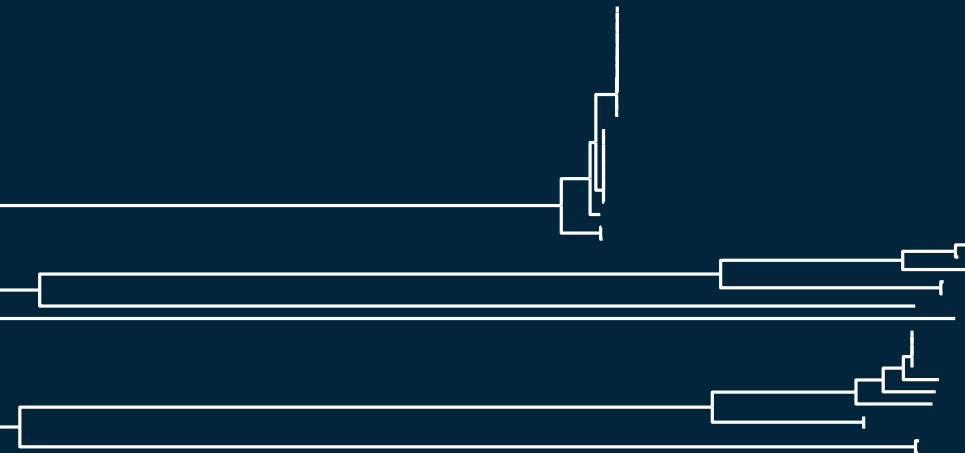




ФКУЗ Иркутский научно-исследовательский
противочумный институт Роспотребнадзора

Анализ корового генома
штаммов *Vibrio cholerae*
разной эпидемической значимости:
разработка подходов к sg-MLST и поиск
специфических молекулярных маркеров



Авторы:
Н.О. Бочалгин
А.В. Миронова, С.В. Балахонов

cgMLST (core genome multi-locus sequence typing)
мультилокусное сиквенс-типирование на основе корового генома:

Преимущества метода

- Портируемость
- Масштабируемость
- Высокая разрешающая способность
- Однозначность результатов

Крупнейшие провайдеры данных

- cgmlst.org
- enterobase.warwick.ac.uk
- pubmlst.org

Ограничения метода

- Зависимость от базы данных
- Недостаточное покрытие пангенома отдельных видов микроорганизмов (в т.ч. возбудителей особо опасных инфекций)
- Низкая дискриминирующая способность при достижении *terra incognita*
- Низкая скорость обработки больших объёмов данных

Цель работы:

разработка и апробация схемы мультилокусного
сиквенс-типирования на основе корового генома
холерного вибриона для углублённой идентификации и
оценки родства различных групп штаммов

Исследованные штаммы:

