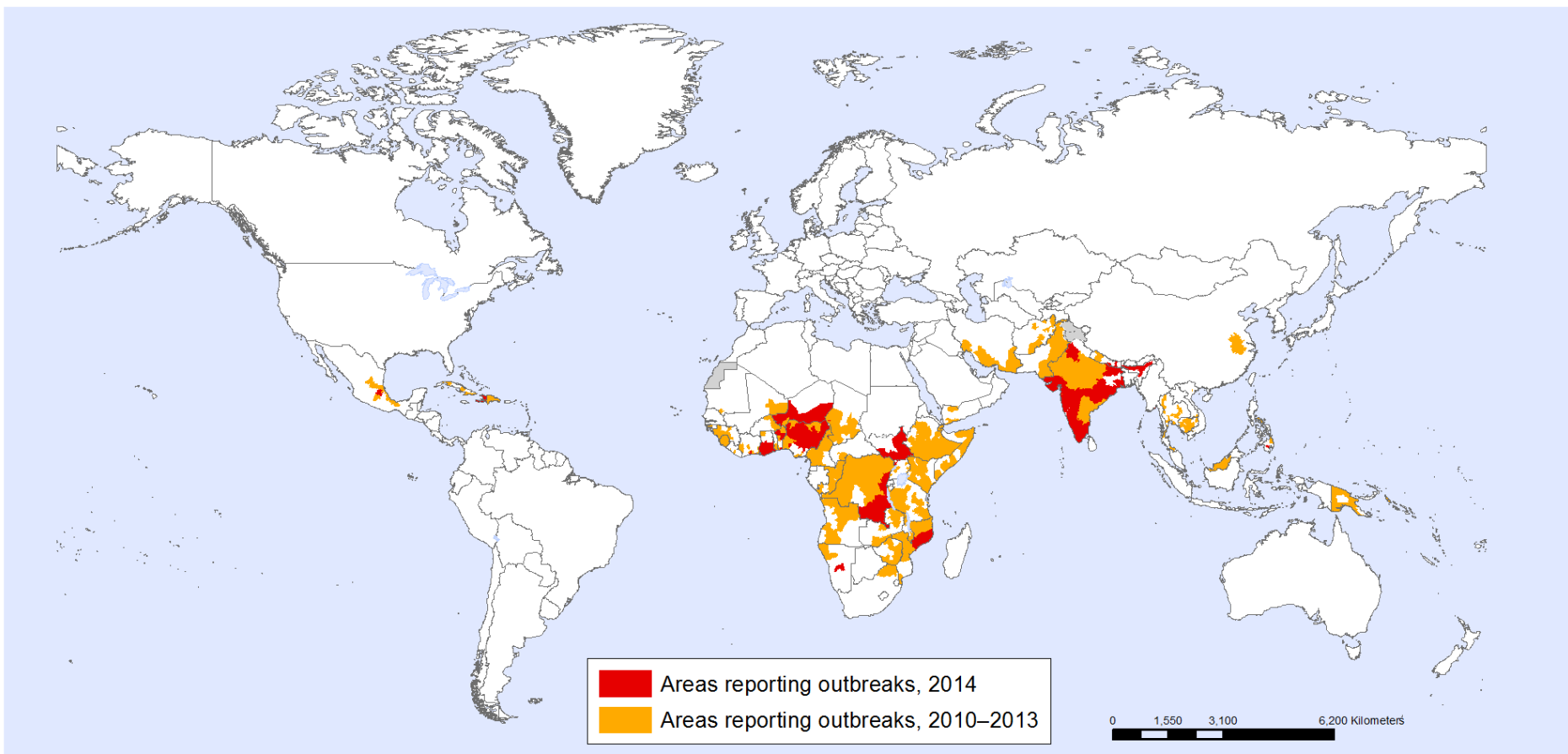


Поиск генов биопленкообразования и системы «quorum sensing» в геномах *Vibrio cholerae* El Tor, выделенных на территории Дальнего Востока России

Носкова Екатерина Эдуардовна
Научный руководитель: Гладких Анна Сергеевна

Территории, на которых в 2010-2014 гг. зарегистрированы вспышки холеры



The boundaries and names shown and the designations used on this map do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement.

