(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
9 September 2016 (09.09.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/141370 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2016/021089

(22) International Filing Date:
5 March 2016 (05.03.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/128,764 5 March 2015 (05.03.2015) US

(71) Applicant: **UNIVERSITY OF NEW HAMPSHIRE** [US/US]; 105 Main Street, Durham, NH 03824 (US).

(72) Inventors: **WHISTLER, Cheryl**; 1 Pinecrest Lane, Durham, NH 03824 (US). **HALL, Jeffrey, Allister**; 4 Alderwood Drive, Stratham, NH 03885 (US). **XU, Feng**; 149 Flag Marsh Road, Guilford, CT 06437 (US).

(74) Agent: **ANDERSON, Marydilys, S.**; Finch & Maloney PLLC, 50 Commercial Street, Manchester, NH 03101 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

— *with sequence listing part of description (Rule 5.2(a))*

(54) Title: METHODS AND COMPOSITIONS FOR IDENTIFYING PATHOGENIC VIBRIO PARAHAEMOLYTICUS

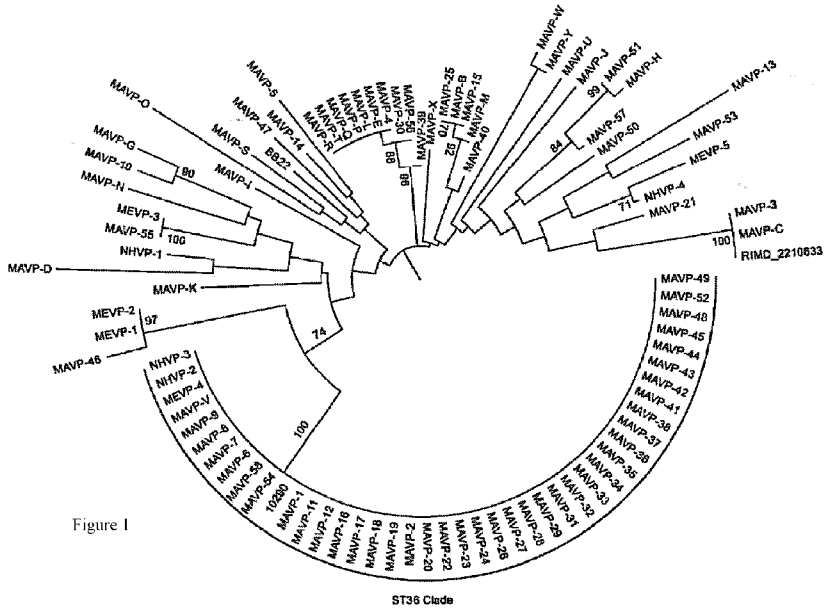


Figure 1

(57) Abstract: The invention relates, in part, to methods to identify the presence and/or level of a pathogenic *V. parahaemolyticus* bacterium in a subject or in or on a substrate. Methods are provided that, in part, permit detection of infection of a subject, or contamination of a substrate by pathogenic *V. parahaemolyticus* bacteria.

WO 2016/141370 A2

METHODS AND COMPOSITIONS FOR IDENTIFYING PATHOGENIC *VIBRIO PARAHAEMOLYTICUS*

Related Applications

This application claims benefit under 35 U.S.C. § 119(e) of U.S. Provisional Application No. 62/128,764, filed March 5, 2015 the content of which is incorporated by reference herein in its entirety.

Field of the Invention

The invention relates, in part, to methods and compositions for identifying pathogenic Vibrio parahaemolyticus contamination and infections.

Background

Rare pathogenic variants of *Vibrio parahaemolyticus*, a ubiquitous yet typically harmless estuarine bacterium, can cause human gastric infections most often from the consumption of raw or improperly handled seafood, and wound infections from recreational aquatic activities [Daniels NA, et al. (2000) J Infect Dis 181: 1661-1666; Scallan E, et al. (2011) Emerg Infect Dis 2011 Jan. DOI:10.3201/eid1701.P11101]. A better understanding of conditions that promote emergence and relative abundance of pathogens is necessary to develop appropriate strategies for disease prevention but an obstacle for the study of emergent pathogenic strains of *V. parahaemolyticus* is a lack of understanding of factors that define virulence and that could be used to detect pathogens within mostly non-pathogenic populations

