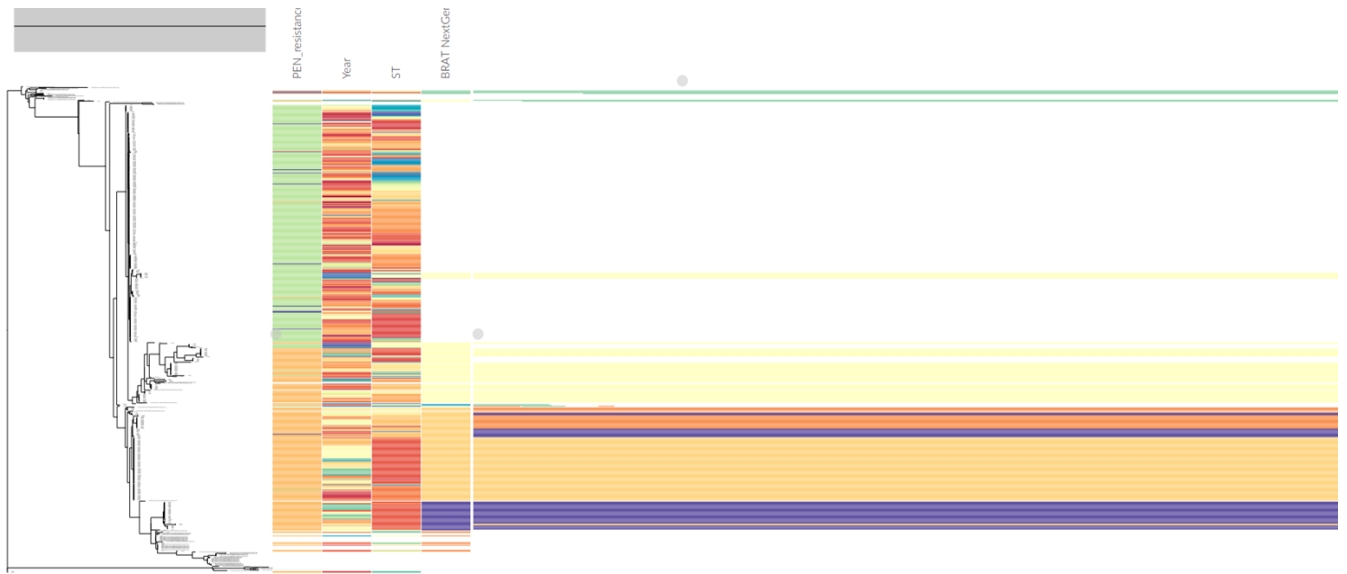


**Рисунок 40 – Карта рекомбинационных событий в последовательностях транспептидазного домена гена *pbp1a***

События горизонтального переноса, идентифицированные с помощью BratNEXTGEN между *S.pneumonia* ST90 и ST81, а также *S.pneumonia* ST230, ST236, ST1094 и *S.mitis*.



Анализ результатов кластеризации изолятов по последовательностям гена *pbp2b* показал наличие семи кластеров, ассоциированных с резистентностью к пенициллину (Рисунок 41). Кластер 5 представлен изолятами ST90, ST63, ST315, ST230, ST199, ST1094 (а также редкие чувствительные изоляты). Кластер 6 представлен изолятами ST376, ST344, ST236, ST81, ST156. Кластер 1 представлен изолятами ST90. Предположение о том, кто выступал возможным донором, сделать сложно.

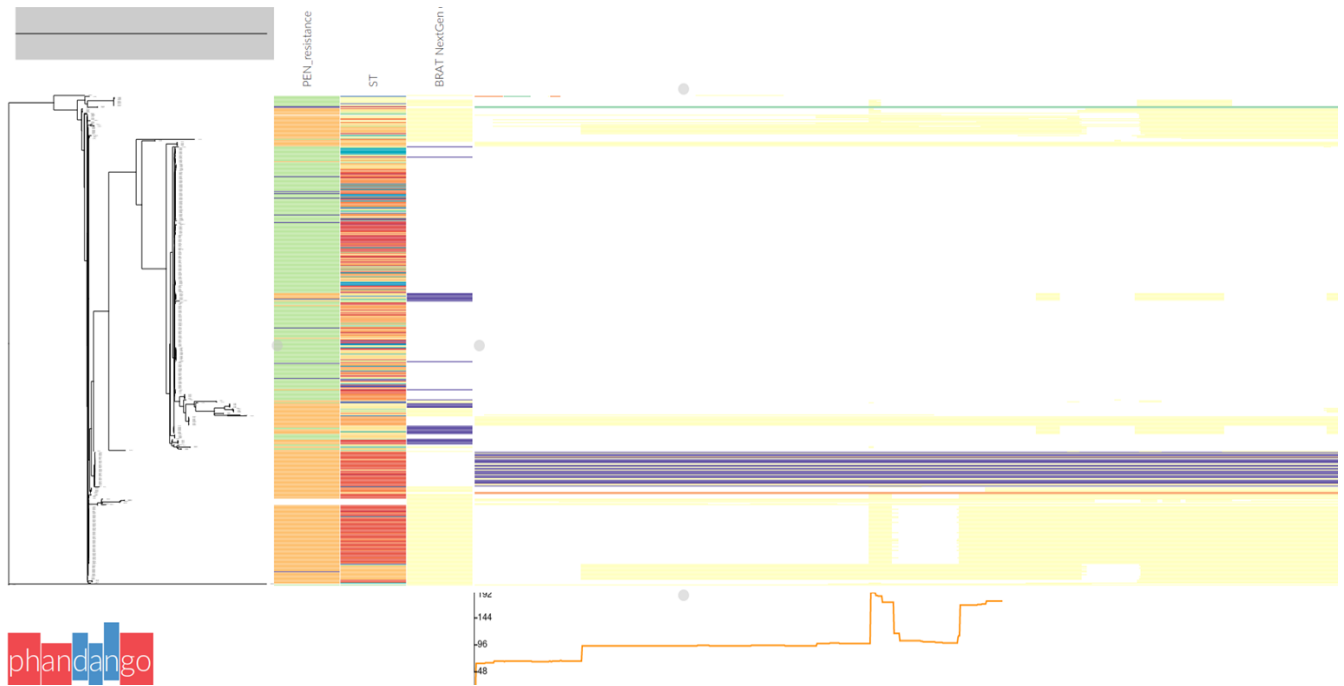


**Рисунок 41 - Карта рекомбинационных событий в последовательностях транспептидазного домена гена *pbp2b* (BratNextGen)**

События горизонтального переноса, идентифицированные с помощью BratNEXTGEN.

Анализ результатов кластеризации изолятов по последовательностям гена *pbp2x* позволил выявить два больших кластера, коррелирующих с антибиотикорезистентностью (Рисунок 42). Кластер 1 включал ST289, ST315, ST344; Кластер 2 включал ST81, ST90, ST153 и *S. mitis*. При этом интересно, что некоторые изоляты ST90, близкие ко второму кластеру, имеют регионы рекомбинации первого кластера.

Наблюдаются рекомбинации у отдельных резистентных к пенициллину изолятов ST199, хотя большинство представителей генетической линии ST199 являются чувствительными к бета-лактамам антибиотикам.



**Рисунок 42 – Карта рекомбинационных событий в последовательностях транспептидазного домена гена *pbp2x* (BratNextGen)**  
События горизонтального переноса, идентифицированные с помощью BratNEXTGEN.

### Выводы по разделу:

1. Полученные результаты подтверждают данные предыдущих исследований о событиях рекомбинации между *S. mitis* и *S. pneumoniae*.
2. Донорами для рекомбинаций в генах *pbp1a* и *pbp2x* являлись представители других стрептококков.
3. Изоляты CC236/CC271/CC320 в гене *pbp1a* имеют также следы независимых рекомбинаций.
4. Изоляты CC90 в гене *pbp2x* имеют также следы независимых рекомбинаций.
5. Для гена *pbp2b* сложно сделать предположение о том, кто выступал возможным донором.

### 3.5. Гены вирулентности, продукты которых могут быть потенциальными мишенями для создания белковой антипневмококковой вакцины

Анализируемые изоляты хорошо кластеризовались по групповой переменной «Источник выделения» в зависимости от вариантов генов (Раздел 3.1.3 «Оценка вклада генов ядерного генома в формирование структуры популяции *S. pneumoniae* (групп А/В1/В2, сиквенс-кластеров, серотипов, инвазивных штаммов резистентных к антибиотикам штаммов)»). По 214 генам (Том 2 – Приложения. Приложение 5),

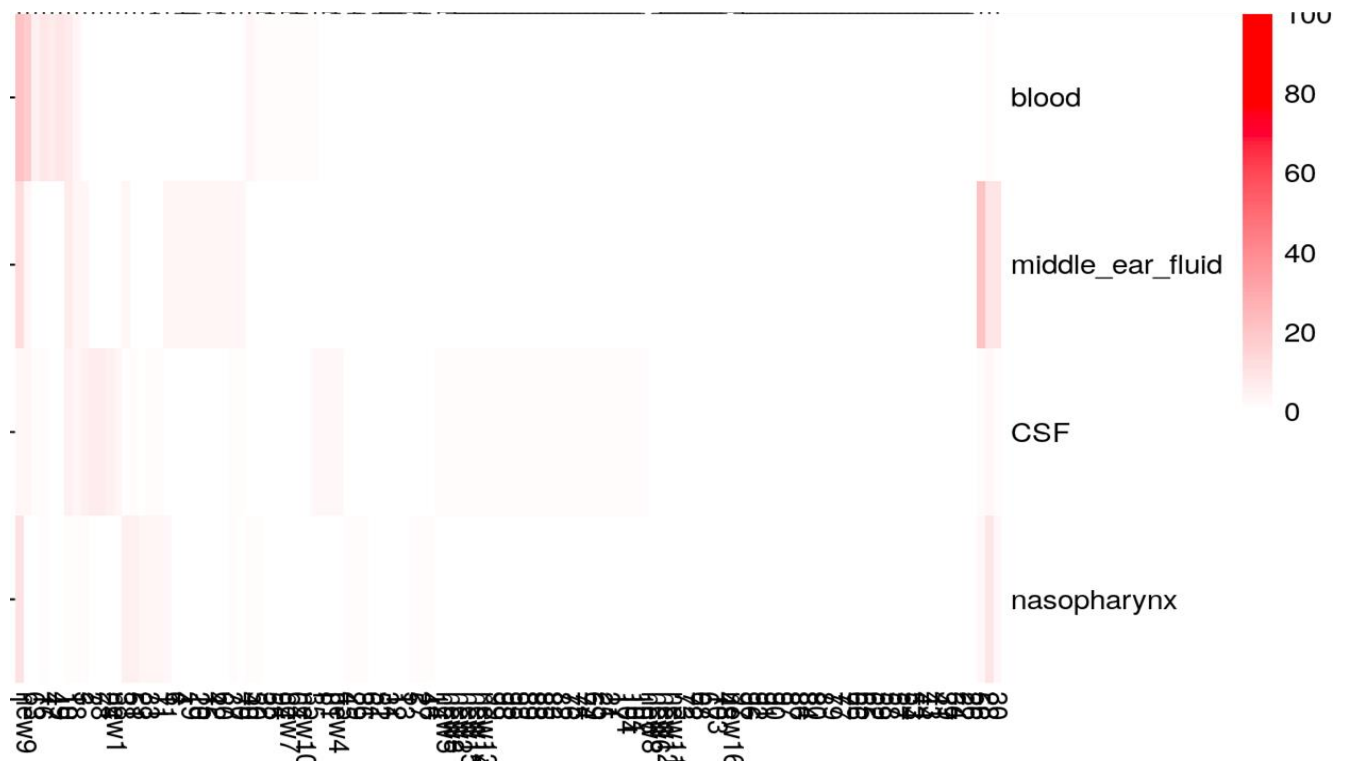
характеризующимся высоким качеством вариативности ( $\cos 2 > 0,5$ ), была получена наименьшая ошибка предсказания по источнику выделения (4,28% - ошибка предсказания изолятов, выделенных из носоглотки от бессимптомных носителей, Таблица 12). К наиболее значимым генам, вносящим значительный вклад в формирование групп инвазивных изолятов и кодирующим поверхностные и мембранные белки, относились: *strH*, *nanA*, *phts*, *nanB*, *oppC*, *pcpA*, *wzd* (участвует в синтезе капсулы).

Ген *strH* показал максимальную ассоциацию с инвазивностью и источником выделения (кровь, ликвор, жидкость среднего уха, носоглотка). С источником выделения «кровь» ассоциируется ограниченное число вариантов гена *strH* и генетических линий (Рисунок 43). Этот ген кодирует экспрессируемую на поверхности клеточной стенки экзо- $\beta$ -D-N-ацетилглюкозаминидазу, участвующую в деградации N-связанных гликанов. Белок StrH экспрессируется вместе с другими важнейшими факторами вирулентности: ферментами NanA (Нейроминидаза А), GH2 (BgaA, бета-галактозидаза А), GH85 (EndoD) и GH125. Уровень экспрессии StrH зависит от условий окружения, в котором находится пневмококк. Экспрессия белка регулируется уровнем углеводов в среде. В условиях дефицита углеводов экспрессия StrH повышается и пневмококк начинает разрушать ткани хозяина.

Выше было показано, что особенности метаболизма пневмококка детерминируют его значительный вирулентный потенциал и белок StrH оказывается на стыке метаболических путей. Так, расщепление популяции по вариантам гена *strH* коррелирует с расщеплением пневмококков по компонентам углеводного метаболизма (*gnd*, *dexB*), путям синтеза ароматических аминокислот (*aroE*), регуляторному гену *relA* (ГТФ-пирофосфокиназа), генам синтеза пептидогликана и клеточного деления (*murD*, *pbp1A*), регуляторным генам экспрессии полисахаридной капсулы (*wzg-wzh-wze*) (Раздел 3.1.3.5 «Взаимодействие генов, участвующих в делении популяции на группы A/B1/B2, SC\_MLST и серотипы, инвазивные, резистентные и чувствительные к антибиотикам штаммы»).

Известно, что основной трудностью создания универсальных антипневмококковых белковых вакцин является вариабельность белков (что снижает эффективность таких вакцин). Однако варианты белка StrH могут быть использованы в

качестве одной из мишеней химерной белковой вакцины. Возможно, это позволит модулировать вирулентный потенциал пневмококка.



**Рисунок 43 – Гистограмма ассоциаций между вариантами гена *strH* и инвазивными свойствами изолята**

Источник выделения: blood - кровь, CSF - ликвор, nasopharynx - носоглотка / носитель, middle\_ear\_fluid – жидкость среднего уха

### Выводы по разделу:

1. Белок StrH, кодирующий экспрессируемую на поверхности клеточной стенки экзо-β-D-N-ацетилглюкозаминидазу, может быть использован в качестве одной из мишеней для белковой вакцины. Возможно, это позволит модулировать вирулентный потенциал пневмококка.
2. Необходимы дополнительные исследования для оценки антигенного потенциала StrH.

## ЗАКЛЮЧЕНИЕ И ВЫВОДЫ

Применение антипневмококковых конъюгированных вакцин оказывает влияние на структуру популяции *S. pneumoniae*. В мире проводилось несколько исследований по изучению эффекта антипневмококковой вакцинации с использованием данных полногеномного секвенирования пневмококка. В частности, были проанализированы популяции пневмококка до и после применения вакцин: в лагере для беженцев МаеЛа, Таиланд (3085 изолятов) [19, 344-347]; Саутгемптон [347], Великобритания (672 изолята); Массачусетс, США (616 изолятов) [115, 136, 347]; Блантайр, Малави (585 изолятов); Неймеген, Нидерланды (346 изолятов) [137, 303, 347]. В данных исследованиях использовался байесовский анализ структуры популяции, который позволяет получить надежную и специфическую кластеризацию набора данных. В опубликованных за последние 15 лет работах сообщалось, что генотип может ассоциироваться с инвазивностью для некоторых генетических линий пневмококка. Международная программа по изучению структуры популяции пневмококка, созданная в последние годы и использующая данные полногеномного секвенирования, позволила охарактеризовать и сравнить генетические линии пневмококка из разных стран, оценить международные данные по серотиповому составу, устойчивости к антибиотикам и географическому распространению [303]. В этом исследовании были установлены глобальные генетические линии пневмококка, их ассоциация с инвазивностью и резистентностью, приведены доказательства того, что генотип может влиять на инвазивность.