V. cholerae non O1/O139

n=10

V. cholerae P-вар.

n=15

ctx+ tcp+

ctx- tcp+

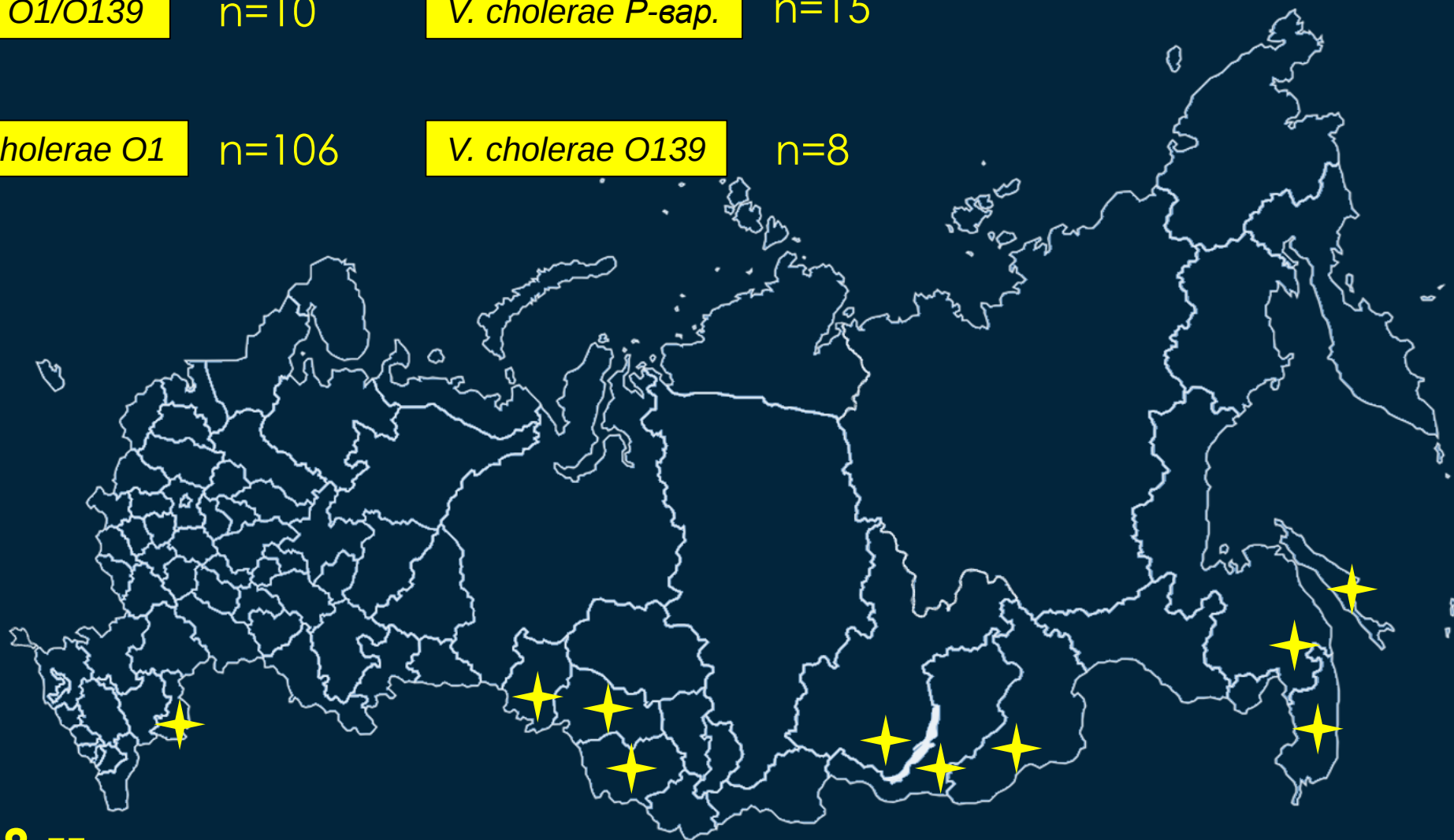
V. cholerae O1

n=106

V. cholerae O139

n=8

ctx- tcp-



1970 – 2018 гг.

В работе проведено сравнение двух подходов к cgMLST: полноценный программный конвейер chewBBACA и программа для определения аллельных вариантов MentaLiST в комбинации с авторскими R- и bash-скриптами.

Сопоставление результатов cgMLST с расщеплённой филогенетической сетью NeighbourNet (Genome Profiler - GeP) и эпидемиологическими данными говорит о преимуществе подхода с применением MentaLiST, поскольку в этом случае в большей степени прослеживается временная и территориальная приуроченность группирования штаммов на сиквенс-типы.

Апробация схемы sgMLST для *Vibrio cholerae* (106 ST)