Summary of the Invention

According to one aspect of the invention, methods for identifying the status of a pathogenic *V. parahaemolyticus* in a sample are provided. The methods include: (a) detecting in a sample, a level of at least one ST36*prp* polynucleotide comprising a nucleic acid sequence set forth as SEQ ID NO:1, 2, 3, 4, 5, 6, 7, 21, 22, 23, 24, 25 or 26, or a variant thereof; (b) comparing the level of the detected ST36*prp* polynucleotide to a control level of the detected ST36*prp* polynucleotide; and (c) identifying the status of the pathogenic *V. parahaemolyticus* in the sample, based at least in part on a difference between the level of the detected ST36*prp* polynucleotide and the control level of the detected ST36*prp* polynucleotide. In some embodiments, a higher level of the detected ST36*prp* polynucleotide

compared to the control level of the detected ST36*prp* polynucleotide identifies the status of the pathogenic *V. parahaemolyticus* as present in the sample. In certain embodiments, the control level of the detected ST36*prp* polynucleotide is zero. In some embodiments, the method also includes: (d) detecting in the sample a level of at least one of a ST36*cps* polynucleotide comprising a nucleic acid sequence set forth as SEQ ID NO:8, 9, 10, 27 or 28, or a variant thereof; (e) comparing the level of the detected ST36*cps* polynucleotide to a control level of the detected ST36*cps* polynucleotide; and (f) identifying the status of the pathogenic *V. parahaemolyticus* in the sample, based at least in part on the difference between the level of the detected ST36*cps* polynucleotide and the control level of the detected ST36*cps* polynucleotide. In some embodiments, a higher level of the detected ST36*cps* polynucleotide compared to the control level of the detected ST36*cps* polynucleotide identifies the status of the pathogenic *V. parahaemolyticus* as present in the sample. In some embodiments, the control level of the detected ST36*cps* polynucleotide is zero. In certain embodiments, the method also includes (g) detecting in the sample a level of at least one of a *tlh* polynucleotide comprising a nucleic acid sequence set forth as SEQ ID NO:11, 12, 13, 14, 29, 30 or 31, or a variant thereof; and (h) comparing the level of the detected *tlh* polynucleotide to a control level of the detected *tlh* polynucleotide. In some embodiments, a higher level of the detected *tlh* polynucleotide compared to the control level of the detected *tlh* polynucleotide aids in identifying the status of the pathogenic *V. parahaemolyticus* in the sample. In certain embodiments, the control level of the detected *tlh* polynucleotide is zero. In some embodiments, the method also includes (i) detecting in the sample a level of at least one of a *tdh* and *trh* polynucleotide comprising a nucleic acid sequence set forth as SEQ ID NO:15, 16, 17, 18, 19, 20, 32, 33, 34 or 35, or a variant thereof; and (j) comparing the level of the detected polynucleotide to a control level of the detected polynucleotide. In some embodiments, a higher level of the detected *tdh* or *trh* polynucleotide compared to the control level of the detected polynucleotide aids in identifying the status of the pathogenic *V. parahaemolyticus* in the sample. In some embodiments, the control level of the detected *tdh* or *trh* polynucleotide is zero. In certain embodiments, the method also includes: (k) detecting in the sample a level of at least one of a ST36*flp* polynucleotide comprising a nucleic acid sequence set forth as SEQ ID NO:36, 37, 38, 39 or 40, or a variant thereof; (l) comparing the level of the detected ST36*flp* polynucleotide to a control level of the detected ST36*flp* polynucleotide; and (m) identifying the status of the pathogenic *V. parahaemolyticus* in the sample, based at least in part on difference between the level of the detected ST36*flp* polynucleotide and the control level of

the detected ST36*flp* polynucleotide. In some embodiments, a higher level of the detected ST36*flp* polynucleotide compared to the control level of the detected ST36*flp* polynucleotide aids in identifying the status of the pathogenic *V. parahaemolyticus* in the sample. In some embodiments, the control level of the detected ST36*flp* polynucleotide is zero. In certain embodiments, the variant of the SEQ ID NO:1, 2, 3, 4, 5, 6, 7, 21, 22, 23, 24, 25 or 26 nucleic acid sequence has at least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the nucleic acid sequence of SEQ ID NO:1, 2, 3, 4, 5, 6, 7, 21, 22, 23, 24, 25 or 26, respectively. In some embodiments, the variant of the SEQ ID NO:8, 9, 10, 27 or 28 nucleic acid sequence has at least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the nucleic acid sequence of SEQ ID NO:8, 9, 10, 27 or 28, respectively. In some embodiments, the variant of the SEQ ID NO:11, 12, 13, 14, 29, 30 or 31 nucleic acid sequence has at least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the nucleic acid sequence of SEQ ID NO:11, 12, 13, 14, 29, 30 or 31, respectively. In some embodiments, the variant of the SEQ ID NO:15, 16, 17, 18, 19, 20, 32, 33, 34 or 35 nucleic acid sequence has at least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the nucleic acid sequence of SEQ ID NO:15, 16, 17, 18, 19, 20, 32, 33, 34 or 35, respectively. In certain embodiments, the variant of the SEQ ID NO:36, 37, 38, 39 or 40 nucleic acid sequence has at least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the nucleic acid sequence of SEQ ID NO:36, 37, 38, 39 or 40, respectively. In some embodiments, the variant of the nucleic acid sequence comprises a fragment of the nucleic acid sequence. In some embodiments, the variant of the nucleic acid sequence comprises a fragment of the nucleic acid sequence and the fragment has at least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the region of the nucleic acid sequence with which it aligns. In certain embodiments, the sample is obtained from a subject. In some embodiments, the sample is obtained from a substrate.