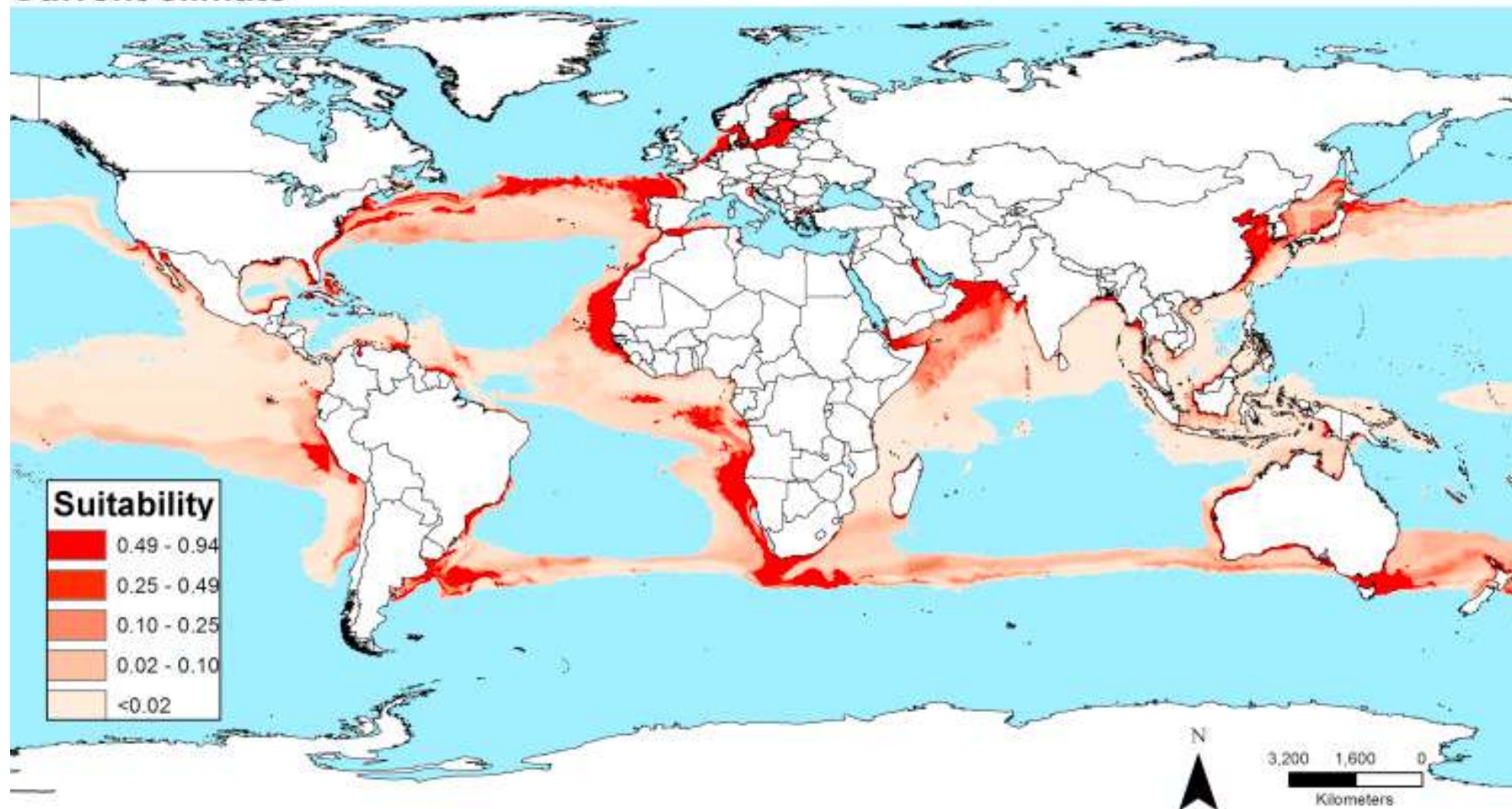
Data Source: World Health Organization
Map Production: Health Statistics and
Information Systems (HSI)
World Health Organization



© WHO 2015. All rights reserved.

Карта выживания *V. cholerae* в прибрежных водах в условиях современного климата

Current climate



Биопленка и система «quorum sensing»

- Бактериальные биопленки – это агрегаты микроорганизмов, объединенных между собой при помощи вырабатываемого ими защитного адгезивного матрикса.
- Биопленки повышают выживаемость микроорганизмов в водной среде.
- Биопленка защищает популяцию от воздействия неблагоприятных факторов окружающей среды.
- «Quorum sensing» - система межклеточного взаимодействия, позволяющая микроорганизму делиться информацией и регулировать экспрессию генов.

Цель и задачи исследования

Цель работы - поиск генов, ответственных за образование биопленок и участвующих в системе «quorum sensing», в геномах *V. cholerae* El Tor, выделенных на территории Дальнего Востока России.

Задачи:

- Изучить литературу и найти гены, ответственные за формирование биопленок и системы «quorum sensing» в геномах *V. cholerae* El Tor.
- Провести сборку геномов штаммов *V. cholerae* El Tor, выделенных на территории Дальнего Востока России.
- Проанализировать структуру и локализацию генов в сборках по сравнению с геномом референсного штамма *V. cholerae* El Tor N16961.
- Провести поиск полиморфизмов (SNPs) в анализируемых геномах.
- Сделать заключение о характерных особенностях генов биопленкообразования и системы «quorum sensing» в геномах *V. cholerae* El Tor, выделенных на территории Дальнего Востока.