Гетерогенные кластеры

V. cholerae O1
(*ctxAB*⁻*tcpA*⁻)

V. cholerae O1
(*ctxAB*⁺*tcpA*⁺/*ctxAB*⁻*tcpA*⁺)

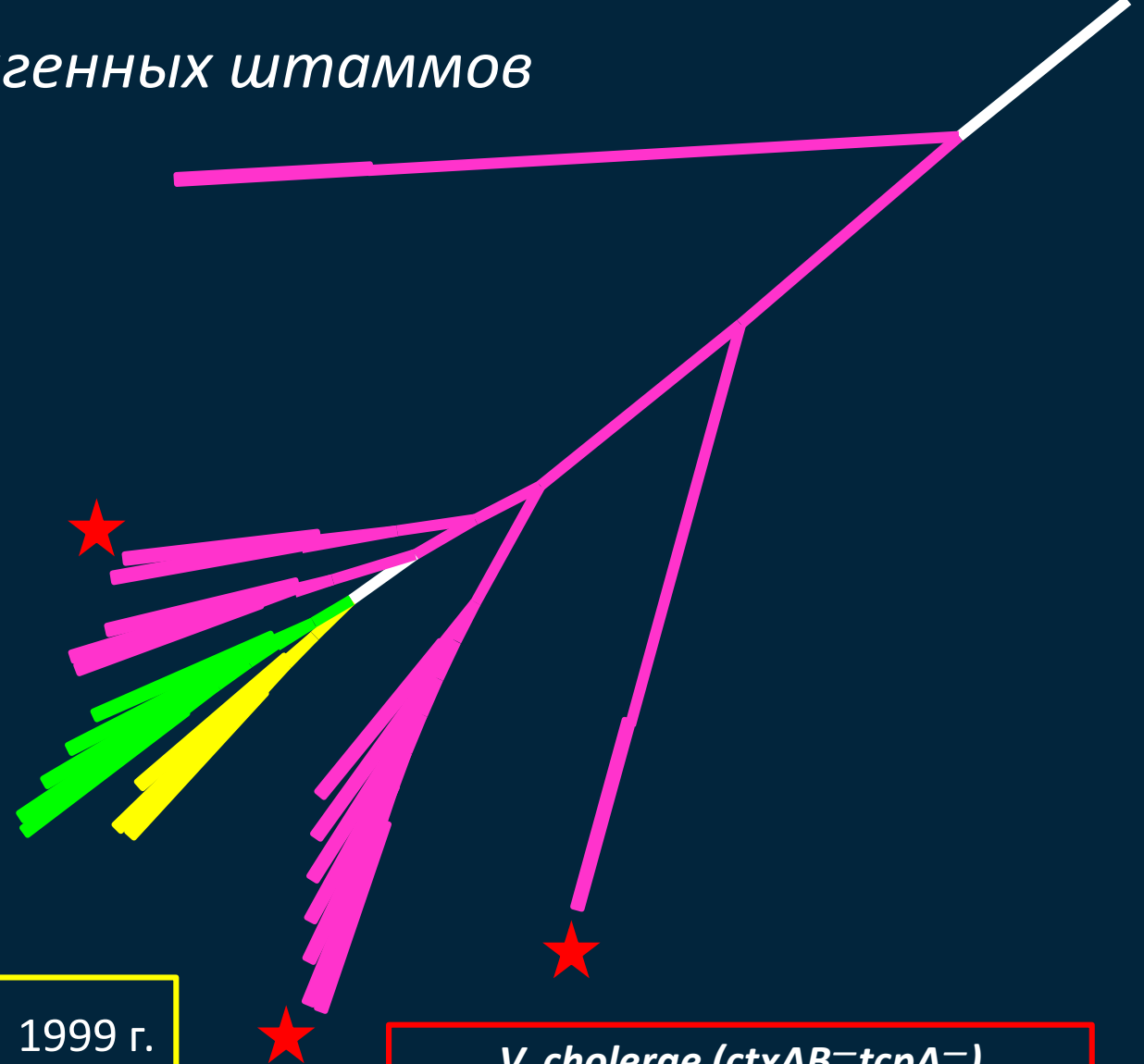
*P-варианты
V. cholerae*

Филогенетическое родство токсигенных штаммов *Vibrio cholerae* O1 серогруппы

Алтайский край, 1973, 1994 гг.
Астраханская область, 1970 г.
Омская область, 1972, 1973, 1994 гг.
Новосибирск, 1973, 1998 гг.

Владивосток, 1999 г.
(5 СТ, 12 штаммов)

Южно-Сахалинск, 1999 г.
(3 СТ, 11 штаммов)



***V. cholerae* (*ctxAB*⁻*tcpA*⁻)**
Мутантные клоны вспышечных штаммов, утратившие за время хранения на питательных средах протяжённые участки СТХ-профага

Нетоксигенные
V. cholerae
(*ctxAB*⁻*tcpA*⁻)

Нетоксигенные
V. cholerae O1
(*ctxAB*⁻*tcpA*⁺)