According to another aspect of the invention methods of identifying the status of a pathogenic *V. parahaemolyticus* in a sample are provided. The methods include: (a) detecting in a sample a level of at least one ST36*cps* polynucleotide comprising a nucleic acid sequence set forth as SEQ ID NO:8, 9, 10, 27 or 28, or a variant thereof; (b) comparing the level of the detected ST36*cps* polynucleotide to a control level of the detected ST36*cps* polynucleotide; and (c) identifying the status of a pathogenic *V. parahaemolyticus* in the sample, based at least in part on a difference in the level of the detected ST36*cps* polynucleotide and the control level of the detected ST36*cps* polynucleotide. In some embodiments, a higher level of the detected ST36*cps* polynucleotide compared to the control level of the detected ST36*cps* polynucleotide identifies the status of the pathogenic *V.*

parahaemolyticus as present in the sample. In certain embodiments, the control level of the detected ST36*cps* polynucleotide is zero. In some embodiments, the method also includes detecting in the sample a level of at least one of a polynucleotide comprising a ST36*prp* nucleic acid sequence or variant thereof, a *tlh* nucleic acid sequence or variant thereof, a *tdh* nucleic acid sequence or variant thereof, a *trh* nucleic acid sequence or variant thereof, and an ST36*flp* nucleic acid sequence or variant thereof. In some embodiments, the variant of the SEQ ID NO: 8, 9, 10, 27, or 28 nucleic acid sequence has at least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the nucleic acid sequence of SEQ ID NO:8, 9, 10, 27 or 28, respectively. In certain embodiments, the variant of the nucleic acid sequence comprises a fragment of the nucleic acid sequence. In some embodiments, the variant of the nucleic acid sequence comprises a fragment of the nucleic acid sequence and the fragment has at least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the region of the nucleic acid sequence with which it aligns. In some embodiments, the sample is obtained from a subject. In some embodiments, the sample is obtained from a substrate.

According to yet another aspect of the invention, methods of identifying the presence of a pathogenic *V. parahaemolyticus* in a sample are provided. The methods include (a) detecting in a sample a level of at least one ST36*flp* polynucleotide comprising a nucleic acid sequence set forth as SEQ ID NO: 36, 37, 38, 39 or 40, or a variant thereof; (b) comparing the level of the detected ST36*flp* polynucleotide to a control level of the detected ST36*flp* polynucleotide; and (c) identifying the status of a pathogenic *V. parahaemolyticus* in the sample, based at least in part on the difference between the level of the detected ST36*flp* polynucleotide and the control level of the detected ST36*flp* polynucleotide. In certain embodiments, a higher level of the detected ST36*flp* polynucleotide compared to the control level of the detected ST36*flp* polynucleotide identifies the status of the pathogenic *V. parahaemolyticus* as present in the sample. In some embodiments, the control level of the detected ST36*flp* polynucleotide is zero. In some embodiments, the method also includes detecting in the sample a level of at least one of a polynucleotide comprising a ST36*prp* nucleic acid sequence or variant thereof, a *tlh* nucleic acid sequence or variant thereof, a *tdh* nucleic acid sequence or variant thereof, a *trh* nucleic acid sequence or variant thereof, and an ST36*cps* nucleic acid sequence or variant thereof. In certain embodiments, the variant of the SEQ ID NO:36, 37, 38, 39 or 40 nucleic acid sequence has at least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the nucleic acid sequence of SEQ ID NO:36, 37, 38, 39 or 40, respectively. In some embodiments, the variant of the nucleic acid sequence comprises a fragment of the nucleic acid sequence. In some embodiments, the variant of the

nucleic acid sequence comprises a fragment of the nucleic acid sequence and the fragment has at least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the region of the nucleic acid sequence with which it aligns. In some embodiments, the sample is obtained from a subject. In certain embodiments, the sample is obtained from a substrate.