Для характеристики популяции *S. pneumoniae*, циркулирующих в России, в нашем исследовании были использованы дополнительные методы, которые позволили получить новую информацию. Удалось установить филогенетические взаимосвязи для редких сиквенс-типов, которыми представлена более половины популяции пневмококков в России. Благодаря этому была получена связанная филогенетическая картина. Установлена взаимосвязь циркулирующих в России генетических линий с распространенными в различное время в разных регионах эпидемически значимыми клонами. Выявлена тенденция к преимущественному распространению в России антибиотикорезистентности среди отдельных сиквенс-кластеров: SC2, SC4. Установлено деление глобальной популяции пневмококков на 3 группы. Был проведен анализ взаимосвязей между глобальными группами A, B1 и B2 и их субструктурами

(кладами дерева и ассоциирующимися с ними сиквенс-кластерами). Были выявлены и охарактеризованы 11 сиквенс-кластеров. Были идентифицированы гены, полиморфизм которых ассоциируется с развитием резистентности к бета-лактамам у *Streptococcus pneumoniae*.

Увеличение знаний о географическом распространении, распределении устойчивости к антибиотикам, существовании невакцинных серотипов и вкладе различных генетических линий в развитие инвазивных заболеваний обеспечивает полезный контекст для оценки воздействия вакцин. Полученные геномные данные предоставляют значительный ресурс для дальнейшего изучения биологии пневмококка и осуществления контроля над пневмококковой инфекцией.

### **Выводы:**

1. Мировая популяция *S. pneumoniae* представлена тремя глобальными группами А, В1 и В2. В настоящее время в популяции пневмококка эволюционирует генетически гетерогенная группа В, представители которой ассоциируются преимущественно с разными серотипами, но тем не менее, имеют характерные общие метаболические особенности. В группе А доминируют серотипы 23F, 19F, 14, 23A; в группе В1 – серотипы 11A, 19F, 19A, 1, 9N; в группе В2 – серотипы 6A/B/E, 3, 19A, 7F, 5. Группа В2 ассоциируется преимущественно с серотипами, не входящими в состав вакцины ПКВ7. Наибольший вклад в деление на глобальные группы вносят гены, кодирующие: сигнальную пептидазу I, участвующую в процессинге секретируемых белков; гены глюкокиназы и глюкозо-6-фосфат-1-дегидрогеназы, регулирующие поток глюкозы в клетку. Принадлежность к генетическим линиям определяется структурой базовых систем метаболизма: типа системы рестрикции-модификации ДНК; CiaH сенсорной гистидин-киназы (компонент двухкомпонентной сигнальной системы CiaHR, регулятор компетентности / вирулентности); АссС - ацетил-СоА-карбоксилазы (участвует в первых реакциях синтеза жирных кислот).
2. Происходившие в 2000 – 2010 гг. изменения в структуре популяции пневмококков, циркулировавших в Российской Федерации, были обусловлены изменениями в структуре глобальной популяции в ответ на массовую вакцинацию ПКВ7 в различных регионах мира. Происходивший в этот период рост устойчивости к бета-лактамам и макролидным антибиотикам связан с преимущественным глобальным распространением серотипа 19A, не входящего в состав ПКВ7.

3. Возможно, АТФ-синтазный комплекс и компоненты фосфотрансферазной системы детерминируют формирование типов капсулы и метаболический тип *agaD-1* и *atpG-1* является оптимальным для роста и адаптации большинства пневмококков. Серотип-специфичная вакцинация, приводящая к исключению из циркуляции распространенных генетических линий пневмококка, обладающих большей способностью к адаптации за счет энерго-обеспечивающих систем (АТФ, НАДФН, фосфотрансферазные системы), может приводить к распространению генетических линий, характеризующихся большей вирулентностью.
4. Наличие резистентности к антибиотикам различных классов ассоциируется с принадлежностью изолята к генетической линии и серотипу. Наши данные подтверждают данные других авторов о серотипах, ассоциирующихся с высокой частотой рекомбинаций (6А, 6Е, 23F, 19F, 14, 19А). Резистентные к бета-лактамам антибиотики изоляты ассоциируются с серотипами, отличающимися высокой эффективностью рекомбинаций. Рекомбинации в генах, кодирующих первичные мишени бета-лактамовых антибиотиков, *Pbp1A*, *Pbp2B* и *Pbp2X*, ассоциируются с множеством адаптивных рекомбинаций ядерного генома, в частности, в генах, кодирующих белки клеточного деления, и других. При этом не наблюдается ассоциированных рекомбинаций в *cps*-локусе. Мультирезистентные изоляты (к пенициллину и эритромицину) составляют 75,45% резистентных к пенициллину изолятов. Возможно, интеграция транспозонов Tn916 и Tn5253, обуславливающих резистентность к макролидам, так же как и рекомбинации, была более эффективной на этапе стабилизации генетических линий, у пневмококков, ассоциирующихся с серотипами 6А, 6Е, 23F, 19F, 14, 19А. Донорами для рекомбинаций в генах *pbp1a* и *pbp2x* являлись представители других стрептококков (*S. mitis* и *S. oralis*). Изоляты, принадлежащие к клонам СС236/СС271/СС320, в гене *pbp1a* имеют также следы независимых рекомбинаций. Изоляты СС90 в гене *pbp2x* имеют также следы независимых рекомбинаций. Для гена *pbp2b* сложно сделать предположение о том, кто выступал возможным донором.
5. Особенности метаболизма углеводов и ароматических аминокислот детерминируют регуляцию клеточных процессов пневмококка и фенотипические особенности, в том числе устойчивость к антибиотикам и вирулентность. Расщепление популяции по вариантам гена *strH* коррелирует с кластеризацией по компонентам углеводного

метаболизма (*gnd*, *dexB*), путем синтеза ароматических аминокислот (*aroE*), регуляторному гену *relA* (ГТФ-пирофосфокиназа), генам синтеза пептидогликана и клеточного деления (*murD*, *pbp1A*), регуляторным генам экспрессии полисахаридной капсулы (*wzg-wzh-wze*). Белок StrH, кодирующий экспрессируемую на поверхности клеточной стенки экзо- $\beta$ -D-N-ацетилглюкозаминидазу, оказывается на стыке метаболических путей и может быть использован как мишень для белковой вакцины. Возможно, это позволит модулировать вирулентный потенциал пневмококка. Необходимы дополнительные исследования для оценки антигенного потенциала StrH.

#### **Рекомендации по использованию результатов исследования**

1. Проведенные исследования имеют теоретическое значение, так как дополняют и уточняют существующие представления об эволюции *S. pneumoniae*, его метаболических особенностях и возможностях адаптации.
2. Полученные результаты позволят в дальнейшем оценивать изменения в ходе дальнейшего мониторинга эпидемиологической ситуации в России, а также анализировать генетические процессы, детерминирующие ответ популяции пневмококка на методы профилактики и лечения пневмококковых заболеваний.
3. Подходы филогенетического анализа могут быть использованы в практической и научной работе.
4. Результаты использования предложенной методики исследования могут быть использованы при чтении лекций и проведении практических занятий.



### Список публикаций по теме диссертации

#### Статьи, опубликованные в изданиях, рекомендованных ВАК МОН РФ:

1. **Цветкова, И. А.** SNP-полиморфизм в геномах изолятов *Streptococcus pneumoniae* CC320, устойчивых к бета-лактамам антибиотикам. / Цветкова И. А., Волкова, М. О., Калиногорская О. С. [и др.]. // Антибиотики и химиотерапия. - 2016. - Т. 61, № 11-12. - С. 11-12. – URL: <https://www.antibiotics-chemotherapy.ru/jour/article/view/694>.
2. **Цветкова, И. А.** Клональная структура популяции изолятов *Streptococcus pneumoniae*, циркулирующих в России с 1980 по 2017 гг. / Цветкова И. А., Беланов С. С., Гостев В. В. [и др.]. // Антибиотики и химиотерапия. - 2019. - Т. 64, № 5-6. - С. 22-31. – doi: 10.24411/0235-2990-2019-100027.
3. **Sidorenko, S.** Multicenter study of serotype distribution of *Streptococcus pneumoniae* nasopharyngeal isolates from healthy children in the Russian Federation after introduction of PCV13 into the National Vaccination Calendar. / Sidorenko S., Rennert W., Lobzin Y. [et al]. // Diagnostic Microbiology and Infectious Disease. - 2020. V. 96, №1. - P. 114914. – doi: 10.1016/J.DIAGMICROBIO.2019.114914. – JCR=2.499.

#### Публикации в других изданиях, тезисы докладов:

1. **Belanov, S.** Prevalence of PI-I and PI-2 among *Streptococcus pneumoniae* isolated from healthy nasopharyngeal carriers and children with acute otitis media. / S. Belanov, I. Tsvetkova, S. Sidorenko // ECCMID-XXIV, Barcelona, Spain, 10-13.05.2014. - P1446.
2. **Belanov, S.** Evaluation of the capsular genes of *Streptococcus pneumoniae* serotypes 19A/F, isolated from children with pneumococcal infections: bioinformatic assay. / S. Belanov, I. Tsvetkova, V. Gostev, S. Sidorenko. // ECCMID-XXV, Copenhagen, Denmark, 25-28.04.2015. - P0178.
3. **Tsvetkova, I.** *Streptococcus pneumoniae* clonal complex 320 – evolutionary relationships with other genetic lineages. / I. Tsvetkova, V. Gostev, S. Belanov, S. Sidorenko. // The 10th International Symposium on Pneumococci and Pneumococcal Diseases (ISPPD 2018), Glasgow, Scotland, 26–30.06.2016. - ISPPD-0698.
4. **Mokhov, A.** Quantitative detection of *Streptococcus pneumoniae* in nasopharyngeal fluid using droplet digital PCR method. / Mokhov A., Sopova J. V., Tsvetkova I. [et al]. // ECCMID-XXVII, Vienna, Austria, 22-25.04.2017. - EP0941.
5. **Tsvetkova, I.** Polymorphism in the genomes of the CC320 isolates of *Streptococcus pneumoniae* resistant to beta-lactam antibiotics. / Tsvetkova I., Volkova M.,

- Kalinogorskaya O. [et al]. // 7th European Conference on Prokaryotic and Fungal Genomics (ProkaGENOMICS), 19-22.09.2017, Göttingen, Germany. - Poster ID83. URL: <https://fems-microbiology.org/opportunities/prokagenomics-2017-7th-european-conference-prokaryotic-fungal-genomics/>
6. **Цветкова, И. А.** Клональная структура популяции *Streptococcus pneumoniae*, циркулирующих в Санкт-Петербурге и распространение резистентности к антибактериальным препаратам. / Цветкова И. А., Беланов С. С., Гостев В. В. [и др.]. // Проблемы медицинской микологии. - 2018. - Т. 20, №2 - С. 126.
  7. **Никитина, Е. В.** Динамика серотипового состава *Streptococcus pneumoniae* у детей в г. Санкт-Петербурге на фоне вакцинации. / Никитина Е. В., Мохов А. С., Цветкова И. А. [и др.]. // Проблемы медицинской микологии. – 2018. - Т. 20, №2 - С. 98.
  8. **Цветкова, И. А.** Полногеномный анализ *Streptococcus pneumoniae*, выделенных от пациентов с менингитами. / И. А. Цветкова // 4-5я Российская научно-практическая конференция «Актуальные проблемы менингококковой инфекции и гнойных бактериальных менингитов», 14-15 ноября 2018 г. - Москва, 2018. - Доклад. URL: <https://expodata.info/2018/11/14/vi-rossiyskaya-nauchno-prakticheskaya-konf/>
  9. **Tsvetkova, I.** Antimicrobial Resistance and Clonality of *Streptococcus pneumoniae* Isolates in Russia. / Tsvetkova I., Belanov S., Gostev V. [et al]. // Bioinformatics: from Algorithms to Applications 2019 (BiATA 2019), 20-22.06.2019, Saint-Petersburg, Russia. - Доклад. URL: <http://biata2019.spbu.ru/program/>
  10. **Tsvetkova, I.** Antimicrobial resistance and clonality of *Streptococcus pneumoniae* isolates in Russia. / Tsvetkova I., Belanov S., Gostev V. [et al]. // 2-й Российский микробиологический конгресс, 23-27 сентября, 2019, Саранск. Материалы конгресса. - С. 169.
  11. **Муравьев, А. А.** Эпидемиология серотипов *S. pneumoniae*, выделенных у лиц старше 18 лет: здоровых носителей, пациентов с острым средним отитом, внебольничной пневмонией и инвазивной пневмококковой инфекцией (исследование «SPECTRUM»). / Муравьев А. А., Чагарян А. Н., Иванчик Н. В. [и др.]. // Клиническая микробиология и антимикробная химиотерапия. - 2019. - Т. 21, № 4. - С. 275-281. – URL: <https://cslide.ctimeetingtech.com/isppd20/attendee/eposter/poster/654>.
  12. **Tsvetkova, I. A.** Mechanisms of virulence regulation and global distribution of vaccine candidate antigens in the high virulent *Streptococcus pneumoniae* strains. / Tsvetkova I. A.,

Likholetova D. V., Gostev V. V. [et al]. // 12th International Symposium on Pneumococci and Pneumococcal Diseases (ISPPD-12), 21-25.06.2020, Toronto, Canada. – P654. URL: <https://cslide.ctimeetingtech.com/isppd20/attendee/eposter/poster/654>

**Список используемых сокращений и обозначений**

ИКЭ – Интегративный конъюгативный элемент

МГЭ – Мобильный генетический элемент

ПИ – Пневмококковая инфекция

ПКВ7 – Пневмококковая конъюгированная 7-валентная вакцина

ПКВ10 - Пневмококковая конъюгированная 10-валентная вакцина

ПКВ13 – Пневмококковая конъюгированная 13-валентная вакцина

П.н. – Пары нуклеотидов

ПЦР – Полимеразная цепная реакция

СРМ – Система рестрикции-модификации

СС – Clonal complex (Клональный комплекс)

eBURST – адаптированный алгоритм BURST для оценки диверсификации клональных комплексов, разработанный для анализа больших объемов данных

ECDC - The European Centre for Disease Prevention and Control (Европейский центр по контролю и профилактике заболеваний)

GI – Genome islands (Островки патогенности)

MLST – Multilocus sequence typing (Мультилокусное сиквенс-типирование)

PMEN – Pneumococcal Molecular Epidemiology Network (Пневмококковая молекулярно-эпидемиологическая сеть)

PPI-1 - Pneumococcal pathogenicity island 1 (Пневмококковый островок патогенности 1)

PubMLST – публичная база данных для молекулярного типирования и анализа геномного разнообразия микроорганизмов (один из подходов – по данным Multi Locus Sequence Typing, MLST)

RAST - Rapid Annotations using Subsystems Technology (Быстрое аннотирование с использованием технологии подсистем)

SC Sequence cluster (Сиквенс-кластер)

SC\_MLST – Сиквенс-кластер, идентифицированный по конктенатам генов MLST

SLV – Single locus variant (Однолокусный вариант)

SNPs – Single nucleotide polymorphisms (Однонуклеотидные полиморфизмы)

ST – Sequence type (Сиквенс-тип)