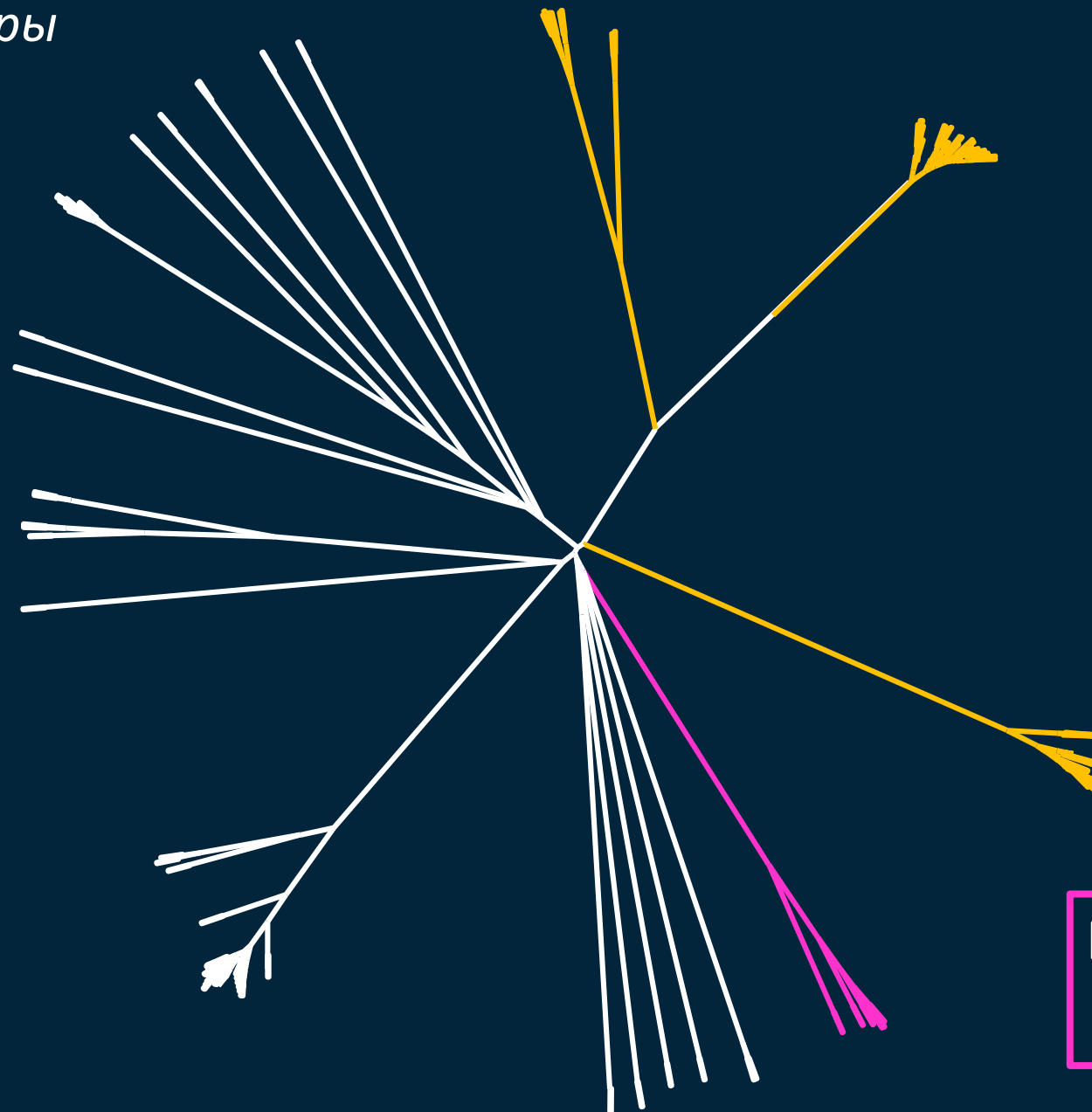
$\Delta \text{ genes} = 2296$

$\Delta \text{ genes} = 483$

Токсигенные
V. cholerae O1
(*ctxAB*⁺*tcpA*⁺)

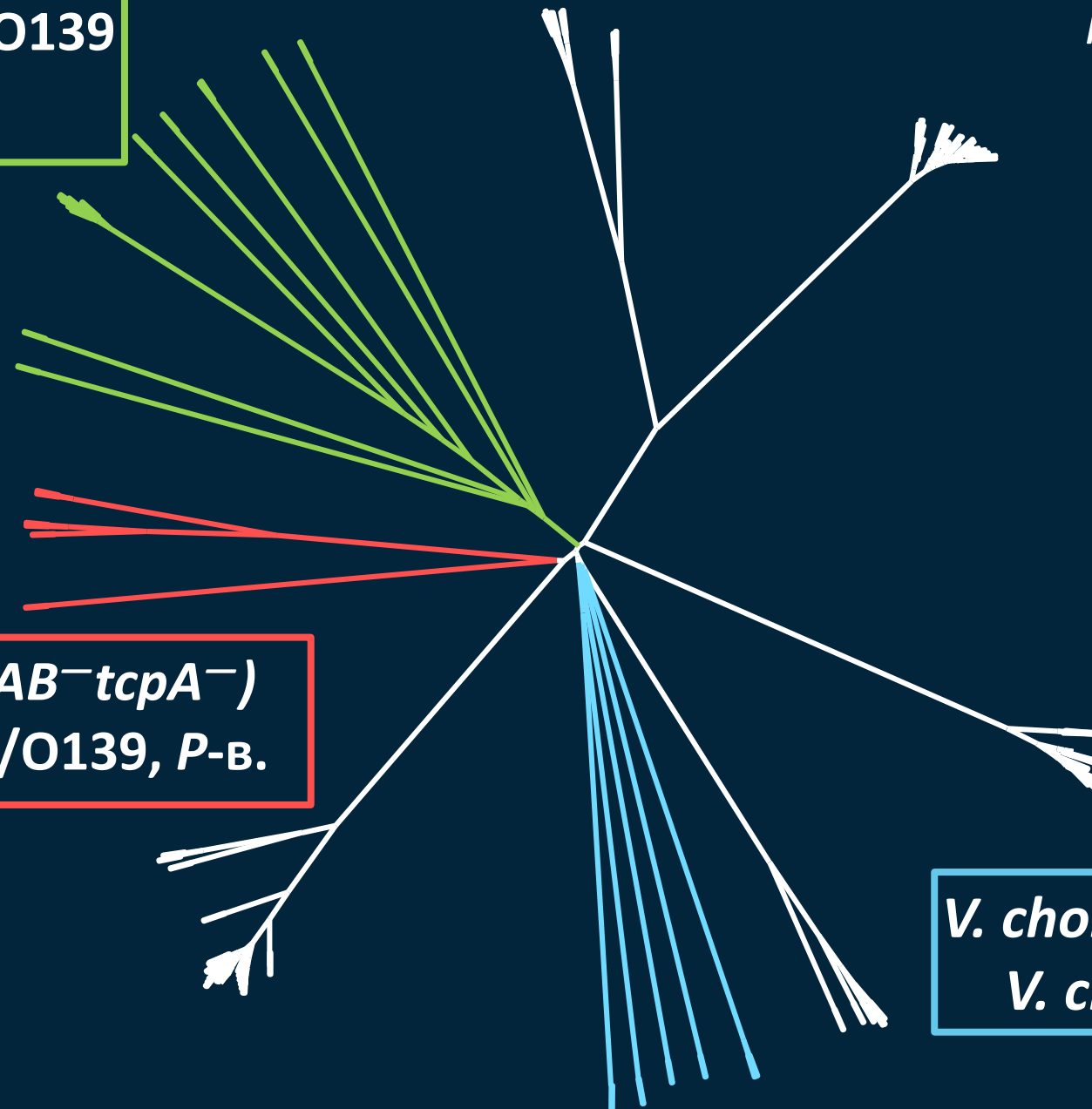
Филогенетическое положение
потенциально эпидемически опасных
штаммов *Vibrio cholerae* O1 серогруппы