According to another aspect of the invention, methods of assaying a sample are provided. The methods include (a) detecting in a sample determined to have at least one tdh and trh polynucleotide comprising a nucleic acid sequence set forth as SEQ ID NO: 15, 16, 17, 18, 19, 20, 32, 33, 34, 35, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120 or 121, or a variant thereof, and a level of at least one: (i) ST36*prp* polynucleotide comprising at least one of the nucleic acid sequences set forth as SEQ ID NO:1, 2, 3, 4, 5, 6, 7, 21, 22, 23, 24, 25, 26, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138 or 139, or a variant thereof; (ii) ST36*flp* polynucleotide comprising at least one of the nucleic acid sequences set forth as SEQ ID NO: 36, 37, 38, 39, 40, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149 or 150, or a variant thereof; (iii) ST631*end* polynucleotide comprising at least one of the nucleic acid sequences set forth as: SEQ ID NO: 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162 or 163, or a variant thereof; and (iv) 631-ENV polynucleotide comprising at least one of the nucleic acid sequences set forth as: SEQ ID NO: 59, 60, 61, 62, 170, 171, 172, 173, 174 or 175, or a variant thereof; and (b) comparing the level of the detected polynucleotide in at least one of steps (i), (ii), (iii), and (iv) to a control level of the detected polynucleotide in steps (i), (ii), (iii), and (iv), respectively. In some embodiments, the method also includes (c) determining the presence of a difference between the detected polynucleotide level and the control level of the detected polynucleotide as compared in (b); and (d) identifying the presence or absence of the pathogenic *V. parahaemolyticus* in the sample, based at least in part on difference between the level of the detected polynucleotide of at least one of steps (i), (ii), (iii), and (iii) and the respective control level of the detected polynucleotide. In some embodiments, assaying the sample comprises identifying the presence of a pathogenic *V. parahaemolyticus* in a sample. In some embodiments, the method also includes detecting an additional amplification control polynucleotide in the sample. In certain embodiments, the nucleic acid sequence of the variant of the tdh, trh, ST36*prp*, ST36*fl*, ST631*end* and ST631-ENV polynucleotide has at least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the nucleic acid sequence of the polynucleotide of which it is a variant thereof. In some embodiments, the variant of the nucleic acid sequence comprises a fragment of the nucleic acid sequence. In some embodiments, the variant of the nucleic acid sequence comprises a fragment of the nucleic

acid sequence and the fragment has at least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the region of the nucleic acid sequence with which it aligns. In some embodiments, the sample is obtained from at subject. In certain embodiments, the sample is obtained from a substrate.

According to another aspect of the invention, methods of identifying the status of a pathogenic *V. parahaemolyticus* in a sample are provided. The methods including: (a) detecting in a sample a level of at least one of (a) a ST36Phage polynucleotide comprising a nucleic acid sequence set forth as SEQ ID NO 81, 82, 83 or 84, or a or a variant thereof; (b) a ST36PhHypD-Orf9 polynucleotide comprising a nucleic acid sequence set forth as SEQ ID NO:94, 95, 96 or 97, or a variant thereof; (c) a ST36NEOrf10-Hyp polynucleotide comprising a nucleic acid sequence set forth as SEQ ID NO: 87, 88, 89 or 90, or a variant thereof; and (d) a tlh polynucleotide comprising a nucleic acid sequence set forth as SEQ ID NO: 12 or 14, or a variant thereof; (b) comparing the level of the detected ST36Phage polynucleotide, ST36PhHypD-Orf9 polynucleotide, ST36NEOrf10-Hyp polynucleotide, and tlh polynucleotide to a control level of the detected ST36Phage polynucleotide, ST36PhHypD-Orf9 polynucleotide, ST36NEOrf10-Hyp polynucleotide, and tlh polynucleotide, respectively; and (c) identifying the status of a pathogenic *V. parahaemolyticus* in the sample, based at least in part on the difference between the level of the detected ST36Phage polynucleotide, ST36PhHypD-Orf9 polynucleotide, ST36NEOrf10-Hyp polynucleotide, and tlh polynucleotide, and the control level of ST36Phage polynucleotide, ST36PhHypD-Orf9 polynucleotide, ST36NEOrf10-Hyp polynucleotide, and tlh polynucleotide, respectively. In some embodiments, a higher level of the detected ST36Phage polynucleotide compared to the control level of the detected ST36Phage polynucleotide and the presence of a higher level of one or more of the ST36PhHypD-Orf9 polynucleotide and the ST36NEOrf10-Hyp polynucleotide identifies the status of the pathogenic *V. parahaemolyticus* as present in the sample. In some embodiments, the control level of one or more of the detected ST36Phage polynucleotide, ST36PhHypD-Orf9 polynucleotide, and the ST36NEOrf10-Hyp polynucleotide is zero. In some embodiments, the method also includes detecting in the sample a level of at least one of a polynucleotide comprising a ST36*prp* nucleic acid sequence or variant thereof, a *tlh* nucleic acid sequence or variant thereof, a *tdh* nucleic acid sequence or variant thereof, a *trh* nucleic acid sequence or variant thereof, and an ST36*cps* nucleic acid sequence or variant thereof. In certain embodiments, the nucleic acid sequence of the variant of the SEQ ID NO: 81, 82, 83, 84, 94, 95, 96, 97, 87, 88, 89, 90, 12 or 14 nucleic acid sequence has at least 85%, 90%, 95%, 97%,