### Список использованной литературы

1. Centers for Disease Control and Prevention. Active Bacterial Core Surveillance Report, Emerging Infections Program Network, *Streptococcus pneumoniae*, 2018 / CDC // <http://www.cdc.gov> : Медицина. Медицинский сайт: 2018. – URL: <http://www.cdc.gov/abcs/reportsfindings/survreports/spneu18.html> (дата обращения: 20.08.2021)
2. European Centre for Disease Prevention and Control. Invasive pneumococcal disease. In: ECDC. Annual epidemiological report for 2017. Stockholm: ECDC; 2019 / ECDC // <https://www.ecdc.europa.eu> : Медицина. Медицинский сайт: 2019. – URL: <https://www.ecdc.europa.eu/en/publications-data/invasive-pneumococcal-disease-annual-epidemiological-report-2017> (дата обращения: 20.08.2021).
3. **Сидоренко, С. В.** Популяционная структура пневмококков со сниженной чувствительностью к пенициллину и перспективы антипневмококковой вакцинации для сдерживания распространения антибактериальной резистентности. / С. В. Сидоренко, Т. А. Савинова, Е. М. Ильина, М. А. Сырочкина // Антибиотики и химиотерапия. – 2011. – Т. 5. – С. 11-18.
4. **Королева, И. С.** Серотиповая характеристика пневмококков, выделенных от больных пневмококковым менингитом. / И. С. Королева, Г. В. Белошицкий, К. О. Миронов. // Вопросы современной педиатрии. – 2012. – Т. 11, № 4. – С. 122-127.
5. **Калиногорская, О. С.** Антибиотикорезистентность и серотиповый состав *Streptococcus pneumoniae*, выделенных у детей в Санкт-Петербурге в 2010-2013 гг. / Калиногорская О. С., Беланов С. С., Волкова М. О. [и др.]. // Антибиотики и химиотерапия. – 2015. – Т. 60, № 1-2. – С. 10.
6. **Ubukata, K.** Serotype changes and drug resistance in invasive pneumococcal diseases in adults after vaccinations in children, Japan, 2010–2013. / Ubukata K., Chiba N., Hanada S. [et al.]. // Emerging Infectious Diseases. – 2015. – V. 21, № 11. – P. 1956-1965.
7. **Griffin, M. R.** U.S. Hospitalizations for pneumonia after a decade of pneumococcal vaccination. / Griffin M. R., Zhu Y., Moore M. R. [et al.]. // New England Journal of Medicine. – 2013. – V. 369, № 7. – P. 683-683.
8. **Маянский, Н. А.** Антибиотикорезистентность и клональная эволюция *Streptococcus pneumoniae* серотипа 19А в России, 2002-2013 гг. / Маянский Н. А., Савинова Т. А.,

- Алябьева Н. М. [и др.]. // Клиническая микробиология и антимикробная химиотерапия. – 2017. – Т. 19, № 2. – С. 145-151.
9. **Sidorenko, S.** Multicenter study of serotype distribution of *Streptococcus pneumoniae* nasopharyngeal isolates from healthy children in the Russian Federation after introduction of PCV13 into the National Vaccination Calendar. / Sidorenko S., Rennert W., Lobzin Y. [et al]. // Diagnostic microbiology and infectious disease. – 2020. – V. 96, № 1. – P.114914.
10. **Муравьев, А. А.** Эпидемиология серотипов *S. pneumoniae*, выделенных у лиц старше 18 лет: здоровых носителей, пациентов с острым средним отитом, внебольничной пневмонией и инвазивной пневмококковой инфекцией (исследование «SPECTRUM»). / Муравьев А. А., Чагарян А. Н., Иванчик Н. В. [и др.]. // Клиническая микробиология и антимикробная химиотерапия. – 2019. – Т. 21, № 4. – С. 275-281.
11. **Баранов, А. А.** Роль *Streptococcus pneumoniae* в структуре бактериальных инфекций у детей, госпитализированных в стационары г. Москвы в 2011–2012 гг. / Баранов А. А., Намазова-Баранова Л. С., Маянский Н. А. [и др.]. // Педиатрическая фармакология. – 2013. – Т. 10, № 5. – С. 6-12.
12. **Sandgren, A.** Effect of clonal and serotype-specific properties on the invasive capacity of *Streptococcus pneumoniae*. / Sandgren A., Sjostrom K., Olsson-Liljequist B. [et al]. // J Infect Dis. – 2004. – V. 189, № 5. – P. 785-796.
13. **Pletz, M. W. R.** Levofloxacin-resistant invasive *Streptococcus pneumoniae* in the United States: evidence for clonal spread and the impact of conjugate pneumococcal vaccine. / Pletz M. W. R., McGee L., Jorgensen J. [et al]. // Antimicrobial agents and chemotherapy. – 2004. – V. 48, № 9. – P. 3491-3497.
14. **Reinert, R. R.** Molecular epidemiology of macrolide-resistant *Streptococcus pneumoniae* isolates in Europe. / Reinert R. R., Ringelstein A., van der Linden M. [et al]. // Journal of Clinical Microbiology. – 2005. – V. 43, № 3. – P. 1294-1300.
15. **Riley, L. W.** Molecular epidemiology of infectious diseases: principles and practices. / Riley L. W. – Washington, DC: ASM Press, 2004. – 350 pp.
16. **Klugman, K. P.** The successful clone: the vector of dissemination of resistance in *Streptococcus pneumoniae*. / K. P. Klugman // The Journal of antimicrobial chemotherapy. – 2002. – V. 50 Suppl S2. – P. 1-5.

17. **Beall, B. W.** Shifting genetic structure of invasive serotype 19A pneumococci in the United States. / Beall B. W., Gertz R. E., Hulkower R. L. [et al]. // *J Infect Dis.* – 2011. – V. 203. – P. 1360–1368.
18. **Brown, J.** *Streptococcus Pneumoniae*: Molecular mechanisms of host-pathogen interactions. / Brown J., Hammerschmidt S., Orihuela C. [et al] – Academic Press, 2015. – 470 pp.
19. **Croucher, N. J.** Evidence for soft selective sweeps in the evolution of pneumococcal multidrug resistance and vaccine escape. / Croucher N. J., Chewappreecha C., Hanage W. P. [et al]. // *Genome Biol Evol.* – 2014. – V. 6, № 7. – P. 1589-602.
20. **Watkins, E. R.** Vaccination drives changes in metabolic and virulence profiles of *Streptococcus pneumoniae*. / Watkins E. R., Penman B. S., Lourenço J. [et al]. // *PLoS Pathogens.* – 2015. – V. 11, № 7. – P. e1005034.
21. **Brueggemann, A. B.** Clonal relationships between invasive and carriage *Streptococcus pneumoniae* and serotype- and clone-specific differences in invasive disease potential. / Brueggemann A. B., Griffiths D. T., Meats E. [et al]. // *The Journal of Infectious Diseases.* – 2003. – V. 187, № 9. – P. 1424-1432.
22. **Brueggemann, A. B.** Geographic distribution and clonal diversity of *Streptococcus pneumoniae* serotype 1 isolates. / A. B. Brueggemann, B. G. Spratt // *Journal of clinical microbiology.* – 2003. – V. 41, № 11. – P. 4966-4970.
23. **Kawamura, Y.** Determination of 16s rRNA sequences of *Streptococcus mitis* and *Streptococcus gordonii* and phylogenetic relationships among members of the genus *Streptococcus*. / Kawamura Y., Hou X. G., Sultana F. [et al]. // *International Journal of Systematic Bacteriol.* – 1995. – V. 45, № 2. – P. 406-408.
24. **Vinderola, G.** Lactic acid bacteria: microbiological and functional aspects. / Vinderola G., Ouwehand A., Salminen S., von Wright A. [et al]. – CRC Press. – 764 pp.
25. **Hoshino, T.** Use of phylogenetic and phenotypic analyses to identify nonhemolytic *Streptococci* isolated from bacteremic patients. / T. Hoshino, T Fujiwara, M. Kilian. // *Journal of Clinical Microbiology.* – 2005. – V. 43, № 12. – P. 6073-6085.
26. **Kilian, M.** Evolution of *Streptococcus pneumoniae* and its close commensal relatives. / Kilian M., Poulsen K., Blomqvist T. [et al]. // *PloS one.* – 2008. – V. 3, № 7. – P. e2683.

27. **Kilian, M.** Parallel evolution of *Streptococcus pneumoniae* and *Streptococcus mitis* to pathogenic and mutualistic lifestyles. / Kilian M., Riley D. R., Jensen A. [et al]. // mBio. – 2014. – V. 5, № 4. – P. e01490-e01414.
28. **Hoskins, J.** Genome of the bacterium *Streptococcus pneumoniae* strain R6. / Hoskins J., Alborn W. E., Arnold J. [et al]. // Journal of bacteriology. – 2001. – V. 183, № 19. – P. 5709-5717.
29. **Tettelin, H.** Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. / Tettelin H., Nelson K. E., Paulsen I. T. [et al]. // Science. – 2001. – V. 293. – P. 498–506.
30. **Härtel, T.** Impact of glutamine transporters on pneumococcal fitness under infection-related conditions. / Härtel T., Klein M., Koedel U. [et al]. // Infect. Immun. – 2011. – V. 79, № 1. – P. 44-58.
31. **Kerr, A. R.** The Ami-AliA / AliB permease of *Streptococcus pneumoniae* is involved in nasopharyngeal colonization but not in invasive disease. / Kerr A. R., Adrian P. V., Estevao S. [et al]. // Infect. Immun. – 2004. – V. 72, № 7. – P. 3902-3906.
32. **Kloosterman, T. G.** Regulation of glutamine and glutamate metabolism by GlnR and GlnA in *Streptococcus pneumoniae*. / Kloosterman T. G., Hendriksen W. T., Bijlsma J. J. E. [et al]. // J. Biol. Chem. – 2006. – V. 281, № 35. – P. 25097-25109.
33. **Hammerschmidt, S.** Adherence molecules of pathogenic pneumococci. / S. Hammerschmidt // Current Opinion in Microbiology. – 2006. – V. 9, № 1. – P. 12-20.
34. **Paulsen, I. T.** Microbial genome analyses: comparative transport capabilities in eighteen prokaryotes. / Paulsen I. T., Nguyen L., Sliwinski M. K. [et al]. // J. Mol. Biol. – 2000. – V. 301. – P. 75–100.
35. **Ruoff, K. L.** *Streptococcus*. In Manual of clinical microbiology. / Ruoff K. L. [et al]. – ASM Press, Washington D.C. – 1995.
36. **Duane, P. G.** Identification of hydrogen peroxide as a *Streptococcus pneumoniae* toxin for rat alveolar epithelial cells. / P. G. Duane, J. B. Rubins, H. R. Weisel, E. N. Janoff // Infect Immun. – 1993. – V. 61, № 10. – P. 4392-4397.
37. **Agar, N. S.** Erythrocyte catalase. A somatic oxidant defense? / N. S. Agar, S. M. Sadrzadeh, P. E. Hallaway, J. W. Eaton // J. Clin. Invest. – 1986. – V. 77, № 1. – P. 319-321.
38. **Aguiar, S. I.** Emergence of optochin resistance among *Streptococcus pneumoniae* in Portugal. / Aguiar S. I., Frias M. J., Santos L. [et al]. // Microbial Drug Resistance. – 2007. – V. 12, № 4.



39. **Maccannell, D.** Bacterial strain typing. / D. Maccannell // Clin. Lab. Med. – 2013. – V. 33. – P. 629-650.
40. **Enright, M. C.** Extensive variation in the *ddl* gene of penicillin-resistant *Streptococcus pneumoniae* results from a hitchhiking effect driven by the penicillin-binding protein 2b gene. / M. C. Enright, B. G. Spratt // Mol. Biol. Evol. – 1999. – V. 16, № 2. – P. 1687–1695.
41. **Aguiar, S. I.** Displaying the relatedness among isolates of bacterial species – the eBURST approach. / Spratt B. G., Hanage W. P., Li B., [et al]. // FEMS microbiology letters. – 2004. – V. 241. – P. 129-134.
42. **McGee, L.** Nomenclature of major antimicrobial-resistant clones of *Streptococcus pneumoniae* defined by the pneumococcal molecular epidemiology network. / McGee L., McDougal L., Zhou J. [et al]. // Journal of Clinical Microbiology. – 2001. – V. 39. – P. 2565-71.
43. **Enright, M. C.** A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. / M. C. Enright, B. G. Spratt // Microbiology. – 1998. – V. 144. – P. 3049-3060.
44. **Thomas, J. C.** *Streptococcus pneumoniae* clonal complex 199: genetic diversity and tissue-specific virulence. / Thomas J. C., Figueira M., Fennie K. P. [et al]. // PloS one. – 2011. – V. 6, № 4. – P. e18649.
45. **Pai, R.** Postvaccine genetic structure of *Streptococcus pneumoniae* serotype 19A from children in the United States. / Pai R., Moore M. R., Pilishvili T. [et al]. // The Journal of Infectious Diseases. – 2005. – V. 192. – P. 1988-1995.
46. **Hanage, W. P.** Invasiveness of serotypes and clones of *Streptococcus pneumoniae* among children in Finland. / Hanage W. P., Kaijalainen T. H., Syrjanen R. K. [et al]. // Infection and immunity. – 2005. – V. 73, № 1. – P. 431-435.
47. **Sa-Leao, R.** Analysis of invasiveness of pneumococcal serotypes and clones circulating in Portugal before widespread use of conjugate vaccines reveals heterogeneous behavior of clones expressing the same serotype. / Sa-Leao R., Pinto F., Aguiar S. [et al]. // Journal of Clinical Microbiology. – 2011. – V. 49, № 4. – P. 1369-1375.
48. **Langille, M. G. I.** Detecting genomic islands using bioinformatics approaches. / M. G. I. Langille, W. W. L. Hsiao, F. S. L. Brinkman // Nature Reviews Microbiology. – 2010. – V. 8, № 5. – P. 373–382.

49. **Frost, L. S.** Mobile genetic elements: the agents of open source evolution. / L. S. Frost, R. Leplae, A. O. Summers, A. Toussaint. // *Nature Reviews Microbiology*. – 2005. – V. 3, № 9. – P. 722-732.
50. **Croucher, N. J.** Rapid pneumococcal evolution in response to clinical interventions. / Croucher N. J., Simon R. H., Fraser C. [et al]. // *Science*. – 2011. – V. 331, № 6016. – P. 430-4.
51. **Hiller, N. L.** Generation of genic diversity among *Streptococcus pneumoniae* strains via horizontal gene transfer during a chronic polyclonal pediatric infection. / N. L. Hiller, A. Ahmed, E. Powell [et al]. // *PLoS Pathog*. – 2010. – V. 6, № 9. – P. e1001108.
52. **Golubchik, T.** Pneumococcal genome sequencing tracks a vaccine escape variant formed through a multi-fragment recombination event. / T. Golubchik, A. B. Brueggemann, T. Street [et al]. // *Nature Genetics*. – 2012. – V. 44. – P. 352–355.
53. **Toussaint, A.** Mobile elements as a combination of functional modules. / A. Toussaint, C. Merlin // *Plasmid*. – 2002. – V. 47. – P. 26-35.
54. **Johnsborg, O.** Regulation of natural genetic transformation and acquisition of transforming DNA in *Streptococcus pneumoniae*. / O. Johnsborg, L. S. Havarstein // *FEMS Microbiol Rev*. – 2009. – V. 33. – P. 627-642.
55. **Salvadori, G.** Competence in *Streptococcus pneumoniae* and close commensal relatives: mechanisms and implications. / G. Salvadori, R. Junges, D. A. Morrison, F. C. Petersen // *Front. Cell. Infect. Microbiol*. – 2019. – V. 9, № 4. – P. 94.
56. **Prudhomme, M.** Pneumococcal competence coordination relies on a cell-contact sensing mechanism. / M. Prudhomme, M. Berge, B. Martin, P. Polard // *12*. – 2016. – V. 6. – P. e1006113.
57. **Dagkessamanskaia, A.** Interconnection of competence, stress and CiaR regulons in *Streptococcus pneumoniae*: competence triggers stationary phase autolysis of *ciaR* mutant cells. / A. Dagkessamanskaia, M. Moscoso, V. Hénard [et al]. // *Molecular microbiology*. – 2004. – V. 51, № 4. – P. 1071-1086.
58. **Shoemaker, N. B.** Organization and transfer of heterologous chloramphenicol and tetracycline resistance genes in pneumococcus. / N. B. Shoemaker, M. D. Smith, W. R. Guild // *J Bacteriol*. – 1979. – V. 139. – P. 432-441.