cgMLST Vibrio cholerae.
Гомогенные кластеры



Нетоксигенные
V. cholerae O1
(*ctxAB*⁻*tcpA*⁻)

Р-варианты
V. cholerae



V. cholerae O139
V. cholerae non O1/O139
P-варианты

V. cholerae O1 (*ctxAB*⁻*tcpA*⁻)
V. cholerae non O1/O139, *P*-В.

V. cholerae O1 (*ctxAB*⁻*tcpA*⁻)
V. cholerae non O1/O139

Оптимизация схемы типирования

- В целом, данный метод как таковой пригоден для анализа филогенетических взаимосвязей, он характеризуется высокой дискриминирующей способностью как на уровне вида, так и с высокой точностью улавливает внутригрупповые различия. Результаты типирования этим методом обладают хорошей эпидемиологической согласованностью.
- Однако, он неэффективен с точки зрения затрачиваемых ресурсов и процессорного времени
(> 48 часов на 2x Intel Xeon E5-2620v4, 2,1 GHz)

=> необходима оптимизация схемы для уменьшения количества мишеней в анализе

=> Задача: углублённый поиск генов, ...

- ...эволюционная динамика которых совпадает с общими трендами, характерными для *Vibrio cholerae*
- ...отражающих внутригрупповые различия

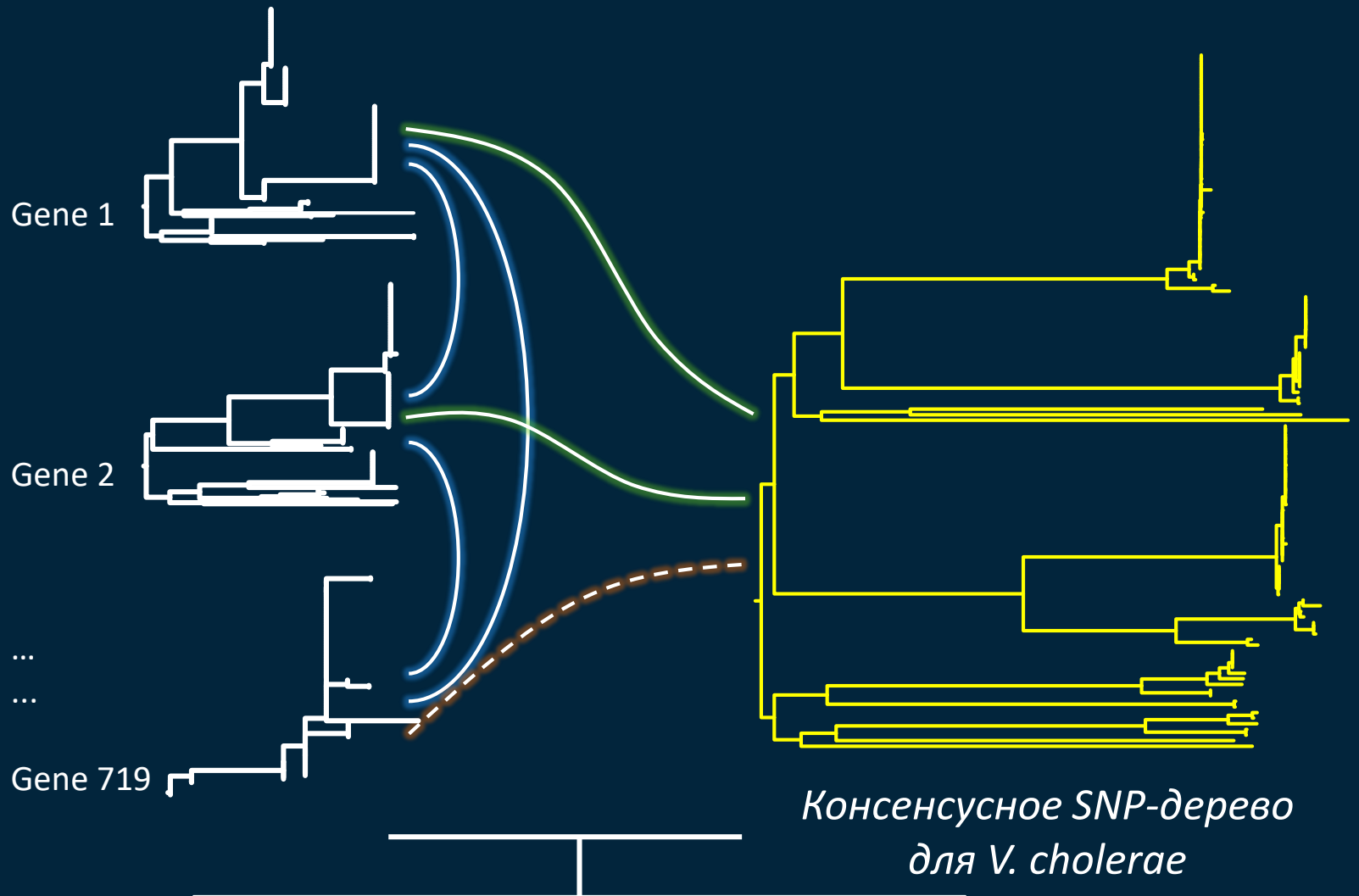
Эволюционные стратегии холерного вибриона различаются в ряду таксономических категорий, поэтому разведочный поиск производился на основе многофакторного анализа пула филогенетических деревьев для выборки штаммов по следующим генам.