98%, 99%, or 100% sequence identity to the nucleic acid sequence of which it is a variant. In some embodiments, the variant of the nucleic acid sequence comprises a fragment of the nucleic acid sequence. In some embodiments, the variant of the nucleic acid sequence comprises a fragment of the nucleic acid sequence and the fragment has at least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the region of the nucleic acid sequence with which it aligns. In some embodiments, the sample is obtained from a subject. In certain embodiments, the sample is obtained from a substrate.

According to an aspect of the invention, methods of identifying the status of a pathogenic *V. parahaemolyticus* in a sample are provided, the methods including: (a) detecting in the sample, the level of at least one ST631*end* polynucleotide comprising at least one nucleic acid sequence set forth as SEQ ID NO: 53, 54, 55, 56, 57, 58, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162 or 163, or a variant thereof; (b) comparing the level of the detected ST631*end* polynucleotide to a control level of the detected ST631*end* polynucleotide; and (c) identifying the presence of the pathogenic *V. parahaemolyticus* in the sample, based at least in part on a difference between the level of the detected ST631*end* polynucleotide and the control level of the detected ST631*end* polynucleotide. In some embodiments, a higher level of the detected ST631*end* polynucleotide compared to the control level of the detected ST631*end* polynucleotide identifies the status of the pathogenic *V. parahaemolyticus* as present in the sample. In some embodiments, the control level of the detected ST631*end* polynucleotide is zero. In some embodiments, the method also includes detecting in the sample a level of at least one of a polynucleotide comprising an ST631-ENV nucleic acid sequence or variant thereof, an ST34*reg* nucleic acid sequence or variant thereof, a ST674*hyp* nucleic acid sequence or variant thereof, a ST1127*hyp* nucleic acid sequence or variant thereof, a ST36Phage nucleic acid sequence or variant thereof, an ST36NEOrf1O-Hyp nucleic acid sequence or variant thereof, a ST36PhHypD-Orf9 nucleic acid sequence or variant thereof, a TdhUreG nucleic acid sequence or variant thereof, a ST36*cps* nucleic acid sequence or variant thereof, a ST36*prp* nucleic acid sequence or variant thereof, a *tlh* nucleic acid sequence or variant thereof, a *tdh* nucleic acid sequence or variant thereof, a *trh* nucleic acid sequence or variant thereof, a ST36*flp* nucleic acid sequence or variant thereof, and an ORF8 nucleic acid sequence or variant thereof. In some embodiments, the nucleic acid sequence of the variant of the ST631*end,* ST631-ENV, ST34*reg*, ST674*hyp*, ST1127*hyp*, ST36Phage, ST36NEOrf1O-Hyp, ST36PhHypD-Orf9, TdhUreG, ST36*cps*, ST36*prp*, *tlh, tdh*, *trh*, ST36*flp*, ORF8 polynucleotide has at least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the nucleic acid sequence of the polynucleotide of which it is a