59. **Croucher, N. J.** Role of conjugative elements in the evolution of the multidrug-resistant pandemic clone *Streptococcus pneumoniae* Spain23FST81. / Croucher N. J., D. Walker, P. Romero [et al]. // Journal of bacteriology. – 2009. – V. 191, № 5. – P. 1480-1489.
60. **Cornejo, O. E.** Polymorphic competence peptides do not restrict recombination in *Streptococcus pneumoniae*. / O. E. Cornejo, L. McGee, D. E. Rozen // Molecular Biology and Evolution. – 2010. – V. 27, № 3. – P. 694-702.
61. **Rivera-Olivero, I. A.** Multiplex PCR reveals a high rate of nasopharyngeal pneumococcal 7-valent conjugate vaccine serotypes co-colonizing indigenous Warao children in Venezuel. / Rivera-Olivero I. A., Blommaert M., Bogaert D. [et al]. // Journal of medical microbiology. – 2009. – V. 58. – P. 584-587.
62. **Johnsborg, O.** A predatory mechanism dramatically increases the efficiency of lateral gene transfer in *Streptococcus pneumoniae* and related commensal species. / O. Johnsborg, V. Eldholm, M. L. Bjørnstad, L. S. Håvarstein // Molecular microbiology. – 2008. – V. 69, № 1. – P. 245-253.
63. **Guiral, S.** Competence-programmed predation of noncompetent cells in the human pathogen *Streptococcus pneumoniae*: genetic requirements. / S. Guiral, T. J. Mitchell, B. Martin, J.-P. Claverys // Proc Natl Acad Sci USA. – 2005. – V. 102, № 24. – P. 8710–8715.
64. **Claverys, J. P.** Competence-induced fratricide in streptococci. / J. P. Claverys, B. Martin, L. S. Håvarstein // Molecular microbiology. – 2007. – V. 64. – P. 1423-1433.
65. **Lacks, S.** Role of a deoxyribonuclease in the genetic transformation of *Diplococcus pneumoniae*. / S. Lacks, B. Greenberg, M. Neuberger // Proc. Nat. Acad. Sci. USA. – 1974. – V. 71, № 6. – P. 2305-2309.
66. **Morrison, D. A.** Breakage prior to entry of donor DNA in *Pneumococcus* transformation. / D. A. Morrison, W. R. Guild // Biochimica et Biophysica Acta. – 1973. – V. 299, № 4. – P. 545-556.
67. **Chen, I.** DNA uptake during bacterial transformation. / I. Chen, D. Dubnau // Nat Rev Microbiol. – 2004. – V. 2, № 3. – P. 241-249.
68. **Chen, I.** The ins and outs of DNA transfer in bacteria. / I. Chen, P. J. Christie, D. Dubnau // Science. – 2005. – V. 310, № 5753. – P. 1456-1460.
69. **Bergé, M.** Transformation of *Streptococcus pneumoniae* relies on DprA- and RecA-dependent protection of incoming DNA single strands. / M. Bergé, I. Mortier-Barrière, B. Martin, J.-P. Claverys // Mol Microbiol. – 2003. – V. 50. – P. 527-536.

70. **Morrison, D. A.** Transformation and deoxyribonucleic acid size: extent of degradation on entry varies with size of donor. / D. A. Morrison, W. R. Guild // *J Bacteriol.* – 1972. – V. 112, № 3. – P. 1157-1168.
71. **Attaiech, L.** Role of the single-stranded DNA – binding protein SsbB in pneumococcal transformation: maintenance of a reservoir for genetic plasticity. / Attaiech L., Olivier A., Mortier-Barrière I. [et al]. // *PLoS Genet.* – 2011. – V. 7, № 6. – P. e1002156.
72. **Mortier-Barriere, I.** Control of recombination rate during transformation of *Streptococcus pneumoniae*: an overview. / Mortier-Barriere I., Humbert O., Martin B., Prudhomme M., Claverys J. P. // *Microb Drug Resist.* – 1997. – V. 3, № 3. – P. 233-242.
73. **Mcdonnell, M.** “ Diplophage”: a bacteriophage of *Diplococcus pneumoniae*. / M. Mcdonnell, C. Ronda-Lain, A. Tomasz // *Virology.* – 1975. – V. 63. – P. 577-582.
74. **Romero, P.** Comparative genomic analysis of ten *Streptococcus pneumoniae* temperate bacteriophages. / Romero P., Croucher N. J., Hiller L. [et al]. // *Journal of bacteriology.* – 2009. – V. 151, № 15. – P. 4854-4862.
75. **Smith, M. D.** A plasmid in *Streptococcus pneumoniae*. / M. D. Smith, W. R. Guild // *Journal of bacteriology.* – 1979. – V. 137, № 2. – P. 735-739.
76. **Dagkessamanskaia, A.** Interconnection of competence, stress and CiaR regulons in *Streptococcus pneumoniae*: competence triggers stationary phase autolysis of ciaR mutant cells. / Dagkessamanskaia A., Moscoso M., Hénard V. [et al]. // *Molecular microbiology.* – 2004. – V. 51, № 4. – P. 1071-1086.
77. **Croucher, N. J.** Role of conjugative elements in the evolution of the multidrug-resistant pandemic clone *Streptococcus pneumoniae* Spain23FST81. / Croucher N. J., D. Walker, P. Romero [et al]. // *Journal of bacteriology.* – 2009. – V. 191, № 5. – P. 1480-1489.
78. **Henderson-Begg, S. K.** Diversity of putative Tn5253-like elements in *Streptococcus pneumoniae*. / S. K. Henderson-Begg, A. P. Roberts, L. M. C. Hall // *International journal of antimicrobial agents.* – 2009. – V. 33, № 4. – P. 364-367.
79. **Roberts, A. P.** A modular master on the move: the Tn 916 family of mobile genetic elements. / A. P. Roberts, P. Mullany // *Trends in Microbiology.* – 2009. – V. 17, № 6. – P. 251-258.
80. **Li, Y.** Molecular characterization of erm(B)- and mef(E)-mediated erythromycin-resistant *Streptococcus pneumoniae* in China and complete DNA sequence of Tn2010. / Li Y., Tomita H., Lv Y., [et al]. // *Journal of Applied Microbiology.* – 2010. – V. 110. – P. 254-265.

81. **Chancey, S. T.** Composite mobile genetic elements disseminating macrolide resistance in *Streptococcus pneumoniae*. / Chancey S. T., Agrawal S., Schroeder M. R. [et al]. // *Frontiers in microbiology*. – 2015. – V. 6. – P. 26.
82. **Møller, P.** Adduct formation, mutagenesis and nucleotide excision repair of DNA damage produced by reactive oxygen species and lipid peroxidation product. / P. Møller, H. Wallin // *Mutat. Res.* – 1998. – V. 410. – P. 271–290.
83. **Pericone, C. D.** Inhibitory and bactericidal effects of hydrogen peroxide production by *Streptococcus pneumoniae* on other inhabitants of the upper respiratory tract. / C. D. Pericone, K. Overweg, P. W. Hermans, J. N. Weiser // *Infect. Immun.* – 2000. – V. 68. – P. 3990–3997.
84. **Munoz-Najar, U.** An operon that confers UV resistance by evoking the SOS mutagenic response in streptococcal conjugative transposon Tn5252. / U. Munoz-Najar, M. N. Vijayakumar // *J. Bacteriol.* – 1999. – V. 181. – P. 2782–2788.
85. **Bergmann, S.** Versatility of pneumococcal surface proteins. / S. Bergmann, S. Hammerschmidt // *Microbiology*. – 2006. – V. 152. – P. 295–303.
86. **Brown, J. S.** A locus contained within a variable region of pneumococcal pathogenicity island 1 contributes to virulence in mice. / J. S. Brown, S. M. Gilliland, B. G. Spratt, and D. W. Holden // *Infection and immunity*. – 2004. – V. 72, № 3. – P. 1587–1593.
87. **Brown, J. S.** A *Streptococcus pneumoniae* pathogenicity island encoding an ABC transporter involved in iron uptake and virulence. / J. S. Brown, S. M. Gilliland, D. W. Holden // *Molecular microbiology*. – 2001. – V. 40, № 3. – P. 572–585.
88. **Lux, T.** Diversity of bacteriocins and activity spectrum in *Streptococcus pneumoniae*. / T. Lux, M. Nuhn, R. Hakenbeck, P. Reichmann // *J. Bacteriol.* – 2007. – V. 189. – P. 7741–7751.
89. **Willey, J. M.** Lantibiotics: peptides of diverse structure and function. / J. M. Willey, W. A van der Donk // *Annu. Rev. Microbiol.* – 2007. – V. 61. – P. 477–501.
90. **Slamti, L.** A cell-cell signaling peptide activates the PlcR virulence regulon in bacteria of the *Bacillus cereus* group. / L. Slamti, D. Lereclus // *EMBO J.* – 2002. – V. 21. – P. 4550–4559.
91. **Ramirez, M.** A high incidence of prophage carriage among natural isolates of *Streptococcus pneumoniae*. / M. Ramirez, E. Severina, A. Tomasz // *Journal of bacteriology*. – 1999. – V. 181, № 12. – P. 3618–3625.

92. **Garcia, P.** Bacteriophages of *Streptococcus pneumoniae*: a molecular approach. / P. Garcia, A. C. Martin, R. Lopez // Microbial Drug Resistance. – 1997. – V. 3, № 2. – P. 165-176.
93. **Martin, A. C.** Analysis of the complete nucleotide sequence and functional organization of the genome of *Streptococcus pneumoniae* bacteriophage Cp-1. / A. C. Martin, R. Lopez, Garcia P. // Journal of virology. – 1996. – V. 70, № 6. – P. 3678-3687.
94. **Obrego, V.** Genome organization and molecular analysis of the temperate bacteriophage MM1 of *Streptococcus pneumoniae*. / Obrego V., García J. L., Garcí E. [et al]. // Journal of bacteriology. – 2003. – V. 185, № 7. – P. 2362-2368.
95. **Pelletier, J.** Unated States Patent USOO678393OB1. Development of novel anti-microbial agents based on bacteriophage genomics. / J. Pelletier, P. Gros, M. Dubow // Patent № – US 6,783,930 B1. – Aug. 31, 2004
96. **Bensing, B. A.** Genetic loci of *Streptococcus mitis* that mediate binding to human platelets. / B. A. Bensing, C. E. Rubens, P. M. Sullam // Infection and immunity. – 2001. – V. 69, № 3. – P. 1373-1380.
97. **Bensing, B. A.** Proteins PblA and PblB of *Streptococcus mitis*, which promote binding to human platelets, are encoded within a lysogenic bacteriophage. / B. A. Bensing, I. R. Siboo, P. M. Sullam // Infection and immunity. – 2001. – V. 69, № 10. – P. 6186-6192.
98. **Wagner, P. L.** Bacteriophage control of bacterial virulence. / P. L. Wagner, M. K. Waldor // Infection and immunity. – 2002. – V. 70, № 8. – P. 3985-3993.
99. **Loeffler, J. M.** Lysogeny of *Streptococcus pneumoniae* with MM1 phage: improved adherence and other phenotypic changes. / J. M. Loeffler, V. A. Fischetti // Infection and immunity. – 2006. – V. 74, № 8. – P. 4486-4495.
100. **Obregón, V.** VO1, a temperate bacteriophage of the type 19A multiresistant epidemic 8249 strain of *Streptococcus pneumoniae*: analysis of variability of lytic and putative C5 methyltransferase genes. / V. Obregón, P. García, R. López, J. L. García // Microbial Drug Resistance. – 2003. – V. 9, № 1. – P. 7-15.
101. **Sampath, J.** Identification of a DNA cytosine methyltransferase gene in conjugative transposon Tn 5252. / J. Sampath, M. N. Vijayakumar // Plasmid. – 1998. – V. 39. – P. 63-76.
102. **Rocha, E. P. C.** Order and disorder in bacterial genomes. / E. P. C. Rocha // Current opinion in microbiology. – 2004. – V. 7, № 5. – P. 519-527.

103. **Dybvig, K.** A family of phase-variable restriction enzymes with differing specificities generated by high-frequency gene rearrangements. / K. Dybvig, R. Sitaraman, C. T. French // PNAS. – 1998. – V. 95, № 23. – P. 13923-13928.
104. **Vasu, K.** Diverse functions of restriction-modification systems in addition to cellular defense. / K. Vasu, V. Nagaraja // Microbiology and Molecular Biology Reviews. – 2013. – V. 77, № 1. – P. 53-72.
105. **Stein, D. C.** Restriction and modification systems of *Neisseria gonorrhoeae*. / D. C. Stein, J. S. Gunn, M. Radlinska, A. Piekarowicz // Gene. – 1995. – V. 157, № 1-2. – P. 19-22.
106. **Loenen, W. A. M.** Tracking EcoKI and DNA fifty years on: a golden story full of surprises. / W. A. M. Loenen // Nucleic Acids Res. – 2003. – V. 31, № 24. – P. 7059–7069.
107. **Warren, R. A.** Modified bases in bacteriophage DNAs. / R. A. Warren // Annu Rev Microbiol. – 1980. – V. 34. – P. 137-158.
108. **Krugert, D. H.** Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. / D. H. Krugert, T. A. Bickle // Microbiological Reviews. – 1983. – V. 47, № 3. – P. 345-360.
109. **Gelfand, M. S.** Avoidance of palindromic words in bacterial and archaeal genomes: a close connection with restriction enzymes. / M. S. Gelfand, E. V. Koonin // Nucleic Acids Research. – 1997. – V. 25, № 12. – P. 2430-2439.
110. **Nechaev, S.** The elusive object of desire - interactions of bacteriophages and their hosts. / S. Nechaev, K. Severinov // Curr Opin Microbiol. – 2008. – V. 11, № 2. – P. 186-193.
111. **Johnston, C.** The DpnI / DpnII pneumococcal system, defense against foreign attack without compromising genetic exchange. / C. Johnston, P. Polard, J.-P. Claverys // Mobile Genetic Elements. – 2013. – V. 3, № 4. – P. e25582.
112. **Loenen, W. A.** Highlights of the DNA cutters: a short history of the restriction enzymes. / Loenen W. A. M., Dryden D. T. F., Raleigh E. A. [et al]. // Nucleic Acids Research. – 2014. – V. 42, № 1. – P. 3-19.
113. **Lacks, S.** Complementary specificity of restriction endonucleases of *Diplococcus pneumoniae* with respect to DNA methylation. / S. Lacks, B. Greenberg // Journal of molecular biology. – 1977. – V. 114, № 1. – P. 153-168.
114. **Eutsey, R. A.** Genetic stabilization of the drug-resistant PMEN1 pneumococcus lineage by its distinctive dpniii restriction-modification system. / Eutsey R. A., Powell E., Dordel J. [et al]. // mBio. – 2015. – V. 6, № 5. – P. e00173-15.