Целевые гены:

protein export	14
phosphotransferase system	22
biofilm formation	53
bacterial chemotaxis	17
quorum sensing	16
flagellar assembly	34
bacterial secretion system	31
stress adaptation	532
[total]:	719

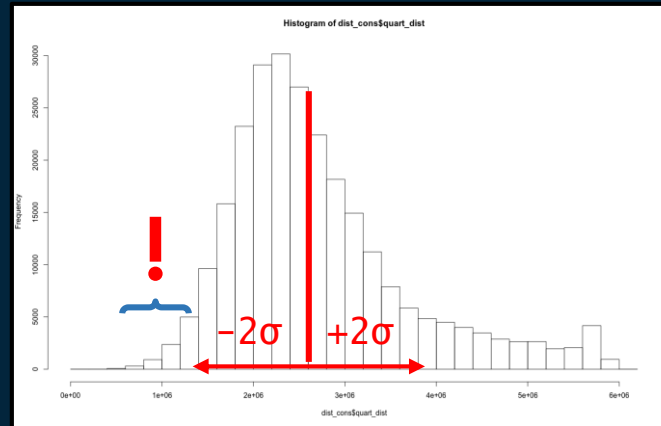
Построение ML-филогении
по каждому из генов
для выборки штаммов



Tree distance metrics:
> SPR (subtree pruning and
regraft) distance
> Path distance
> Quartet distance

Попарное сравнение топологий между
собой и с консенсусным snp-деревом
для *V. cholerae* с использованием
метрик дистанций