variant thereof. In certain embodiments, the variant of the nucleic acid sequence comprises a fragment of the nucleic acid sequence. In some embodiments, the variant of the nucleic acid sequence comprises a fragment of the nucleic acid sequence and the fragment has at least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the region of the nucleic acid sequence with which it aligns. In some embodiments, the sample is obtained from a subject. In certain embodiments, the sample is obtained from a substrate. In some embodiments of any of the aforementioned aspects of the invention, detecting comprises performing a polymerase chain reaction (PCR) to detect the level of the least one polynucleotide comprising an ST36*prp* nucleic acid sequence or variant thereof, an ST36*cps* nucleic acid sequence or variant thereof, a ST36*flp* nucleic acid sequence or variant thereof, *ST631end* nucleic acid sequence or variant thereof, *ST631*-ENV nucleic acid sequence or variant thereof, a ST34*reg* nucleic acid sequence or variant thereof, a ST674*hyp* nucleic acid sequence or variant thereof, a ST1127*hyp* nucleic acid sequence or variant thereof, a ST36Phage nucleic acid sequence or variant thereof, a ST36NEOrf1O-Hyp nucleic acid sequence or variant thereof, a ST36PhHypD-Orf9 nucleic acid sequence or variant thereof, and a TdhUreG nucleic acid sequence or variant thereof. In some embodiments, the PCR comprises qPCR. In certain embodiments, an oligonucleotide probe in the qPCR comprises a detectable label. In some embodiments, the oligonucleotide probe comprises a nucleic acid sequence set forth herein as: SEQ ID NO: 127, 128, 129, 130, 143, 144, 155, 156 or 170, or a variant thereof, wherein the variant has an 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the nucleic acid sequence of the polynucleotide of which it is a variant thereof. In some embodiments of any of the aforementioned aspects of the invention, detecting includes performing in situ hybridization to measure the level of the at least one polynucleotide comprising an ST36*prp* nucleic acid sequence or variant thereof, an ST36*cps* nucleic acid sequence or variant thereof, a ST36*flp* nucleic acid sequence or variant thereof, *ST631end* nucleic acid sequence or variant thereof, *ST631*-ENV nucleic acid sequence or variant thereof, a ST34*reg* nucleic acid sequence or variant thereof, a ST674*hyp* nucleic acid sequence or variant thereof, a ST1127*hyp* nucleic acid sequence or variant thereof, a ST36Phage nucleic acid sequence or variant thereof, a ST36NEOrf1O-Hyp nucleic acid sequence or variant thereof, a ST36PhHypD-Orf9 nucleic acid sequence or variant thereof, and a TdhUreG nucleic acid sequence or variant thereof. In some embodiments of any of the aforementioned aspects of the invention, detecting includes performing DNA hybridization to measure the level of the at least one polynucleotide comprising a ST36*prp* nucleic acid sequence or variant thereof, an ST36*cps* nucleic acid sequence or variant thereof, a ST36*flp*

nucleic acid sequence or variant thereof, *ST631end* nucleic acid sequence or variant thereof, *ST631*-ENV nucleic acid sequence or variant thereof, a ST34*reg* nucleic acid sequence or variant thereof, a ST674*hyp* nucleic acid sequence or variant thereof, a ST1127*hyp* nucleic acid sequence or variant thereof, a ST36Phage nucleic acid sequence or variant thereof, a ST36NEOrf1O-Hyp nucleic acid sequence or variant thereof, a ST36PhHypD-Orf9 nucleic acid sequence or variant thereof, and a TdhUreG nucleic acid sequence or variant thereof. In certain embodiments, the DNA hybridization is DNA microarray hybridization. In some embodiments of any of the aforementioned aspects of the invention, the method also includes one or more of selecting a therapeutic agent to reduce a pathogenic *V. parahaemolyticus* infection in the subject; and administering a therapeutic agent to treat a pathogenic *V. parahaemolyticus* infection in the subject, based at least in part on the identified status of the pathogenic *V. parahaemolyticus* in the sample. In some embodiments of any of the aforementioned aspects of the invention, the method also includes one or more of selecting an agent to reduce a pathogenic *V. parahaemolyticus* contamination of the substrate; and contacting the substrate with an agent selected to reduce a pathogenic *V. parahaemolyticus* contamination of the substrate, based at least in part on the identified status of the pathogenic *V. parahaemolyticus* in the sample obtained from the substrate. In certain embodiments of any of the aforementioned aspects of the invention, the presence of the pathogenic *V. parahaemolyticus* in the sample comprises the presence of one or more pathogenic *V. parahaemolyticus* bacteria in the sample. In some embodiments, the agent is an anti-*V. parahaemolyticus* agent. In some embodiments of any of the aforementioned aspects of the invention, the subject is at least one of: suspected of having a pathogenic *V. parahaemolyticus* infection and diagnosed with a pathogenic *V. parahaemolyticus* infection. In some embodiments of any of the aforementioned aspects of the invention, the subject is a vertebrate, and optionally is a mammal. In certain embodiments of any of the aforementioned aspects of the invention, the subject is an invertebrate. In some embodiments of any of the aforementioned aspects of the invention, the sample comprises one or more of a skin sample, fluid sample, tissue sample, stool sample, pus sample, gastric sample, emesis sample, inflammatory exudate sample, blood sample, or lymph sample. In some embodiments of any of the aforementioned aspects of the invention, the sample is a cultured sample. In certain embodiments of any of the aforementioned aspects of the invention, the substrate is at least one of: known to have been exposed to a *V. parahaemolyticus* bacteria, suspected of being contaminated with a *V. parahaemolyticus* bacteria; and identified as being contaminated with a *V. parahaemolyticus* bacteria. In some embodiments of any of the