115. **Croucher, N. J.** Diversification of bacterial genome content through distinct mechanisms over different timescales. / Croucher N. J., Coupland P. G., Stevenson A. E. [et al]. // *Nat Commun.* – 2014. – V. 5. – P. 5471.
116. **Wyres, K. L.** The multidrug-resistant PMEN1 pneumococcus is a paradigm for genetic success. / Wyres K. L., Lambertsen L. M., Croucher N. J. [et al]. // *Genome biology.* – 2012. – V. 13, № 11. – P. R103.
117. **Kobayashi, I.** Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. / I. Kobayashi // *Nucleic Acids Research.* – 2001. – V. 29, № 18. – P. 3742-3756.
118. **Kulakauskas, S.** DNA restriction-modification systems mediate plasmid maintenance. / S. Kulakauskas, A. Lubys, S. D. Ehrlich // *Journal of bacteriology.* – 1995. – V. 177, № 12. – P. 3451-3454.
119. **Ishikawa, K.** Conflicts targeting epigenetic systems and their resolution by cell death: novel concepts for methyl-specific and other restriction systems. / K. Ishikawa, E. Fukuda, I. Kobayashi // *DNA research.* – 2010. – V. 17. – P. 325-342.
120. **Fukuda, E.** Cell death upon epigenetic genome methylation: a novel function of methyl-specific deoxyribonucleases. / E. Fukuda, K. H. Kaminska, J. M. Bujnicki, I. Kobayashi // *Genome Biology.* – 2008. – V. 9, № 11. – P. R163.
121. **Ishikawa, K.** Cleavage of a model DNA replication fork by a Type I restriction endonuclease. / K. Ishikawa, N. Handa, I. Kobayashi // *Nucleic Acids Research.* – 2009. – V. 37, № 11. – P. 3531-3544.
122. **De Ste Croix, M.** Phase-variable methylation and epigenetic regulation by type I restriction-modification systems. / De Ste Croix M., Vacca I., Kwun M. [et al]. // *FEMS microbiology reviews.* – 2017. – V. 41, № 1. – P. S3-S15.
123. **Murray, N. E.** Type I restriction systems: sophisticated molecular machines (a legacy of Bertani type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). / N. E. Murray // *Microbiology and Molecular Biology Reviews.* – 2000. – V. 64, № 2. – P. 412–434.
124. **Manso, A. S.** A random six-phase switch regulates pneumococcal virulence via global epigenetic changes. / Manso A. S., Chai M. H., Atack J. M. [et al]. // *Nature Communications.* – 2014. – V. 5. – P. 1-9.



125. **Blow, M. J.** The Epigenomic landscape of prokaryotes. / Blow M. J., Clark T. A., Daum C. G. [et al]. // PLOS Genetics. – 2016. – V. 12, № 5. – P. e1006064.
126. **Li, J.** Epigenetic switch driven by DNA inversions dictates phase variation in *Streptococcus pneumoniae*. / Li J., Li J.-W., Feng Z. [et al]. // PLoS Pathog. – 2016. – V. 12, № 7. – P. 1-36.
127. **Harvey, R. M.** A variable region within the genome of *Streptococcus pneumoniae* contributes to strain-strain variation in virulence. / Harvey R. M., Stroehner U. H., Ogunniyi A. D., [et al]. // PloS one. – 2011. – V. 6, № 5. – P. e19650.
128. **Lees, J. A.** Large scale genomic analysis shows no evidence for pathogen adaptation between the blood and cerebrospinal fluid niches during bacterial meningitis. / Lees J. A., Kremer P. H. C., Manso A. S. [et al]. // 3. – 2017.
129. **Weiser, J. N.** Phase variation in pneumococcal opacity: relationship between colonial morphology and nasopharyngeal colonization. / J. N. Weiser, R. Austrian, P. K. Sreenivasan, H. R. Masure // Infection and immunity. – 1994. – V. 62, № 6. – P. 2582-2589.
130. **Martin, B.** A highly conserved repeated DNA element located in the chromosome of *Streptococcus pneumoniae*. / Martin B., Humbert O., Camara M., [et al]. // Nucleic Acids Research. – 1992. – V. 20, № 13. – P. 3479-3483.
131. **Oggioni, M. R.** Repeated extragenic sequences in prokaryotic genomes: a proposal for the origin and dynamics of the RUP element in *Streptococcus pneumoniae*. / M. R. Oggioni, J.-P. Claverys // Microbiology. – 1999. – V. 145. – P. 2647–2653.
132. **Knutsen, E.** BOX elements modulate gene expression in *Streptococcus pneumoniae*: impact on the fine-tuning of competence development. / Knutsen E., Johnsborg O., Quentin Y. [et al]. // Journal of bacteriology. – 2006. – V. 188, № 23. – P. 8307-8312.
133. **Croucher, N. J.** Identification, variation and transcription of pneumococcal repeat sequences. / N. J. Croucher, G. S. Vernikos, J. Parkhill, S. D. Bentley // BMC Genomics. – 2011. – V. 12. – P. 120.
134. **Saluja, S. K.** The genetic basis of colony opacity in *Streptococcus pneumoniae*: evidence for the effect of box elements on the frequency of phenotypic variation. / S. K. Saluja, J. N. Weiser // Molecular microbiology. – 1995. – V. 16, № 2. – P. 215-227.
135. **de Hoon, M. J. L.** Prediction of transcriptional terminators in *Bacillus subtilis* and related species. / M. J. L. de Hoon, Y. Makita, K. Nakai, S. Miyano // PLoS Comput Biol. – 2005. – V. 1, № 3. – P. e25.

136. **Croucher, N. J.** Population genomics of post-vaccine changes in pneumococcal epidemiology / Croucher N. J., Finkelstein J. A., Pelton S. I. [et al]. // *Nat Genet.* – 2013. – V. 45, № 6. – P. 656-63.
137. **Croucher, N. J.** Population genomic datasets describing the post-vaccine evolutionary epidemiology of *Streptococcus pneumoniae* / Croucher N. J., Finkelstein J. A., Pelton S. I. [et al]. // *Sci Data.* – 2015. – V. 2. – P. 150058.
138. **Bentley, S. D.** Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. / Bentley S. D., Aanensen D. M., Mavroidi A. [et al]. // *PLoS Genet.* – 2006. – V. 2, № 3. – P. e31.
139. **Llull, D.** Tts, a processive beta-glucosyltransferase of *Streptococcus pneumoniae*, directs the synthesis of the branched type 37 capsular polysaccharide in pneumococcus and other gram-positive species. / D. Llull, E. Garcia, R. Lopez // *J Biol Chem.* – 2001. – V. 276, № 24. – P. 21053-21061.
140. **Llull, D.** A single gene (*tts*) located outside the cap locus directs the formation of *Streptococcus pneumoniae* type 37 capsular polysaccharide. Type 37 pneumococci are natural, genetically binary strains. / D. Llull, R. Muñoz, R. López, E. García [et al]. // *J Exp Med.* – 1999. – V. 190, № 2. – P. 241-251.
141. **Varvio, S.-L.** Evolution of the capsular regulatory genes in *Streptococcus pneumoniae*. / S.-L. Varvio, K. Auranen, E. Arjas, P. H. Mäkelä // *The Journal of Infectious Diseases.* – 2009. – V. 200, № 7. – P. 1144–1151.
142. **Kadioglu, A.** The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease. / A. Kadioglu, J. Weiser, J. Paton, P. Andrew (2008). // *Nature Reviews Microbiology.* – 2008. – V. 6, № 4. – P. 288-301.
143. **Coffey, T. J.** Horizontal transfer of multiple penicillin-binding protein genes, and capsular biosynthetic genes, in natural populations of *Streptococcus pneumoniae*. / Coffey T. J., Dowson C. G., Daniels M., [et al]. // *Molecular microbiology.* – 1991. – V. 5, № 9. – P. 2255-2260.
144. **Coffey, T. J.** Recombinational exchanges at the capsular polysaccharide biosynthetic locus lead to frequent serotype changes among natural isolates of *Streptococcus pneumoniae*. / Coffey T. J., Enright M. C., Daniels M. [et al]. // *Molecular microbiology.* – 1998. – V. 27, № 1. – P. 73-83.

145. **Brueggemann, A. B.** Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. / A. B. Brueggemann, R. Pai, D. W. Crook, B. Beall // *PLoS Pathog.* – 2007. – V. 3, № 11. – P. e168.
146. **Sánchez-Beato, A. R.** Identification and characterization of IS1381, a new insertion sequence in *Streptococcus pneumoniae*. / A. R. Sánchez-Beato, E. García, R. López, J. L. García // *Journal of bacteriology.* – 1997. – V. 179, № 7. – P. 2459-2463.
147. **Lambowitz, A. M.** Mobile group II introns. / A. M. Lambowitz, S. Zimmerly // *Annual Review of Genetics.* – 2004. – V. 38, № 1. – P. 1-35.
148. **Sánchez-Beato, A. R.** Molecular organization of the genes required for the synthesis of type 1 capsular polysaccharide of *Streptococcus pneumoniae*: formation of binary encapsulated pneumococci and identification of cryptic dTDP-rhamnose biosynthesis genes. / R. Muñoz, M. Mollerach, R. López, E. García // *Mol Microbiol.* – 1997. – V. 25, № 1. – P. 79-92.
149. **Ramirez, M.** Molecular characterization of the complete 23F capsular polysaccharide locus of *Streptococcus pneumoniae*. / M. Ramirez, A. Tomasz // *J Bacteriol.* – 1998. – V. 180, № 19. – P. 5273-5278.
150. **Morona, J. K.** Comparative genetics of capsular polysaccharide biosynthesis in *Streptococcus pneumoniae* types belonging to serogroup 19. / J. K. Morona, R. Morona, J. C. Paton // *J Bacteriol.* – 1999. – V. 181, № 17. – P. 5355-5364.
151. **Morona, J. K.** Molecular and genetic characterization of the capsule biosynthesis locus of *Streptococcus pneumoniae* type 19B. / J. K. Morona, R. Morona, J. C. Paton // *Journal of bacteriology.* – 1997. – V. 179, № 15. – P. 4953-4958.
152. **Mavroidi, A.** Evolutionary genetics of the capsular locus of serogroup 6 pneumococci. / Mavroidi A., Godoy D., Aanensen D. M., [et al]. // *Journal of bacteriology.* – 2004. – V. 186, № 24. – P. 8181-8192.
153. **Bratcher, P.** Evolution of the capsular gene locus of *Streptococcus pneumoniae* serogroup 6. / Bratcher P., Park E. I. H., Oliver M. B. [et al]. // *Microbiology.* – 2011. – V. 157. – P. 189-198.
154. **Elberse, K.** Sequence diversity within the capsular genes of *Streptococcus pneumoniae* serogroup 6 and 19. / Elberse K., Witteveen S., van der Heide H. [et al]. // *PloS one.* – 2011. – V. 6, № 9. – P. e25018.

155. **Salter, S. J.** Variation at the capsule locus, *cps*, of mistyped and non-typable *Streptococcus pneumoniae* isolates. / Salter S. J., Hinds J., Gould K. A. [et al]. // *Microbiology*. – 2012. – V. 158. – P. 1560-1569.
156. **Park, I. H.** Nontypeable pneumococci can be divided into multiple *cps* types, including one type expressing the novel gene *pspK*. / Park I. H., Kim K.-H., Andrade A. L. [et al]. // *mBio*. – 2012. – V. 3, № 3. – P. e00035-12.
157. **Barry, M. A.** Serotypes of *Streptococcus pneumoniae* isolated from blood cultures at Boston city hospital between 1979 and 1982. / M. A. Barry, D. E. Craven, M. Finland // *Journal of Infectious Diseases*. – 1984. – V. 149, № 3. – P. 449-452.
158. **Butler, J. C.** Serotype distribution of *Streptococcus pneumoniae* infections among preschool children in the United States, 1978-1994: implications for development of a conjugate vaccine. / Butler J. C., Breiman R. F., Lipman H. B. [et al]. // *The Journal of Infectious Diseases*. – 1995. – V. 171. – P. 885-889.
159. **Choi, E. H.** *Streptococcus pneumoniae* serotype 19A in children, South Korea. / Choi E. H., Kim S. H., Eun B. W. [et al]. // *Emerging Infectious Diseases*. – 2008. – V. 14, № 2. – P. 275-281.
160. **Foster, D.** Invasive pneumococcal disease: epidemiology in children and adults prior to implementation of the conjugate vaccine in the Oxfordshire region, England. / Foster D., Kyle Knox 3, A S Walker [et al]. // *J Med Microbiol*. – 2008. – V. 57. – P. 480-487.
161. **Harboe, Z. B.** Temporal trends in invasive pneumococcal disease and pneumococcal serotypes over 7 decades. / Harboe Z. B., Benfield T. L., Valentiner-Branth P. [et al]. // *Clin Infect Dis*. – 2010. – V. 50. – P. 329-337.
162. **Kyaw, M. H.** Serotypes/groups distribution and antimicrobial resistance of invasive pneumococcal isolates: implications for vaccine strategies. / Kyaw M. H., Clarke S., Edwards G. F. [et al]. // *Epidemiol Infect*. – 2000. – V. 125, № 3. – P. 561-572.
163. **Lagos, R.** Age- and serotype-specific pediatric invasive pneumococcal disease: insights from systematic surveillance in Santiago, Chile, 1994 – 2007. / Lagos R., Muñoz A., Martin O. S. [et al]. // *J Infect Dis*. – 2008. – V. 198. – P. 1809-1817.
164. **Mufson, M. A.** Bacteremic pneumococcal pneumonia in one American city: a 20-year longitudinal study, 1978–1997. / M. A. Mufson, R. J. Stanek // *The American Journal of Medicine*. – 1999. – V. 107, № 1. – P. 34–43.

165. **Scheifele, D.** Invasive pneumococcal infections in Canadian children, 1991-1998: implications for new vaccination strategies. Canadian Paediatric Society/Laboratory Centre for Disease Control Immunization Monitoring Program, Active (IMPACT). / Scheifele D., Halperin S., Pelletier L., Talbot J. [et al]. // *Clinical Infectious Diseases*. – 2000. – V. 31, № 1. – P. 58-64.
166. **Weinberger, D. M.** Serotype replacement in disease after pneumococcal vaccination. / D. M. Weinberger, R. Malley, M. Lipsitch // *The Lancet*. – 2011. – V. 378, № 9807. – P. 1962-1973.
167. **Feikin, D. R.** Historical changes in pneumococcal serogroup distribution: implications for the era of pneumococcal conjugate vaccines. / D. R. Feikin, K. P. Klugman // *Clinical Infectious Diseases*. – 2002. – V. 35, № 5. – P. 547–555.
168. **Fenoll, A.** Temporal trends of invasive *Streptococcus pneumoniae* serotypes and antimicrobial resistance patterns in Spain from 1979 to 2007. / Fenoll A., Granizo J. J., Aguilar L., et al. // *J Clin Microbiol*. – 2007. – V. 47. – P. 1012-1020.
169. **Fenoll, N.** Evolution of *Streptococcus pneumoniae* serotypes and antibiotic resistance in Spain: update (1990 to 1996). / Fenoll N., Jado I., Vicioso D. [et al]. // *J Clin Microbiol*. – 1998. – V. 36, № 12. – P. 3447-3454.
170. **Jefferies, J. M. C.** Genetic analysis of diverse disease-causing pneumococci indicates high levels of diversity within serotypes and capsule switching. / Jefferies J. M. C., Smith A., Clarke S. C. [et al]. // *Journal of clinical microbiology*. – 2004. – V. 42, № 12. – P. 5681-5688.
171. **Marttinen, P.** Detection of recombination events in bacterial genomes from large population samples. / Marttinen P., Hanage W. P., Croucher N. J. [et al]. // *Nucleic Acids Research*. – 2012. – V. 40, № 1. – P. e6.
172. **Pillai, D. R.** Genome-wide dissection of globally emergent multi-drug resistant serotype 19A *Streptococcus pneumoniae*. / Pillai D. R., Shahinas D., Buzina A. [et al]. // *BMC Genomics*. – 2009. – V. 10. – P. 642.
173. **Wyres, K. L.** Pneumococcal capsular switching: a historical perspective. / Wyres K. L., L. M. Lambertsen, N. J. Croucher [et al]. // *The Journal of Infectious Diseases*. – 2013. – V. 207, № 3. – P. 439-449.