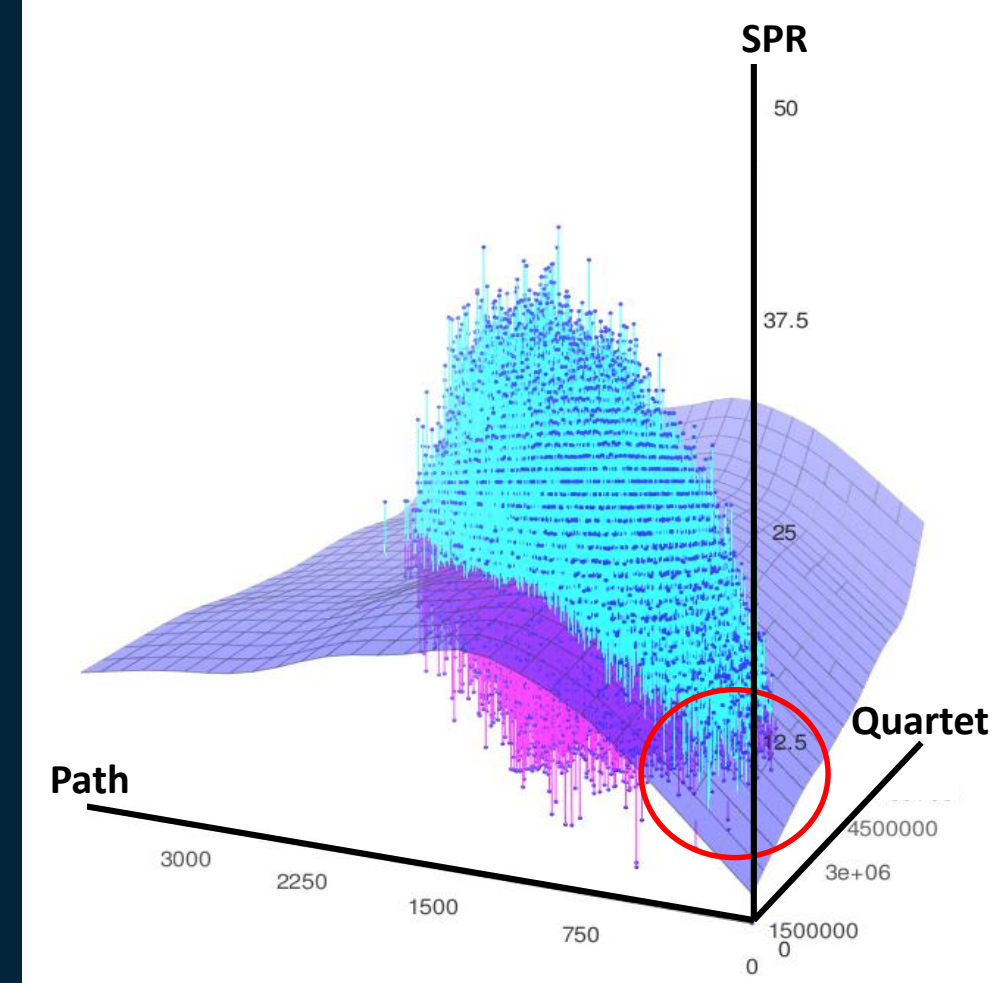
Оптимизация схемы типирования. Выбор целевых генов



Матрица попарных дистанций
для каждой исследуемой метрики

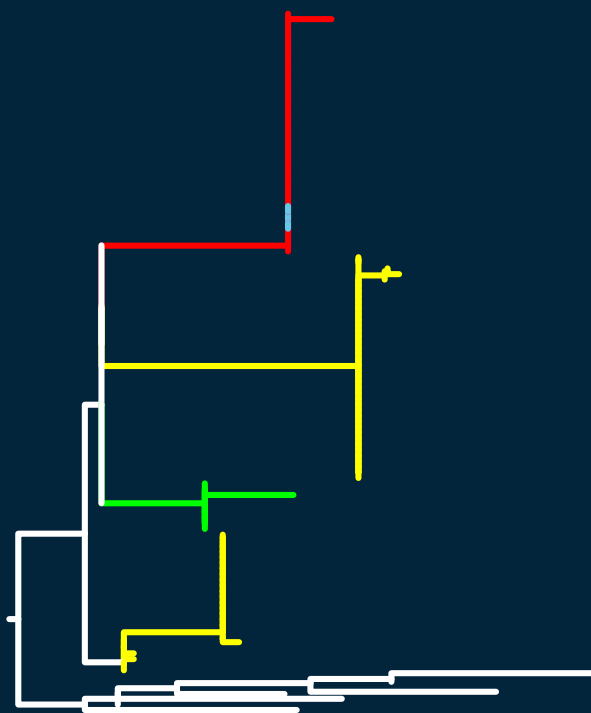
Распределение значений
схожести топологий

[! – целевая область с генами,
топология которых максимально
соответствует консенсусу]



Многомерное шкалирование
по 3 дистанциям для 719 целевых генов

Усреднение эволюционной динамики по 15 генам демонстрирует сохранение кластеризации штаммов в основные таксономические группы, но **не позволяет отследить внутригрупповые различия**.