aforementioned aspects of the invneiton, reducing the pathogenic *V. parahaemolyticus* contamination of the substrate comprises reducing the level of the pathogenic *V. parahaemolyticus* bacteria on the substrate. In some embodiments of any of the aforementioned aspects of the invention, the substrate comprises a liquid and the liquid optionally comprises water. In certain embodiments of any of the aforementioned aspects of the invention, the substrate comprises one or more of a metal, wood, plastic, glass, cork, fiber, a polymer, or a fabric. In some embodiments of any of the aforementioned aspects of the invention, the substrate comprises a food substance, and wherein the food substance optionally comprises shellfish. In some embodiments of any of the aforementioned aspects of the invention, the substrate comprises at least a portion of a tool, work surface, a medical device, body of water, clothing, skin, tissue, an edible substance, a beverage, or a food. In some embodiments of any of the aforementioned aspects of the invention, the substrate is a contaminated substrate and comprises one or more pathogenic *V. parahaemolyticus* bacteria. In certain embodiments of any of the aforementioned aspects of the invention, the sample comprises a fluid sample, semisolid sample, aqueous sample, or tissue sample.

According to another aspect of the invention, kits for performing a method of any embodiment of any aforementioned aspect of the invention are provided, the kits including: at least one a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to at least one of SEQ ID NO: 2-7, 9, 10, 36, 37, 52, 53, 59, 60, 63, 64, 75, 76, 81, 82, 87, 88, 94, 95, 122-126, 140-142, 151-154, 170, 171 and 172; wherein the kit additionally comprises instructions for detecting a status of a *V. parahaemolyticus* polynucleotide in a sample. In some embodiments, the *V. parahaemolyticus* polynucleotide is a pathogenic *V. parahaemolyticus* polynucleotide. In certain embodiments, the status comprises one or more of: the presence or absence of the polynucleotide; and a level of the polynucleotide. In some embodiments, the kit comprises one or more of: (a) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:2 and a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:3; (b) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:4 and a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:5; (c) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:6 and a container comprising an oligonucleotide

primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:7; (d) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:9 and a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:10; (e) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:36 and a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:37; (f) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:53 and a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:54; (g) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:59 and a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:60; (h) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:63 and a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:64; (i) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:69 and a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:70; (j) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:75 and a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:76; (k) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:122 and a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:123; (l) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:124 and a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:125; (m) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:126; (n) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or

100% sequence identity to SEQ ID NO:140 and one or more of a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:141, and a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:142; (o) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:151 and a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:152; (p) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:153 and a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:154; and (q) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:171 and a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:172. In some embodiments, the kit also includes one or more oligonucleotide probes. In some embodiments, one or more of the oligonucleotide primers and probes is linked to one or more detectable labels. In certain embodiments, the detectable label comprises one or more of a radioactive molecule, a luminescent molecule, a chemiluminescent molecule, biotin, an enzyme, a His tag, or an exogenous nucleic acid sequence. In some embodiments, the link is an indirect link. In some embodiments, the detectable label comprises: Hexachloro-Fluorescein (HEX), VIC fluorescent dye, 4-5-Dichloro carboxy fluorescein (JOE), Cy3 fluorescent dye, or TexasRed (TxRed).

According to yet another aspect of the invention, compositions that include a polynucleotide having a sequence set forth as SEQ ID NO: 2-7, 9, 10, 12, 13, 21-30, 36, 37, 40, 53-56, 59, 60, 63-66, 69-72, 81-84, 87-91, 94-97, 100-103, 122-174, 175, or another polynucleotide sequence disclosed herein, or a variant of any thereof are provided. In certain embodiments, the polynucleotide is linked to one or more detectable labels. In some embodiments, the detectable label comprises one or more of a radioactive molecule, a luminescent molecule, a chemiluminescent molecule, biotin, an enzyme, a His tag, or an exogenous nucleic acid sequence. In some embodiments, the link is an indirect link. In certain embodiments, the detectable label comprises: Hexachloro-Fluorescein (HEX), VIC fluorescent dye, 4-5-Dichloro carboxy fluorescein (JOE), Cy3 fluorescent dye, or TexasRed (TxRed).