174. **Calix, J. J.** Biochemical, genetic, and serological characterization of two capsule subtypes among *Streptococcus pneumoniae* serotype 20 strains. / Calix J. J., Porambo R. J., Brady A. M. [et al]. // J Biol Chem. – 2012. – V. 287, № 33. – P. 27885-27894.
175. **Calix, J. J.** A new pneumococcal serotype, 11E, has a variably inactivated wcjE gene. / J. Calix, M. H. Nahm // The Journal of Infectious Diseases. – 2010. – V. 202, № 1. – P. 29-38.
176. **Henrichsen, J.** Six newly recognized types of *Streptococcus pneumoniae*. / J. Henrichsen // Journal of Clinical Microbiology. – 1995. – V. 33, № 10. – P. 2759-2762.
177. **Denapaite, D.** A new variant of the capsule 3 cluster occurs in *Streptococcus pneumoniae* from deceased wild chimpanzees. / D. Denapaite, R. Hakenbeck // PloS one. – 2011. – V. 6, № 9. – P. e25119.
178. **Mavroidi, A.** Genetic relatedness of the *Streptococcus pneumoniae* capsular biosynthetic loci. / Mavroidi A., Aanensen D. M., Godoy D., [et al]. // Journal of bacteriology. – 2007. – V. 189, № 21. – P. 7841-7855.
179. **Park, I. H.** Discovery of a new capsular serotype (6C) within serogroup 6 of *Streptococcus pneumoniae*. / Park I. H., Pritchard D. G., Cartee R. [et al]. // Journal of Clinical Microbiology. – 2007. – V. 45, № 4. – P. 1225-1233.
180. **Burton, R. L.** Pneumococcus with the “6E” cps locus produces serotype 6B capsular polysaccharide. / R. L. Burton, K. A. Geno, J. S. Saad, M. H. Nahm // Journal of Clinical Microbiology. – 2016. – V. 54, № 4. – P. 967-971.
181. **Paton, J. C.** *Streptococcus pneumoniae* capsular polysaccharide. / J. C. Paton, C. Trappetti // Microbiology spectrum. – 2019. – V. 7, № 2. – P. GPP3-0019.
182. **Johnson, H. L.** Systematic evaluation of serotypes causing invasive pneumococcal disease among children under five: The Pneumococcal Global Serotype Project. / Johnson H. L., Deloria-Knoll M., Levine O. S. [et al]. // PLoS Med. – 2010. – V. 7, № 10. – P. e1000348.
183. **Hausdorff, W. P.** Which pneumococcal serogroups cause the most invasive disease: implications for conjugate vaccine formulation and use, part I. / W. P. Hausdorff, J. Bryant, P. R. Paradiso, G. R. Siber // Clinical Infectious Diseases. – 2000. – V. 30, № 1. – P. 100–121.
184. **Jefferson, T.** *Streptococcus pneumoniae* in western Europe: serotype distribution and incidence in children less than 2 years old. / Jefferson T., Ferroni E., Curtale F. [et al]. // The Lancet Infectious Diseases. – 2006. – V. 6, № 7. – P. 405-410.
185. **Brandileone, M. C.** Distribution of serotypes of *Streptococcus pneumoniae* isolated from invasive infections over a 16-year period in the greater Sao Paulo area, Brazil. / Brandileone

- M. C., Vieira V. S., Zanella R. C. [et al]. // J Clin Microbiol. – 1995. – V. 33, № 10. – P. 2789-2791.
186. **Robinson, K. A.** Epidemiology of invasive *Streptococcus pneumoniae* infections in the United States, 1995-1998. Opportunities for prevention in the conjugate vaccine era. / Robinson K. A., Baughman W., Rothrock G., [et al]. // JAMA. – 2001. – V. 285, № 13. – P. 1729-1735.
187. **Siber, G.** Pneumococcal serotype epidemiology. In Pneumococcal vaccines: the impact of conjugate vaccine. / G. Siber, K.P. Klugman, P.H. Mäkelä (ed.) // ASM Press, Washington D.C., 2008. – 449 pp.
188. **Sleeman, K. L.** Capsular serotype-specific attack rates and duration of carriage of *Streptococcus pneumoniae* in a population of children. / Sleeman K. L., Griffiths D., Shackley F. [et al]. // The Journal of Infectious Diseases. – 2006. – V. 194, № 5. – P. 682-688.
189. **Hathaway, L. J.** Capsule type of *Streptococcus pneumoniae* determines growth phenotype. / Hathaway L. J., Brugger S. D., Morand B. [et al]. // PLoS Pathog. – 2012. – V. 8, № 3. – P. e1002574.
190. **Trzcinski, K.** Effect of serotype on pneumococcal competition in a mouse colonization model. / Trzcinski K., Li Y., Weinberger D. M. [et al]. // mBio. – 2015. – V. 6, № 5. – P. e00902-15.
191. **Battig, P.** Capsule genes of *Streptococcus pneumoniae* influence growth in vitro. / P. Battig, K. Muhlemann // FEMS Immunology & Medical Microbiology. – 2007. – V. 50, № 3. – P. 324-329.
192. **Weinberger, D. M.** Pneumococcal capsular polysaccharide structure predicts serotype prevalence. / Weinberger D. M., Trzciński K., Lu Y.-J., [et al]. // PLoS Pathog. – 2009. – V. 5, № 6. – P. e1000476.
193. **Melin, M.** The capsular serotype of *Streptococcus pneumoniae* is more important than the genetic background for resistance to complement. / Melin M., Trzciński K., Meri S. [et al]. // Infect Immun. – 2010. – V. 78, № 12. – P. 5262-5270.
194. **Rosenow, C.** Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of *Streptococcus pneumoniae*. / Rosenow C., Ryan P., Weiser J. N., Johnson S. [et al]. // Molecular microbiology. – 1997. – V. 25, № 5. – P. 819-829.

195. **Giammarinaro, P.** Role of RegM, a homologue of the catabolite repressor protein CcpA, in the virulence of *Streptococcus pneumoniae*. / P. Giammarinaro, J. C. Paton // Infection and immunity. – 2002. – V. 70, № 10. – P. 5454-5461.
196. **Davis, K. M.** Resistance to mucosal lysozyme compensates for the fitness deficit of peptidoglycan modifications by *Streptococcus pneumoniae*. / K. M. Davis, H. T. Akinbi, A. J. Standish, J. N. Weiser // PLoS Pathog. – 2008. – V. 4, № 12. – P. e1000241.
197. **Lemessurier, K. S.** Differential expression of key pneumococcal virulence genes in vivo. / K. S. Lemessurier, A. D. Ogunniyi, J. C. Paton // Microbiology. – 2006. – V. 152. – P. 305-311.
198. **Gupta, R.** Deletion of *arcD* in *Streptococcus pneumoniae* D39 impairs its capsule and attenuates virulence. / Gupta R., Yang J., Dong Y. [et al]. // Infect Immun. – 2003. – V. 81, № 10. – P. 3903-3911.
199. **Weiser, J. N.** Changes in availability of oxygen accentuate differences in capsular polysaccharide expression by phenotypic variants and clinical isolates of *Streptococcus pneumoniae*. / Weiser J. N., Bae D., Epino H., [et al]. // Infection and immunity. – 2001. – V. 69, № 9. – P. 5430-5439.
200. **Shainheit, M. G.** The core promoter of the capsule operon of *Streptococcus pneumoniae* is necessary for colonization and invasive disease. / M. G. Shainheit, M. Mulé, A. Camilli // Infection and immunity. – 2014. – V. 82, № 2. – P. 694-705.
201. **Yother, J.** Capsules of *Streptococcus pneumoniae* and other bacteria: paradigms for polysaccharide biosynthesis and regulation. / J. Yother // Annual review of microbiology. – 2011. – V. 65, № 1. – P. 563–581.
202. **Morona, J. K.** Tyrosine phosphorylation of CpsD negatively regulates capsular polysaccharide biosynthesis in *Streptococcus pneumoniae*. / J. K. Morona, J. C. Paton, D. C. Miller, R. Morona // Mol Microbiol. – 2000. – V. 35, № 6. – P. 1431-1442.
203. **Bender, M. H.** Positive correlation between tyrosine phosphorylation of CpsD and capsular polysaccharide production in *Streptococcus pneumoniae*. / M. H. Bender, R. T. Cartee, J. Yother // J Bacteriol. – 2003. – V. 185, № 20. – P. 6057-6066.
204. **Morona, J. K.** *Streptococcus pneumoniae* capsule biosynthesis protein CpsB is a novel manganese-dependent phosphotyrosine-protein phosphatase. / J. K. Morona, R. Morona, D. C. Miller, J. C. Paton // J Bacteriol. – 2002. – V. 184, № 2. – P. 577-583.



205. **Morona, J. K.** The effect that mutations in the conserved capsular polysaccharide biosynthesis genes *cpsA*, *cpsB*, and *cpsD* have on virulence of *Streptococcus pneumoniae*. / J. K. Morona, D. C. Miller, R. Morona, J. C. Paton // The Journal of Infectious Diseases. – 2004. – V. 189, № 10. – P. 1905-1913.
206. **Morona, J. K.** Attachment of capsular polysaccharide to the cell wall of *Streptococcus pneumoniae* type 2 is required for invasive disease. / J. K. Morona, R. Morona, J. C. Paton // PNAS. – 2006. – V. 103, № 22. – P. 8505-8510.
207. **Standish, A. J.** Chemical inhibition of bacterial protein tyrosine phosphatase suppresses capsule production. / Standish A. J., Salim A. A., Zhang H. [et al]. // PloS one. – 2012. – V. 7, № 5. – P. e36312.
208. **Morona, R.** Sequence-structure relationships in polysaccharide co-polymerase (PCP) proteins. / Morona R., Purins L., Tocilj A., [et al]. // Trends in Biochemical Sciences. – 2009. – V. 34, № 2. – P. 78–84.
209. **Byrne, J. P.** Identification of *Streptococcus pneumoniae* Cps2C residues that affect capsular polysaccharide polymerisation, cell wall ligation and Cps2D phosphorylation. / J. P. Byrne, J. K. Morona, J. C. Paton, R. Morona // J Bacteriol. – 2011. – V. 193, № 9. – P. 2341-2346.
210. **Hardy, G. G.** Capsule biosynthesis and basic metabolism in *Streptococcus pneumoniae* are linked through the cellular phosphoglucomutase. / G. G. Hardy, M. J. Caimano, J. Yother // Journal of bacteriology. – 2000. – V. 182, № 7. – P. 1854-1863.
211. **Ventura, C. L.** Control of capsular polysaccharide chain length by UDP-sugar substrate concentrations in *Streptococcus pneumoniae*. / C. L. Ventura, R. T. Cartee, W. T. Forsee, J. Yother // Molecular microbiology. – 2006. – V. 61, № 3. – P. 723-733.
212. **Forsee, W. T.** A kinetic model for chain length modulation of *Streptococcus pneumoniae* celluburonan capsular polysaccharide by nucleotide sugar donor concentrations. / W. T. Forsee, R. T. Cartee, J. Yother // The Journal of Biological Chemistry. – 2009. – V. 284, № 18. – P. 11836-11844.
213. **Bender, M. H.** The atypical amino-terminal LPNTG-containing domain of the pneumococcal human IgA1-specific protease is required for proper enzyme localization and function. / M. H. Bender, J. N. Weiser // Molecular microbiology. – 2006. – V. 61, № 2. – P. 526-543.

214. **Hakenbeck, R.** Versatility of choline metabolism and choline-binding proteins in *Streptococcus pneumoniae* and commensal streptococci. / R. Hakenbeck, A. Madhour, D. Denapaite, R. Brückner // FEMS Microbiol Rev. – 2009. – V. 33. – P. 572-586.
215. **Vollmer, W.** Identification of the teichoic acid phosphorylcholine esterase in *Streptococcus pneumoniae*. / W. Vollmer, A. Tomasz // Molecular microbiology. – 2001. – V. 39, № 6. – P. 1610-1622.
216. **Ring, A.** Pneumococcal trafficking across the blood-brain barrier. Molecular analysis of a novel bidirectional pathway. / A. Ring, J. N. Weiser, E. I. Tuomanen // J Clin Invest. – 1998. – V. 102, № 2. – P. 347-360.
217. **Kostyukova, N. N.** A study of pathogenic factors of *Streptococcus pneumoniae* strains causing meningitis. / N. N. Kostyukova, M. O. Volkova, V. V. Ivanova, A. S. Kvetnaya // FEMS Immunology & Medical Microbiology. – 1995. – V. 10, № 2. – P. 133-137.
218. **Volkova, M. O.** The role of hyaluronidase in the occurrence of a generalized pneumococcal infection. / M. O. Volkova, N. N. Kostiukova, A. S. Kvetnaia // Zh Mikrobiol Epidemiol Immunobiol. – 1994. – V. 1. – P. 118-122.
219. **Gillespie, S. H.** Pathogenesis of pneumococcal infection. / S. H. Gillespie, I. Balakrishnan // Journal of medical microbiology. – 2000. – V. 49, № 12. – P. 1057-1067.
220. **King, S. J.** Deglycosylation of human glycoconjugates by the sequential activities of exoglycosidases expressed by *Streptococcus pneumoniae*. / S. J. King, K. R. Hippe, J. N. Weiser // Molecular microbiology. – 2006. – V. 59, № 3. – P. 961-974.
221. **Cámara, M.** A neuraminidase from *Streptococcus pneumoniae* has the features of a surface protein. / M. Cámara, G. J. Boulnois, P. W. Andrew, T. J. Mitchell // Infection and immunity. – 1994. – V. 62, № 9. – P. 3688-3695.
222. **Zahner, D.** The *Streptococcus pneumoniae* beta-galactosidase is a surface protein. / D. Zahner, R. Hakenbeck // Journal of bacteriology. – 2000. – V. 182, № 20. – P. 5919-5921.
223. **Berry, A. M.** Sequence heterogeneity of PsaA, a 37-kilodalton putative adhesin essential for virulence of *Streptococcus pneumoniae*. / A. M. Berry, J. Paton // Infection and immunity. – 1996. – V. 64, № 12. – P. 5255-5262.
224. **Rajam, G.** Pneumococcal surface adhesin A (PsaA): a review. / Rajam G., Anderton J. M., Carlone G. M. [et al]. // Critical Reviews in Microbiology. – 2008. – V. 34. – P. 131-142.

225. **Balachandran, P.** Role of pneumococcal surface protein C in nasopharyngeal carriage and pneumonia and its ability to elicit protection against carriage of *Streptococcus pneumoniae*. / Balachandran P., Brooks-Walter A., Virolainen-Julkunen A. [et al]. // Infection and immunity. – 2002. – V. 70, № 5. – P. 2526-2534.
226. **Zhang, J. R.** The polymeric immunoglobulin receptor translocates pneumococci across human nasopharyngeal epithelial cells. / Zhang J. R., Mostov K. E., Lamm M. E. [et al]. // Cell. – 2000. – V. 102. – P. 827-837.
227. **Phalipon, A.** Novel functions of the polymeric Ig receptor: well beyond transport of immunoglobulins. / A. Phalipon, B. Corthe // TRENDS in Immunology. – 2003. – V. 24, № 2. – P. 55-58.
228. **Hammerschmidt, S.** Impact of pneumococcal microbial surface components recognizing adhesive matrix molecules on colonization. / S. Hammerschmidt // Molecular Oral Microbiology. – 2012. – V. 27, № 4. – P. 246-256.
229. **Orihuela, C. J.** Tissue-specific contributions of pneumococcal virulence factors to pathogenesis. / Orihuela C. J., Gao G., Francis K. P. [et al]. // The Journal of Infectious Diseases. – 2004. – V. 190, № 9. – P. 1661-1669.
230. **Pracht, D.** PavaA of *Streptococcus pneumoniae* modulates adherence, invasion, and meningeal inflammation. / Pracht D., Elm C., Gerber J. [et al]. // Infection and immunity. – 2005. – V. 73, № 5. – P. 2680-2689.
231. **Hermans, P. W. M.** The *Streptococcal* lipoprotein rotamase A (SlrA) is a functional peptidyl-prolyl isomerase involved in pneumococcal colonization. / Hermans P. W. M., Adrian P. V., Albert C., [et al]. // The Journal of Biological Chemistry. – 2006. – V. 281, № 2. – P. 968-976.
232. **Barocchi, M. A.** A pneumococcal pilus influences virulence and host. / Barocchi M. A., Ries J., Zogaj X., Hemsley C. [et al]. // PNAS. – 2006. – V. 103, № 8. – P. 2857-2862.
233. **Hacker, J.** Pathogenicity islands and the evolution of microbes. / J. Hacker, J. B. Kaper // Annu Rev Microbiol. – 2000. – V. 54. – P. 641-679.
234. **Aguiar, S. I.** The presence of the *pilus* locus is a clonal property among pneumococcal invasive isolates. / Aguiar S. I., Serrano I., Pinto F. R., et al. // BMC Microbiol. – 2008. – V. 8. – P. 41.

