

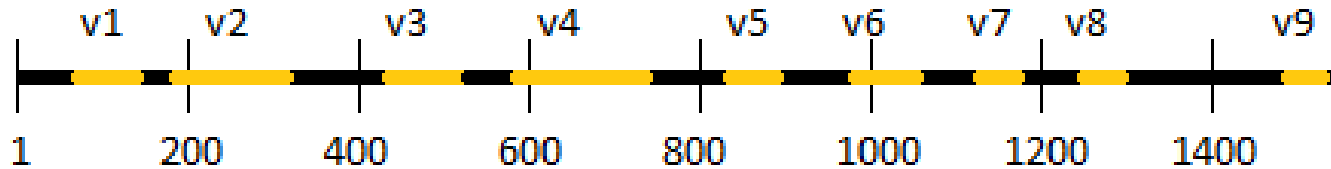
Сравнение результатов метагеномного профилирования посредством секвенирования гена 16S рРНК при использовании сборщиков *idba* и *spades*, а также классификаторов RDP и LCAClassifier

ФГБУ ФНКЦ ФХМ ФМБА России
Лаборатория постгеномных исследований в
биологии

Аспирант: Вахитова Мария
Тимуровна

Научный руководитель: Кострюкова
Елена Сергеевна

Структура гена 16S рРНК



 - переменные области

 - консервативные области

27F	AGAGTTTGATCCTGGCTCAG	V1	338R	TGCTGCCTCCCGTAGGAGT	V2
101F	AGYGGCGNACGGGTGAGTAA	V2	518R	ATTACCGCGGCTGCTGG	V3
341F	CCTACGGGAGGCAGCAG	V3	534R	ATTACCGCGGCTGCTGG	V3
357F	ACTCCTACGGGAGGCAGCAG	V3	785R	TACNVGGGTATCTAATCC	V4
515F	GTGNCAGCMGCCGCGGTAA	V4	806R	GGA CTACHVGGGTWTCTAAT	V4
563F	AYTGGGYDTAAAGNG	V4	926R	CCGYCAATTYMTTTRAGTTT	V5
803F	ATTAGATACCCNGGTAG	V5	1046R	CGACAGCCATGCANCACCT	V6
U789F	TAGATACCCSSGTAGTCC	V5	U1068R	CTGACGRCRGCCATGC	V6
967F	CAACGCGAAGAACCTTACC	V6	1099R	GYAACGAGCGCAACCC	V7
1099F	GYAACGAGCGCAACCC	V7	1407R	GACGGGCGGTGWGTRC	V8
			1492R	GNTACCTTGTTACGACTT	V9

Наиболее частоиспользуемые праймеры для амплификации различных областей гена 16S рРНК

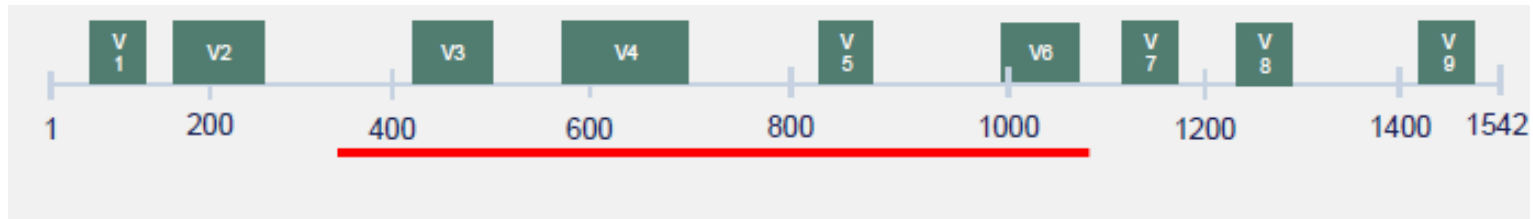
Цель настоящего исследования:

Сравнение достоверности данных о составе бактериального сообщества, полученных при сочетании различных методов компьютерного анализа.

Задачи:

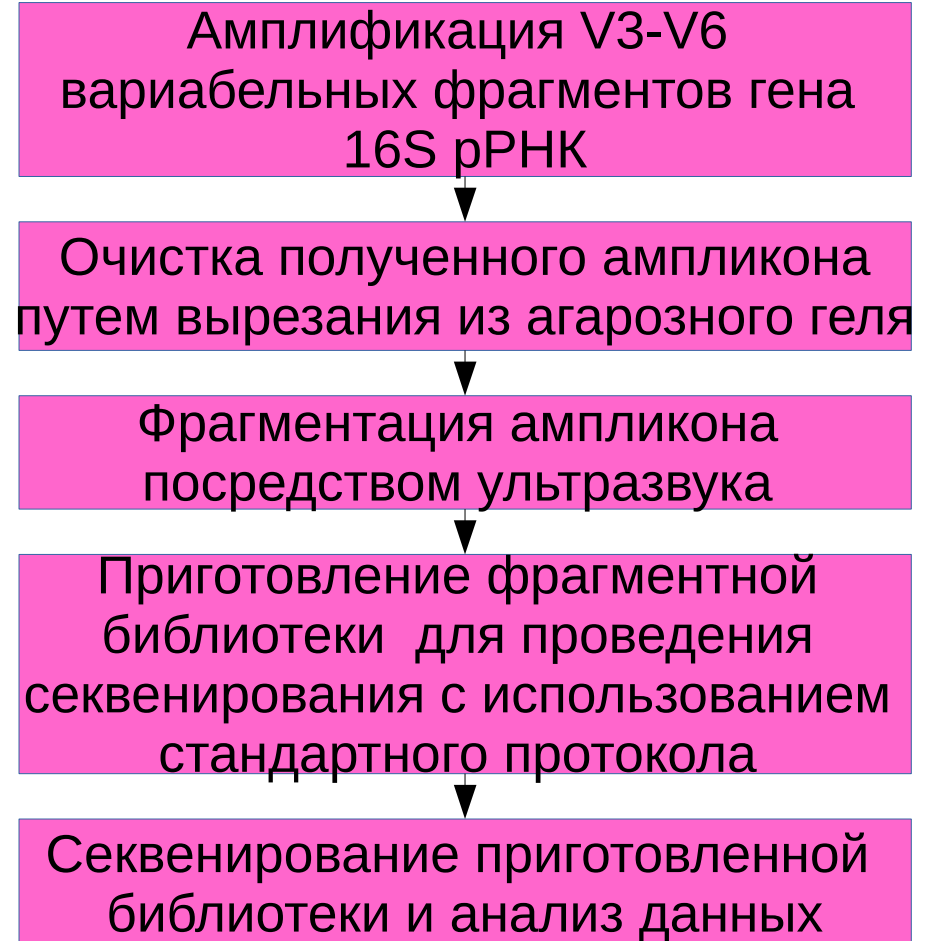
1. Приготовление и секвенирование фрагментной библиотеки ампликона гена 16S рРНК.
2. Проведение компьютерного анализа полученных данных с использованием нескольких алгоритмов.
3. Сравнение результатов и определение оптимального протокола.

Приготовление фрагментной библиотеки ампликона гена 16S рРНК



Для проведения амплификации использовались праймеры 338F и 1061R.

Фрагментация ампликона проводилась посредством воздействия на ДНК ультразвуком.



План эксперимента

Составление
«искусственного метагенома»