Материал для исследования

Данные 454-секвенирования геномов штаммов из коллекции Иркутского НИПЧИ:

Vibrio cholerae El Tor str I-1300 – выделен от человека во время вспышки холеры в г. Южно-Сахалинске, 1999 г.

Vibrio cholerae El Tor str 102-16 – выделен из оз. Соленое в г. Находка, 2016 г.

Референсный геном: *Vibrio cholerae* El Tor str. N16961

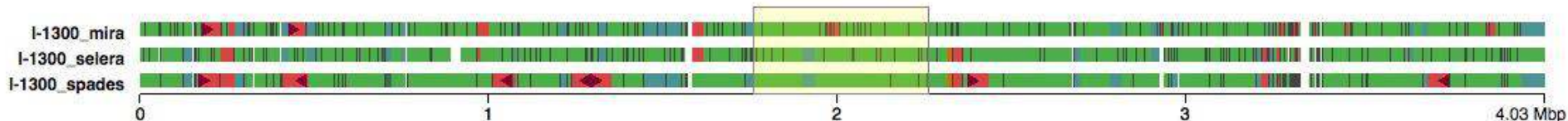
Для анализа было выбрано 35 генов, ответственных за образование биопленки и систему «quorum sensing» у *Vibrio cholerae* El Tor на основании данных литературы.

Пайплайн и используемое ПО

- Фильтрация, корректировка ридов 454-секвенирования (Prinseq, Coral)
- *De novo* сборка геномов (Mira, Celera assembler, SPAdes)
- Выравнивание контигов на референс и получение информации о покрытии генов (QUAST)
- Поиск однолокусных полиморфизмов - SNPs (nucmer + собственный скрипт на python)
- Проверка SNPs на синонимичность (собственный скрипт на python)
- Подтверждение результатов:
 - Референсная сборка генома (bowtie2)
 - Оценка покрытия генов (собственный скрипт на python)
 - Поиск SNPs на основании референсной сборки (nucmer)

Результаты геномных сборок (str. I-1300)

Genome statistics	I-1300_mira	I-1300_selera	I-1300_spades
Genome fraction (%)	97.868	95.386	96.459
Duplication ratio	1.057	1.002	1.003
# genes	32 + 3 part	30 + 4 part	34 + 1 part
Largest alignment	147 663	51 774	135 407
Total aligned length	4 170 901	3 854 374	3 895 124
NGA50	32 305	13 932	56 744
LGA50	40	85	26
Misassemblies			
# misassemblies	19	6	7
Misassembled contigs length	284 200	89 527	179 845
Mismatches			
# mismatches per 100 kbp	14.08	4.94	36.68
# indels per 100 kbp	102.52	10.45	96.1
# N's per 100 kbp	8.41	0.05	0
Statistics without reference			
# contigs	504	403	176
Largest contig	147 663	51 782	135 462
Total length	4 282 018	3 951 137	4 000 586
Total length (>= 1000 bp)	4 116 537	3 951 137	3 984 809
Total length (>= 10000 bp)	3 612 633	2 721 168	3 732 951
Total length (>= 50000 bp)	1 131 038	103 021	2 261 362



Результаты геномных сборок (str. 102-16)

Genome statistics	102-16_mira	102-16_selera	102-16_spades
Genome fraction (%)	89.955	89.255	89.236
Duplication ratio	1.033	1.006	1.003
# genes	26 + 3 part	26 + 3 part	26 + 3 part
Largest alignment	284 425	243 633	283 920
Total aligned length	3 741 799	3 614 681	3 604 595
NGA50	81 248	59 706	80 727
LGA50	16	19	14
Misassemblies			
# misassemblies	45	43	44
Misassembled contigs length	1 850 122	1 653 819	2 493 462
Mismatches			
# mismatches per 100 kbp	1506.11	1506.28	1511.13
# indels per 100 kbp	132.02	128.83	136.08
# N's per 100 kbp	7.04	0	0
Statistics without reference			
# contigs	235	104	106
Largest contig	367 386	244 010	284 299
Total length	4 178 733	4 017 521	3 983 093
Total length (>= 1000 bp)	4 066 004	4 017 521	3 960 151
Total length (>= 10000 bp)	3 957 747	3 875 937	3 871 702
Total length (>= 50000 bp)	3 364 159	3 043 473	3 618 045



Предварительные выводы

- Гены штамма I-1300 близки по структуре к генам референсного штамма. Они полностью покрыты ридами, в референсной сборке полиморфизмы отсутствуют. В сборке *de novo* обнаружались несинонимичные замены в двух локусах: 2 замены в - VC1021 и 1 в - VCA0865.
- В штамме 102-16 5 генов кластера *vpsI* отсутствуют, 2 гена *vpsI* кластера и ген *mshA* представлены лишь частично. В других генах найдены и определены SNPs. Полученные данные свидетельствуют об особенностях процесса биопленкообразования у выделенного штамма и требуют дальнейшего экспериментального подтверждения.

Спасибо за внимание!