235. **Hava, D. L.** Transcriptional regulation in the *Streptococcus pneumoniae* rlrA pathogenicity islet by RlrA. / D. L. Hava, C. J. Hemsley, A. Camilli // Journal of bacteriology. – 2003. – V. 185, № 2. – P. 413-421.
236. **Contreras-Martel, C.** Article structural basis of host cell recognition by the pilus adhesin from *Streptococcus pneumoniae*. / C. Contreras-Martel, L. E. Mortaji, C. Manzano // Structure. – 2010. – V. 18. – P. 106-115.
237. **Nelson, A. L.** RrgA is a pilus-associated adhesin in *Streptococcus pneumoniae*. / Nelson A. L., Ries J., Bagnoli F., Dahlberg S., [et al]. // Molecular microbiology. – 2007. – V. 66, № 2. – P. 329-340.
238. **Bergmann, S.** Identification of a novel plasmin(ogen)-binding motif in surface displayed alpha-enolase of *Streptococcus pneumoniae*. / Bergmann S., Wild D., Diekmann O. [et al]. // Molecular microbiology. – 2003. – V. 49, № 2. – P. 411-423.
239. **Ehinger, S.** Plasmin(ogen)-binding alpha-enolase from *Streptococcus pneumoniae*: crystal structure and evaluation of plasmin(ogen)-binding sites. / Ehinger S., Schubert W.-D., Bergmann S. [et al]. // Journal of molecular biology. – 2004. – V. 343, № 4. – P. 997-1005.
240. **Kolberg, J.** *Streptococcus pneumoniae* enolase is important for plasminogen binding despite low abundance of enolase protein on the bacterial cell surface. / Kolberg J., Aase A., Bergmann S. [et al]. // Microbiology. – 2006. – V. 152, № 5. – P. 1307-1317.
241. **Bergmann, S.** The nine residue plasminogen-binding motif of the pneumococcal enolase is the major cofactor of plasmin-mediated degradation of extracellular matrix, dissolution of fibrin and transmigration. / S. Bergmann, M. Rohde, K. T. Preissner, S. Hammerschmidt // Thrombosis and Haemostasis. – 2005. – V. 94. – P. 304-311.
242. **Tilley, S. J.** Structural basis of pore formation by the bacterial toxin pneumolysin. / Tilley S. J., Orlova E. V., Gilbert R. J. C., [et al]. // Cell. – 2005. – V. 121, № 2. – P. 247-256.
243. **Kislak, J. W.** Susceptibility of pneumococci to nine antibiotics. / J. W. Kislak, L. M. Razavi, A. K. Daly, M. Finland. // Am J Med Sci. – 1965. – V. 250. – P. 261–268.
244. **Laible, G.** Interspecies recombinational events during the evolution of altered PBP 2x genes in penicillinresistant clinical isolates of *Streptococcus pneumoniae*. / G. Laible, B. G. Spratt, R. Hakenbeck // Mol Microbiol. – 1991. – V. 5. – P. 1993–2002.

245. **Zighelboim, S.** Penicillin-binding proteins of multiply antibiotic-resistant South African strains of *Streptococcus pneumoniae*. / S. Zighelboim, A. Tomasz // Antimicrob Agents Chemother. – 1980. – V. 17. – P. 434-442.
246. **du Plessis, M.** Analysis of penicillin-binding protein genes of clinical isolates of *Streptococcus pneumoniae* with reduced susceptibility to amoxicillin. / M. du Plessis, E. Bingen, K. P. Klugman // Antimicrob Agents Chemother. – 2002. – V. 46. – P. 2349–2357.
247. **Metcalf, B. J.** Using whole genome sequencing to identify resistance determinants and predict antimicrobial resistance phenotypes for year 2015 invasive pneumococcal disease isolates recovered in the United States. / Metcalf B. J., Chochua S., Gertz Jr. R. E. [et al]. // Clin Microbiol Infect. – 2016. – V. 22. – P. 1002.e1–1002.e8.
248. **Liñares, J.** Changes in antimicrobial resistance, serotypes and genotypes in *Streptococcus pneumoniae* over a 30-year period. / J. Liñares, C. Ardanuy, R. Pallares, A. Fenoll // Clinical Microbiology and Infection. – 2010. – V. 16, № 5. – P. 402-410.
249. **Felmingham, D.** The Alexander Project: the benefits from a decade of surveillance. / Felmingham D., White A. R., Jacobs M. R. [et al]. // J Antimicrob Chemother. – 2005. – V. 56. – P. ii3-ii21.
250. **Hicks, L. A.** Outpatient antibiotic prescribing and nonsusceptible *Streptococcus pneumoniae* in the United States, 1996-2003. / Hicks L. A., Chien Y.-W., Taylor Jr. T. H. [et al]. // Clin Infect Dis. – 2011. – V. 53. – P. 631-639.
251. **Riedel, S.** Antimicrobial use in Europe and antimicrobial resistance in *Streptococcus pneumoniae*. / Riedel S., Beekmann S. E., Heilmann K. P. [et al]. // Eur J Clin Microbiol Infect Dis. – 2007. – V. 26. – P. 485-490.
252. **Wyres, K.** Genome Evolution in *Streptococcus pneumoniae*. / Wyres K. – Genome Evolution in *Streptococcus pneumoniae*. – 2012.
253. **Whitney, C. G.** Decline in invasive pneumococcal disease after the introduction of protein-polysaccharide conjugate vaccine. / Whitney C. G. Farley M. M., Hadler J. [et al]. // N Engl J Med. – 2003. – V. 348. – P. 1737-1746.
254. **Gertz, R. E.** Increased penicillin nonsusceptibility of nonvaccine-serotype invasive pneumococci other than serotypes 19A and 6A in post-7-valent Conjugate vaccine era. / Gertz R. E., Li Z., Pimenta F. C. [et al]. // J Infect Dis. – 2010. – V. 201. – P. 770-775.
255. **Mera, R. M.** The impact of the pneumococcal conjugate vaccine on antimicrobial resistance in the United States since 1996: Evidence for a significant rebound by 2007 in many

- classes of antibiotics. / R. M. Mera, L. A. Miller, H. Amrine-Madsen, D. F. Sahn // *Microb Drug Resist.* – 2009. – V. 15. – P. 261-268.
256. **Jones, R. N.** Declining antimicrobial susceptibility of *Streptococcus pneumoniae* in the United States: report from the SENTRY antimicrobial surveillance program (1998-2009). / R. N. Jones, H. S. Sader // *Diagn Microbiol Infect Dis.* – 2010. – V. 68. – P. 334-336.
257. **Kaplan, S. L.** Serotype 19A Is the most common serotype causing invasive pneumococcal infections in children. / Kaplan S. L., Barson W. J., Lin P. L. [et al]. // *Pediatrics.* – 2010. – V. 125, № 3. – P. 429-436.
258. **Shin, J.** Predominance of ST320 among *Streptococcus pneumoniae* serotype 19A isolates from 10 Asian countries. / Shin J., Baek J. Y., Kim S. H. [et al]. // *Journal of Antimicrobial Chemotherapy.* – 2011. – V. 66, № 5. – P. 1001-1004.
259. **Murray, P. R.** Antimicrobial agents, in *Manual of clinical microbiology.* / P. R. Murray, E. J. Baron, M. A. Pfaller, F. C. Tenover, and R. H. Tenover (ed.). – *Manual of clinical microbiology* – 8th edition. ASM Press, Washington D. C., 2003.
260. **Pagliari, E.** Biochemical characterization of *Streptococcus pneumoniae* penicillin-binding protein 2b and its implication in beta-lactam resistance. / Pagliero E., Chesnel L., Hopkins J. [et al]. // *Antimicrob Agents Chemother.* – 2004. – V. 48. – P. 1848-1855.
261. **Cafini, F.** Alterations of the penicillin-binding proteins and *murM* alleles of clinical *Streptococcus pneumoniae* isolates with high-level resistance to amoxicillin in Spain. / Cafini F., del Campo R., Alou L. [et al]. // *J Antimicrob Chemother.* – 2006. – V. 57. – P. 224-229.
262. / **Carapito, R.** Pneumococcal b-Lactam resistance due to a conformational change in penicillin-binding protein 2x. / R. Carapito, L. Chesnel, T. Vernet, A. Zapun // *J Biol Chem.* – 2006. – V. 281. – P. 1771-1777.
263. **Chesnel, L.** Identical penicillin-binding domains in penicillin-binding proteins of *Streptococcus pneumoniae* clinical isolates with different levels of b-lactam resistance. / Chesnel L., Carapito R., Croizé J. [et al]. // *Antimicrob Agents Chemother.* – 2005. – V. 49. – P. 2895- 2902.
264. **Dias, R.** Diversity of penicillin binding proteins among clinical *Streptococcus pneumoniae* strains from Portugal. / R. Dias, D. Felix, M. Canica // *Antimicrob Agents Chemother.* – 2008. – V. 52. – P. 2693-2695.

265. **Ferroni, A.** Alterations to penicillin-binding proteins 1A, 2B and 2X amongst penicillin-resistant clinical isolates of *Streptococcus pneumoniae* serotype 23F from the nasopharyngeal flora of children. / A. Ferroni, P. Berche // J Med Microbiol. – 2001. – V. 50. – P. 828-832.
266. **Lambertsen, L.** Molecular characterisation of invasive penicillin non-susceptible *Streptococcus pneumoniae* from Denmark, 2001 to 2005. / L. Lambertsen, M. Brendstrup, H. Friis, J. J. Christensen // Scand J Infect Dis. – 2010. – V. 42. – P. 333-340.
267. **Overweg, K.** Molecular characteristics of penicillin-binding protein genes of penicillin-nonsusceptible *Streptococcus pneumoniae* isolated in The Netherlands. / Overweg K., Bogaert D., Sluijter M. [et al]. // Microb Drug Resist. – 2001. – V. 7. – P. 323-334.
268. **Sanbongi, Y.** Complete sequences of six penicillin-binding protein genes from 40 *Streptococcus pneumoniae* clinical isolates collected in Japan. / Sanbongi Y., Ida T., Ishikawa M. [et al]. // Antimicrob Agents Chemother. – 2004. – V. 48. – P. 2244-2250.
269. **Stanhope, M. J.** Positive selection in penicillin-binding proteins 1a, 2b, and 2x from *Streptococcus pneumoniae* and its correlation with amoxicillin resistance development. / Stanhope M. J., Lefébure T., Walsh S. L. [et al]. // Infect Genet Evol. – 2008. – V. 8. – P. 331-339.
270. **Zhanel, G. G.** Molecular characterisation of Canadian paediatric multidrug-resistant *Streptococcus pneumoniae* from 1998-2004. / Zhanel G. G., Wang X., Nichol K. [et al]. // International journal of antimicrobial agents. – 2006. – V. 28. – P. 465-471.
271. **Sibold, C.** Mosaic *pbpX* genes of major clones of penicillin-resistant *Streptococcus pneumoniae* have evolved from *pbpX* genes of a penicillin-sensitive *Streptococcus oralis*. / Sibold C., Henrichsen J., König A., [et al]. // Molecular microbiology. – 1994. – V. 12, № 6. – P. 1013-1023.
272. **Dowson, C. G.** Evolution of penicillin resistance in *Streptococcus pneumoniae*; the role of *Streptococcus mitis* in the formation of a low affinity PBP2B in *S. pneumoniae*. / C. G. Dowson, T. J. Coffey, C. Kell, R. A. Whiley // Molecular microbiology. – 1993. – V. 9, № 3. – P. 635-643.
273. **Maurer, P.** Penicillin-binding protein 2x of *Streptococcus pneumoniae*: three new mutational pathways for remodelling an essential enzyme into a resistance determinant. / Maurer P., Koch B., Zerfass I., [et al]. // J Mol Biol. – 2008. – V. 376. – P. 1403-1416.

274. **Grebe, T.** Penicillin-binding proteins 2b and 2x of *Streptococcus pneumoniae* are primary resistance determinants for different classes of  $\beta$ -lactam antibiotics. / T. Grebe, R. Hakenbeck // Antimicrob Agents Chemother. – 1996. – V. 40. – P. 829-834.
275. **McDougal, L. K.** Identification of multiple clones of extended-spectrum cephalosporin-resistant *Streptococcus pneumoniae* isolates in the United States. / L. K. McDougal, J. K. Rasheed, J. W. Biddle, F. C. Tenover // Antimicrobial agents and chemotherapy. – 1995. – V. 39, № 10. – P. 2282-2288.
276. **Smith, A. M.** Alterations in PBP 1A essential for high-level penicillin resistance in *Streptococcus pneumoniae*. / A. M. Smith, K. P. Klugman // Antimicrob Agents Chemother. – 1998. – V. 42. – P. 1329-1333.
277. **Filipe, S. R.** The role of the murMN operon in penicillin resistance and antibiotic tolerance of *Streptococcus pneumoniae*. / S. R. Filipe, E. Severina, A. Tomasz // Microb Drug Resist. – 2001. – V. 7. – P. 303-316.
278. **Weber, B.** The *fib* locus in *Streptococcus pneumoniae* is required for peptidoglycan crosslinking and the PBP mediated  $\beta$ -lactam resistance. / Weber B., Ehlert K., Diehl A. [et al]. // FEMS Microbiol Lett. – 2000. – V. 188. – P. 81-85.
279. **Mascher, T.** The CiaRH system of *Streptococcus pneumoniae* prevents lysis during stress induced by treatment with cell wall inhibitors and by mutations in *pbp2x* involved in  $\beta$ -lactam resistance. / Mascher T., Heintz M., Zähler D. [et al]. // J Bacteriol. – 2006. – V. 188. – P. 1959-1968.
280. **Tran, T. D.** Decrease in penicillin susceptibility due to heat shock protein ClpL in *Streptococcus pneumoniae*. / Tran T. D., Kwon H.-Y., Kim E.-H. et al. // Antimicrob Agents Chemother. – 2011. – V. 55. – P. 2714-2728.
281. European Antimicrobial Resistance Surveillance Network (EARS-Net) database. European Centre for Disease Prevention and Control / EARS-Net, ECDC // <https://www.ecdc.europa.eu> : Медицина. Медицинский сайт: URL: <https://www.ecdc.europa.eu/en/about-us/partnerships-and-networks/disease-and-laboratory-networks/ears-net> (дата обращения: 20.08.2021).
282. **Ashley, E. A.** Antimicrobial susceptibility of bacterial isolates from community acquired infections in Sub-Saharan Africa and Asian low and middle income countries. / E. A. Ashley, Y. Lubell, N. J. White, P. Turner // Trop Med Int Health. – 2011. – V. 16. – P. 1167-1179.