Аmplification гена 16S рРНК,
фрагментация ампликона,
приготовление библиотеки

Проведение секвенирования
на MiSeq

MS_R1.fastq
1 500 000 ридов

MS_R2.fastq
1 500 000 ридов

MS_million.fasta
3 000 000 ридов

MS.fasta
400 000 ридов

Проведение сборки с
использованием idba и spades

MS_million_idba.fasta
110 контигов

MS_million_spades.fasta
8227 контигов

MS_200_idba.fasta
24 контига

MS_200_spades.fasta
70 контигов

Проведение классификации в RDP и LCAClassifier

MS_million_idba_rdp

MS_million_spades_lca

MS_200_idba_rdp

MS_200_spades_lca

MS_million_idba_lca

MS_million_spades_rdp

MS_200_idba_lca

MS_200_spades_rdp

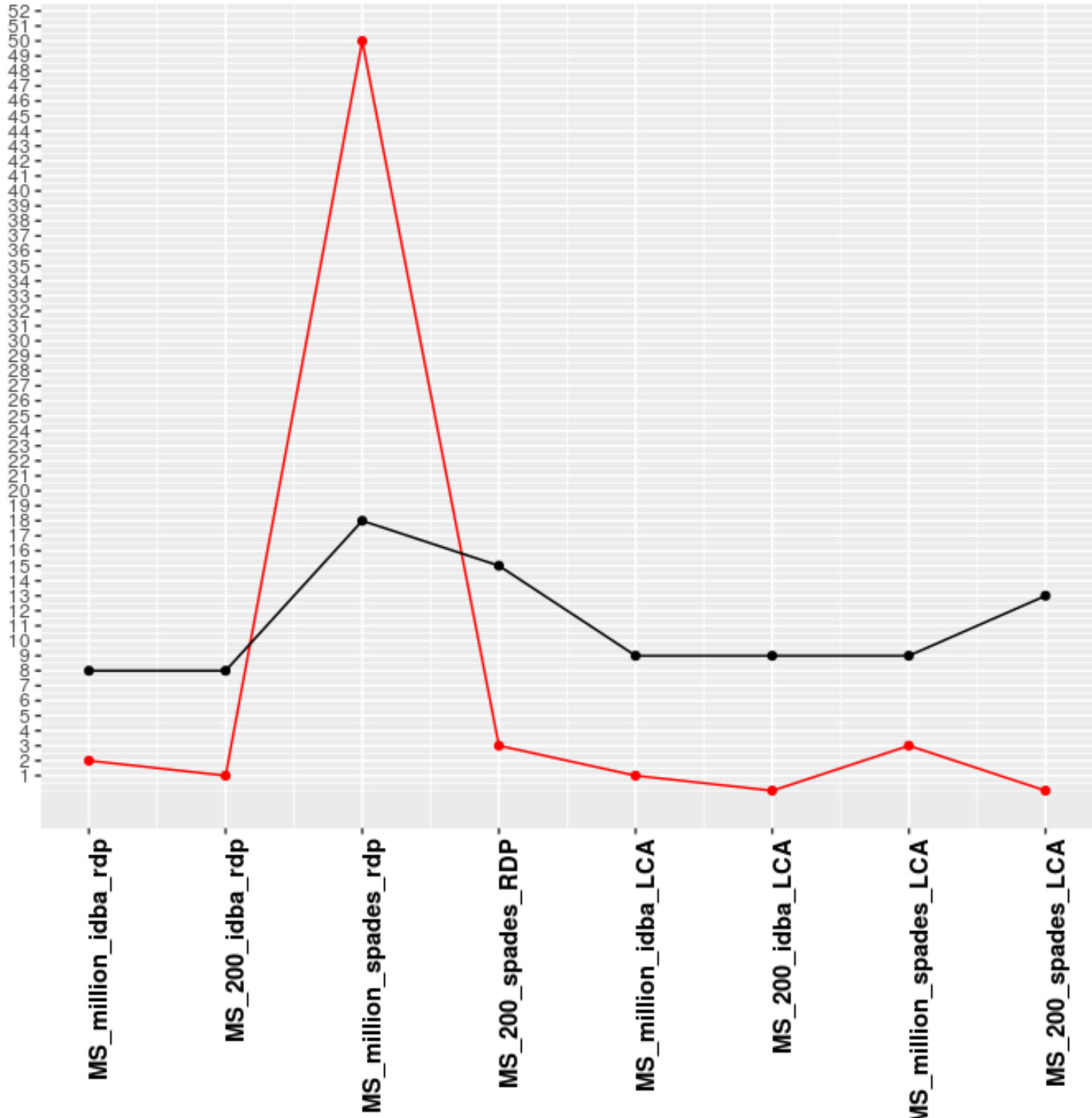
Результаты проведения классификации

	A c t i v e	A c t i v e	B a c t e r i a	B a c t e r i a	C h l o r o p h y t i c a	C h l o r o p h y t i c a	E n t o m o l o g y	E n t o m o l o g y	G e o d e r m o r p h y	H e l i c o b a c t e r i a	K y b a c t e r i a	L y b a c t e r i a	M y c o b i o t a x o n o m y	M y c o b i o t a x o n o m y	N e u r o m i c r o b i o l o g y	P s y c h o l o g y	R a t h o m y	S p h e r o p h y l o g y	S p h e r o p h y l o g y	T h e r m o p h i l o g y	V i r u l o g y	F a l s e p o s s i b l e	
MS million idba rdp	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	2
MS 200 idba rdp	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	1
MS million spades rdp	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	50
MS 200 spades rdp	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	3
MS million idba LCA	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	1
MS 200 idba LCA	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	0
MS million spades LCA	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	3
MS 200 spades LCA	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	0

■ - идентифицированные классификатором бактерии

■ - присутствовавшие в образце, но не идентифицированные классификатором бактерии

number of genres



colour

● detected bacteria

● false positive

algorithm of analysis

Результаты:

- 1) при анализе данных с избыточным покрытием фрагментированного ампликона не происходит улучшения точности идентификации присутствующих в образце бактерий, в некоторых случаях наблюдается ухудшение по сравнению с результатами анализа данных с меньшим на порядок покрытием.
- 2) при анализе данных с избыточным покрытием фрагментированного ампликона происходит увеличение количества ложноположительных результатов по сравнению с результатами анализа данных с меньшим на порядок покрытием.
- 3) применение сборщика `spades` приводит к увеличению точности определения присутствующих в образце бактерий, однако способствует увеличению количества ложноположительных результатов по сравнению с результатами, полученными при использовании сборщика `idba` и единого классификатора.
- 4) применение классификатора `RDP` приводит к увеличению точности определения присутствующих в образце бактерий, однако способствует увеличению количества ложноположительных результатов, по сравнению с результатами, полученными при использовании классификатора `LCAClassifier` и единого сборщика.

Выводы:

- 1) необходим поиск значения оптимального покрытия исследуемых последовательностей ДНК, т.к. данный показатель оказывает сильное влияние на результат и зависит от ряда параметров (тип библиотеки, бактериальное разнообразие исследуемого образца и т.д.)
- 2) не желательно совмещать сборщик `spades` и классификатор RDP при проведении анализа. Наиболее оптимальными являются сочетания `spades` и `LCAClassifier`, либо `idba` и RDP.

Спасибо за внимание!