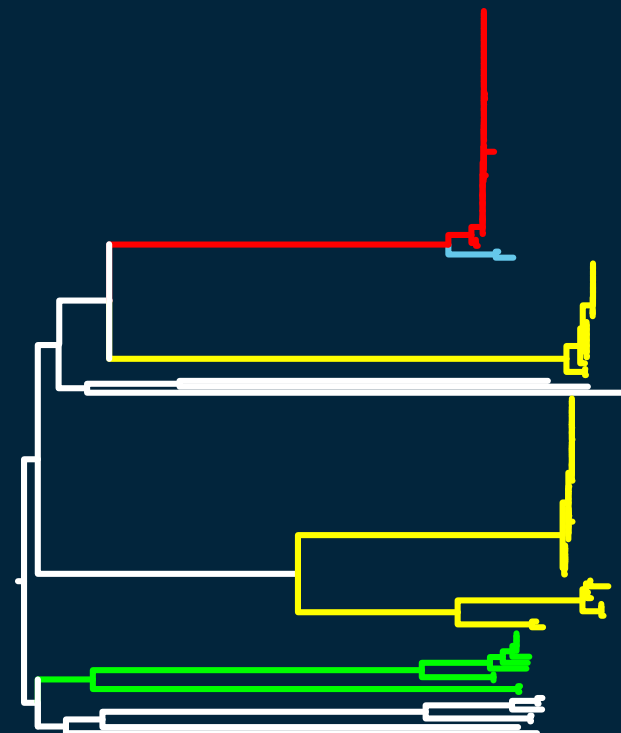
Усреднение целевых деревьев по 15 генам

Токсигенные
V. cholerae O1
(*ctxAB*⁺*tcpA*⁺)

Нетоксигенные
V. cholerae O1
(*ctxAB*⁻*tcpA*⁻)

Нетоксигенные
V. cholerae O1
(*ctxAB*⁻*tcpA*⁻)

P-варианты
V. cholerae



Консенсусное SNP-дерево для *V. cholerae*

Для разрешения данного ограничения необходим дальнейший углублённый поиск погенных различий внутри групп.

Нетоксигенные
V. cholerae
(*ctxAB*⁻*tcpA*⁻)

Нетоксигенные
V. cholerae O1
(*ctxAB*⁻*tcpA*⁺)

Δ genes = 2296

Δ genes = 483

Токсигенные
V. cholerae O1
(*ctxAB*⁺*tcpA*⁺)

Филогенетическое положение
потенциально эпидемически опасных
штаммов *Vibrio cholerae* O1 серогруппы

Внутригрупповые различия

$[ctxAB^-tcpA^+] \leftrightarrow [ctxAB^+tcpA^+]$

39/483

Hypothetical proteins	17
Housekeeping genes	9

39 из 483-х генов разделяют исследуемую подгруппу холерных вибрионов строго на две части: токсигенные и потенциально эпидемически опасные штаммы

Energy metabolism
Oxidative stress
Cold stress
DNA repair

13



Предмет для дальнейшего анализа детальной структуры, поскольку предположительно за счёт данных генов реализовывался адаптивный потенциал штаммов этой группы при освоении новой для них экологической ниши.

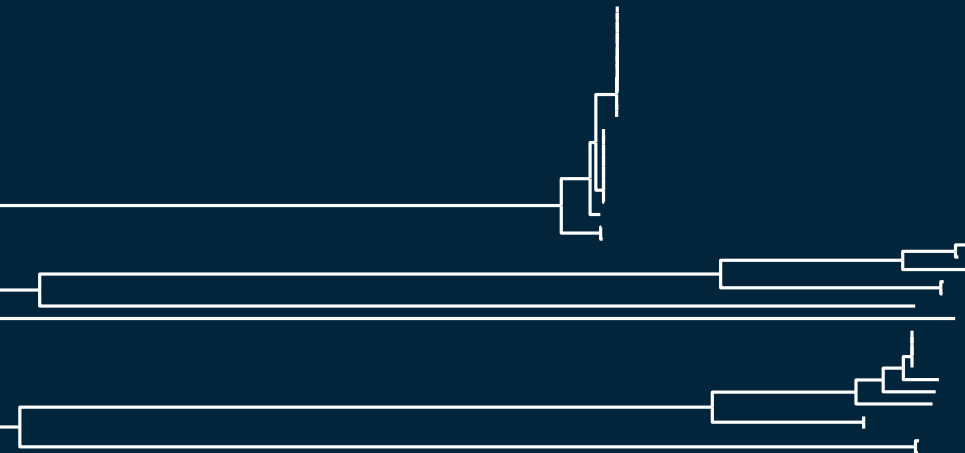
Заключение:

- Выявленные закономерности кластеризации обширной и разнородной выборки штаммов согласуются с эпидемиологическими характеристиками и подтверждаются данными дополнительных методов филогенетического анализа, что говорит об эффективности разработанного подхода для cgMLST *V. cholerae*.
- Идентифицированы группоспецифические локусы корового генома *V. cholerae*, что определяет перспективность дальнейших исследований по разработке и внедрению мультилокусных систем идентификации и типирования возбудителя холеры.
- Произведён подбор участков, уникальных для группы нетоксигенных штаммов *V. cholerae* O1 (*ctxAB*—*tcpA*⁺) и P-вариантов *V. cholerae*. В настоящее время проходит апробацию тест-система RT-PCR для внутривидового типирования холерного вибриона.



ФКУЗ Иркутский научно-исследовательский
противочумный институт Роспотребнадзора

Анализ корового генома
штаммов *Vibrio cholerae*
разной эпидемической значимости:
разработка подходов к sg-MLST и поиск
специфических молекулярных маркеров



Авторы:
Н.О. Бочалгин
А.В. Миронова, С.В. Балахонов