283. **Kim, S. H.** Changing trends in antimicrobial resistance and serotypes of *Streptococcus pneumoniae* isolates in Asian countries: an Asian Network for Surveillance of Resistant Pathogens (ANSORP) study. / Kim S. H., Song J.-H., Chung D. R., [et al]. // Antimicrob Agents Chemother. – 2012. – V. 56. – P. 1418-1426.
284. **Farrell, D. J.** Non-susceptibility trends and serotype distributions among *Streptococcus pneumoniae* from community-acquired respiratory tract infections and from bacteraemias in the UK and Ireland, 1999 to 2007. / Farrell D. J., Felmingham D., Shackcloth J. [et al]. // J Antimicrob Chemother. – 2008. – V. 62. – P. ii87-ii95.
285. **Jones, M. E.** In vitro susceptibility of *Streptococcus pneumoniae*, *Haemophilus influenzae* and 73 *Moraxella catarrhalis*: a European multicentre study during 2000-2001. / Jones M. E., Blosser-Middleton R., Critchley I. [et al]. (2003). // Clin Microbiol Infect. – 2003. – V. 9. – P. 590-599.
286. **Lee, N. Y.** Carriage of antibiotic-resistant pneumococci among Asian children: a multinational surveillance by the Asian Network for Surveillance of Resistant Pathogens (ANSORP). / Lee N. Y., J.-H. Song, S. Kim [et al]. // Clin Infect Dis. – 2001. – V. 32. – P. 1463-1469.
287. **Castanheira, M.** Antimicrobial susceptibility of *Streptococcus pneumoniae* in Latin America: results from five years of the SENTRY antimicrobial surveillance program. / Castanheira M., Gales A., Mendes R. [et al]. // Clin Microbiol Infect. – 2004. – V. 10. – P. 645-651.
288. **Jacobs, M. R.** The Alexander Project 1998-2000: susceptibility of pathogens isolated from community-acquired respiratory tract infection to commonly used antimicrobial agents. / Jacobs M. R., Felmingham D., Appelbaum P. C. [et al]. // J Antimicrob Chemother. – 2003. – V. 52. – P. 229-246.
289. **Tuomanen, E. I.** Macrolide, quinolone, and other non-beta-lactam antibiotic resistance in *Streptococcus pneumoniae*. In The Pneumococcus. / E. I. Tuomanen, T. J. Mitchell, D. A. Morrison, B. G. Spratt The Pneumococcus. – ASM Press. – Washington, USA. – 427 pp.
290. **Weisblum, B.** Erythromycin resistance by ribosome modification. / B. Weisblum // Antimicrob Agents Chemother. – 1995. – V. 39. – P. 577-585.
291. **Tait-Kamradt, A.** *mefE* is necessary for the erythromycin-resistant M phenotype in *Streptococcus pneumoniae*. / Tait-Kamradt A., Clancy J., Cronan M., [et al]. // Antimicrob Agents Chemother. – 1997. – V. 41. – P. 2251-2255.

292. **Doherty, N.** Genetic diversity of the *tet(M)* gene in tetracycline-resistant clonal lineages of *Streptococcus pneumoniae*. / Doherty N., Trzcinski K., Pickerill P. [et al]. // Antimicrob Agents Chemother. – 2000. – V. 44. – P. 2979-2984.
293. **Burdett, V.** Tet(M)-promoted release of tetracycline from ribosomes is GTP dependent. / B. Weisblum // J Bacteriol. – 1996. – V. 178. – P. 3246-3251.
294. **Ayoubi, P.** Tn5253, the pneumococcal *omega (cat tet)* BM6001 element, is a composite structure of two conjugative transposons, Tn5251 and Tn5252. / P. Ayoubi, A. O. Kilic, M. N. Vijayakumar // J Bacteriol. – 1991. – V. 173. – P. 1617-1622.
295. **Ding, F.** Genome evolution driven by host adaptations results in a more virulent and antimicrobial-resistant *Streptococcus pneumoniae* serotype 14. / Ding F., Tang P., Hsu M.-H. [et al]. // BMC Genet. – 2009. – V. 10. – P. 158.
296. **Mingoia, M.** Heterogeneity of Tn5253-like composite elements in clinical *Streptococcus pneumoniae* isolates. / Mingoia M., Tili E., Manso E. [et al]. // Antimicrob Agents Chemother. – 2011. – V. 55. – P. 1453-1459.
297. **Evans, W.** Tetracycline-resistant pneumococcus. / W. Evans, D. Hansman // The Lancet. – 1963. – V. 451.
298. **Roberts, A. P.** Tn916-like genetic elements: a diverse group of modular mobile elements conferring antibiotic resistance. / A. P. Roberts, P. Mullany // FEMS Microbiol. Rev. – 2011. – V. 35. – P. 856-871.
299. **Cochetti, I.** *erm(B)*-carrying elements in tetracycline-resistant pneumococci and correspondence between Tn1545 and Tn6003. / Cochetti I., Tili E., Mingoia M., [et al]. // Antimicrob Agents Chemother. – 2008. – V. 52. – P. 1285-1290.
300. **Del Grosso, M.** Tn2009, a Tn916-like element containing *mef(E)* in *Streptococcus pneumoniae*. / Del Grosso M., d'Abusco A. S., Iannelli F. [et al]. // Antimicrob Agents Chemother. – 2004. – V. 48. – P. 2037-2042.
301. **Santoro, F.** Nucleotide sequence and functional analysis of the *tet(M)*-carrying conjugative transposon Tn5251 of *Streptococcus pneumoniae*. / F. Santoro, M. R. Oggioni, G. Pozzi, F. Iannelli // FEMS Microbiol Lett. – 2010. – V. 308. – P. 150-158.
302. **Bobay, L. M.** The evolution of bacterial genome architecture. / L. M. Bobay, H. Ochman // Front Genet. – 2017. – V. 8. – P. 72.

303. **Gladstone, R. A.** International genomic definition of pneumococcal lineages, to contextualise disease , antibiotic resistance and vaccine impact. / Gladstone R. A., L.o S. W., Lees J. A. [et al]. // EBioMedicine. – 2019. – V. 43. – P. 338-346.
304. **Treangen, T. J.** Rapid core-genome alignment and visualization for thousands of microbial genomes. / T. J. Treangen, B. D. Ondov, S. Koren, A. M. Phillippy // Genome Biol. – 2014. – V. 15. – P. 524.
305. **Stamatakis, A.** RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. / A. Stamatakis // Bioinformatics. – 2014. – V. 30, № 9. – P. 1312-1313.
306. **Letunic, I.** Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. / I. Letunic, P. Bork // Nucleic Acids Research. – 2016. – V. 44, № W1. – P. W242-W245.
307. **Tonkin-Hill, G.** RhierBAPS: An R implementation of the population clustering algorithm hierBAPS. / Tonkin-Hill G., Lees J. A., Bentley S. D. [et al]. // Wellcome Open Research. – 2018. – V. 3, № 0. – P. 93.
308. **Huson, D. H.** Application of phylogenetic networks in evolutionary studies. / D. H. Huson, D. Bryant // Molecular Biology and Evolution. – 2006. – V. 23, № 2. – P. 254-267.
309. **Silva, A. G.** Pairwise SNP difference summary. / A. G. Silva. // <https://github.com> : Биология. Биологический сайт : URL: [https://github.com/MDU-PHL/pairwise\\_snp\\_differences/blob/master/pairwise\\_snp\\_differences.Rmd](https://github.com/MDU-PHL/pairwise_snp_differences/blob/master/pairwise_snp_differences.Rmd) (дата обращения: 20.08.2021).
310. **Simon, A.** FastQC: a quality control tool for high throughput sequence data. / A. Simon // <https://github.com> : Биология. Биологический сайт : URL: <https://github.com/s-andrews/FastQC> (дата обращения: 20.08.2021).
311. **Bolger, A. M.** Trimmomatic: A flexible trimmer for Illumina sequence data. / A. M. Bolger, M. Lohse, B. Usadel // Bioinformatics. – 2014. – V. 30, № 15. – P. 2114-2120.
312. **Bankevich, A.** SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. / Bankevich A., Nurk S., Antipov D. [et al]. // Journal of Computational Biology. – 2012. – V. 19, № 5. – P. 455-477.
313. **Nurk, S.** Assembling genomes and mini-metagenomes from highly chimeric reads. / Nurk S., Bankevich A., Antipov D. [et al]. // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). – 2013. – V. 7821 LNBI. – P. 158-180.

314. **Gurevich, A.** QUASt: Quality assessment tool for genome assemblies. / A. Gurevich, V. Saveliev, N. Vyahhi, G. Tesler // *Bioinformatics*. – 2013. – V. 29, № 8. – P. 1072-1075.
315. **Jolley, K. A.** BIGSdb: Scalable analysis of bacterial genome variation at the population level. / K. A. Jolley, M. C. Maiden // *J. BMC Bioinformatics*. – 2010. – V. 11, № 1. – P. 595.
316. **Seemann, T.** mlst / T. Seemann // <https://github.com> : Биология. Биологический сайт : URL: <https://github.com/tseemann/mlst> (дата обращения: 20.08.2021).
317. **Brettin, T.** RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. / Brettin T., Davis J. J., Disz T. [et al]. // *Sci Rep*. – 2015. – V. 5. – P. 8365.
318. **Edgar, R. C.** MUSCLE: Multiple sequence alignment with high accuracy and high throughput. / R. C. Edgar // *Nucleic Acids Research*. – 2004. – V. 32, № 5. – P. 1792-1797.
319. **Bruen, T. C.** A simple and robust statistical test for detecting the presence of recombination. / T. C. Bruen, H. Philippe, D. Bryant // *Genetics*. – 2006. – V. 172, № 4. – P. 2665-2681.
320. **Croucher, N. J.** Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins / Croucher N. J., Page A. J., Connor T. R. [et al]. // *Nucleic Acids Res*. – 2015. – V. 43, № 3. – P. e15.
321. **Hadfield, J.** Phandango: An interactive viewer for bacterial population genomics. / Hadfield J., Croucher N. J., Goater R. J., [et al]. // *Bioinformatics*. – 2018. – V. 34, № 2. – P. 292-293.
322. **Breiman, L.** Random forests. / L. Breiman // *Machine learning*. – 2001. – V. 45, № 1. – P. 5-32.
323. **Chen, T.** XGBoost: A Scalable Tree Boosting System. / T. Chen, C. Guestrin // *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. – August 2016. – Pages 785–794.
324. The H2O Python Module. / H2O.ai Documentation. // <https://docs.h2o.ai> : Программирование. Сайт. – URL: <https://docs.h2o.ai/h2o/latest-stable/h2o-py/docs/intro.html>. (дата обращения: 20.08.2021).
325. **Roberts, R. J.** REBASE - a database for DNA restriction and modification: enzymes, genes and genomes. / R. J. Roberts, T. Vincze, J. Posfai, D. Macelis // *Nucleic Acids Res*. – 2015. – V. 43. – P. D298–D299.

326. **Arndt, D.** PHASTER: a better, faster version of the PHAST phage search tool. / Arndt D., Grant J. R., Marcu A. [et al]. // *Nucleic Acids Res.* – 2016. – V. 44(W1). – P. W16-W21.
327. Geneious prime / Geneious prime // <https://www.geneious.com> . Биология. Биологические науки. – URL: <https://www.geneious.com> (дата обращения: 20.08.2021).
328. **Bogdanowicz, D.** Matching split distance for unrooted binary phylogenetic trees. / D. Bogdanowicz, K. Giaro // *IEEE/ACM Trans. on Comp. Biol. and Bioinformatics.* – 2012. – V. 9. – P. 150-160.
329. Visual TreeCmp. / Visual TreeCmp. // <https://eti.pg.edu.pl> . Биология. Биологические науки. – URL: <https://eti.pg.edu.pl/TreeCmp/download.html> (дата обращения: 20.08.2021).
330. **Goluch, T.** Visual TreeCmp: comprehensive comparison of phylogenetic trees on the Web. / T. Goluch, D. Bogdanowicz, K. Giaro // *Methods in Ecology and Evolution.* – 2020. – V. 11. – P. 494-499.
331. **Szklarczyk, D.** STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. / Szklarczyk D., Gable A. L., Lyon D. [et al]. // *Nucleic Acids Res.* – 2019. – V. 47. – P. D607-D613.
332. **Aiba, Y.** Mutation of RNA polymerase  $\beta$ -subunit gene promotes heterogeneous-to-homogeneous conversion of  $\beta$ -lactam resistance in methicillin-resistant *Staphylococcus aureus*. / Aiba Y., Katayama Y., Hishinuma T. [et al]. // *Antimicrob Agents Chemother.* – 2013. – V. 57, № 10. – P. 4861-4871.
333. **Panchal, V. V.** Evolving MRSA: High-level  $\beta$ -lactam resistance in *Staphylococcus aureus* is associated with RNA Polymerase alterations and fine tuning of gene expression. / Panchal V. V., Griffiths C., Mosaei H. [et al]. // *PLoS Pathog.* – 2020. – V. 16, № 7. – P. e1008672.
334. **Chaguza, C.** Recombination in *Streptococcus pneumoniae* lineages increase with carriage duration and size of the polysaccharide capsule. / Chaguza C., Andam C. P., Harris S. R. [et al]. // *mBio.* – 2016. – V. 7, № 5. – P. e01053-16.
335. **Mouz, N.** Mutations in the active site of penicillin-binding protein PBP2x from *Streptococcus pneumoniae*. Role in the specificity for beta-lactam antibiotics. / Mouz N., Guilmi A. M. Di., Gordon E. [et al]. // *J Biol Chem.* – 1999. – V. 274. – P. 19175-19180.
336. **De la Iglesia, R.** Novel polymerase chain reaction primers for the specific detection of bacterial copper P-type ATPases gene sequences in environmental isolates and metagenomic

- DNA. / De la Iglesia R., Valenzuela-Heredia D., Pavissich J. P. et al. // *Lett Appl Microbiol.* – 2010. – V. 50. – P. 552-562.
337. **Noirclerc-Savoie, M.** In vitro reconstitution of a trimeric complex of DivIB, DivIC and FtsL, and their transient co-localization at the division site in *Streptococcus pneumoniae*. / Noirclerc-Savoie M., Le Gouëllec A., Morlot C. [et al]. // *Mol Microbiol.* – 2005. – V. 55. – P. 413-424.
338. **Massidda, O.** From models to pathogens: how much have we learned about *Streptococcus pneumoniae* cell division? / O. Massidda, L. Novakova, W. Vollmer // *Environ Microbiol.* – 2013. – V. 15. – P. 3133-3157.
339. **Massidda, O.** Unconventional organization of the division and cell wall gene cluster of *Streptococcus pneumoniae*. / O. Massidda, D. Anderluzzi, L. Friedli, G. Feger // *Microbiology.* – 1998. – V. 144, № Pt 11. – P. 3069-3078.
340. **Fadda, D.** Characterization of divIVA and other genes located in the chromosomal region downstream of the *dcw* cluster in *Streptococcus pneumoniae*. / Fadda D., Pischedda C., Caldara F. [et al]. // *J Bacteriol.* – 2003. – V. 185. – P. 6209–6214.
341. **Chapuy-Regaud, S.** RegR, a global LacI/GalR family regulator, modulates virulence and competence in *Streptococcus pneumoniae*. / Chapuy-Regaud S., Ogunniyi A. D., Diallo N. [et al]. // *Infection and immunity.* – 2003. – V. May. – P. 2615–2625.
342. **Fadda, D.** *Streptococcus pneumoniae* DivIVA: localization and interactions in a MinCD-free context. / Fadda D., Santona A., D'Ulisse V. [et al]. // *J Bacteriol.* – 2007. – V. 189. – P. 1288–1298.
343. **Hamoen, L. W.** SepF, a novel FtsZ-interacting protein required for a late step in cell division. / Hamoen L. W., Meile J.-C., de Jong W. [et al]. // *Mol Microbiol.* – 2006. – V. 59. – P. 989–999.
344. **Chewapreecha, C.** Comprehensive identification of single nucleotide polymorphisms associated with beta-lactam resistance within pneumococcal mosaic genes. / Chewapreecha C., Martinen P., Croucher N. J. [et al]. // *PLOS Genetics.* – 2014. – V. 10, № 8. – P. e1004547.
345. **Chewapreecha, C.** Dense genomic sampling identifies highways of pneumococcal recombination. / Chewapreecha C., Harris S. R., Croucher N. J., [et al]. // *Nature genetics.* – 2014. – V. 46, № 3. – P. 305-309.

346. **Lees, J. A.** Genome-wide identification of lineage and locus specific variation associated with pneumococcal carriage duration. / Lees J. A., Croucher N. J., Goldblatt D. [et al]. // eLife. – 2017. – V. 6. – P. e26255.
347. **Corander, J.** Frequency-dependent selection in vaccine-associated pneumococcal population dynamics. / Corander J., Fraser C., Gutmann M. U. [et al]. // Nature Ecology and Evolution. – 2017. – V. 1, № 12. – P. 1950-1960.