According to yet another aspect of the invention, methods of identifying a test compound as an agent to reduce one or both of pathogenic *V. parahaemolyticus* in subject and pathogenic *V. parahaemolyticus* contamination of a substrate are provided, the methods including: (a) identifying the level of a pathogenic *V. parahaemolyticus* bacteria in a first portion of a sample using a method of any embodiment of any of the aforementioned methods of the invention; (b) contacting a second portion of the sample of (a) with a test compound; (c) incubating the contacted second portion of the sample with the test compound; (d) identifying the level of the pathogenic *V. parahaemolyticus* bacteria in the incubated second portion of the sample; and (e) comparing the level of the pathogenic *V. parahaemolyticus* bacteria in the first portion of the sample to the level of the pathogenic *V. parahaemolyticus* bacteria in the incubated second portion of the sample, wherein a decrease in the level of the pathogenic *V. parahaemolyticus* bacteria in the incubated second portion of the sample compared to the control level identified in step (a) identifies the test compound as a candidate compound to treat a pathogenic *V. parahaemolyticus* infection in a subject or as a candidate compound to reduce a pathogenic *V. parahaemolyticus* contamination of a substrate. In some embodiments, the incubation is at least 1 min, 5 min, 30 min, 1 hour, 6 hours, 12 hours, or 18 hours in duration. In some embodiments, the method also includes: confirming efficacy of the candidate compound for treating a subject having or at risk of having the pathogenic *V. parahaemolyticus* infection or for reducing contamination of a substrate by the pathogenic *V. parahaemolyticus* bacteria. In certain embodiments, the control level detected in step (a) is greater than zero. In some embodiments, detecting the level of the pathogenic *V. parahaemolyticus* bacteria in the sample comprises detecting in the sample, the level of a pathogenic *V. parahaemolyticus* polynucleotide of the pathogenic *V. parahaemolyticus* bacteria.

According to yet another aspect of the invention, methods of identifying the status of a pathogenic *V. parahaemolyticus* infection in a subject are provided wherein the methods include any embodiment of any of the aforementioned methods to identify a *V. parahaemolyticus* in a sample. According to yet another aspect of the invention, methods of identifying the status of a pathogenic *V. parahaemolyticus* contamination of a substrate are provided wherein the methods include any embodiment of any of the aforementioned methods to identify a *V. parahaemolyticus* in a sample. Polynucleotides listed in Table 2, Table 3, and in the Brief Description of the sequences may be used in embodiments of the aforementioned methods, kits, and compositions.

The present invention is not intended to be limited to a system or method that must satisfy one or more of any stated objects or features of the invention. It is also important to note that the present invention is not limited to the exemplary or primary embodiments described herein. Modifications and substitutions by one of ordinary skill in the art are considered to be within the scope of the present invention.

Brief Description of Sequences

SEQ ID NO:1 nucleic acid sequence of ST36prp

gtgaaaataaaaaacgtacgaattaaaaactatcgtctgctcaaagatgtgtcattttcgattgatgaaaaaacaaccattattgttgggcg
taacaatactgggaaaacttcatttgctgaggcttttcgcagctttctgaatcatgctggtcccaaggtgcgttatgaggatttcaatcaatc
ctgtctgtcaggtttcgaagatgcgctgaatgctcatcaaggcggagctgaagacgatgttgtgaggcccatgcttccaactatcgagct
agagttacttatcaattataaagaaaatgcggacgaatatggtgtcctcggcgactttattatcgattttaacgatcaattatttgaaaccatt
atcctgatttcttatcagttaaaagacggcaagatcggcgatttttttagtgggctcgataccatcaaacgaaagcagtatttttcggatctg
agggcgaggatagagcattactacgaagctacagtttacgcggtcgagccgacaaaccagagcaacaaagcacggcttgagttttcg
tcatttaaaaaattgctgctgtcgggtcttataaatgcacaacgcggtttagatgacgaaacgcacaatgaacgtgatgtgcttggtaagt
cgttaggtaatattttcaaaagtgcgagtagtgttggtgcgccagaagcttttaaggcgcaatcagctgaaattcataatgtcgtcgaagg
attacaacaagttgttgatactgactttcaggcaagagtgaaagctctacttcccaccttaaatatttttggatatccgggtctgcacgatcc
aaatttaagcgcagcaacagagcttaatgtgaaatcactgctggaaagtcacacccgagtgttttatcaacgagatgaccattttacattg
ccagaaacctataacgggctgggaatgcggaatctgatctttatcctcttccgaatttatgagtatttccgtgagtttcaaagccatcaaac
gccaccgaaagggcatgtaattttcatagaagagcccgaggcacatcttcacccgcaaatgcaggaagtatttattcgacagcttgagc
aaattgttgagcagttccagcgcgaactaaataaccagcaggtgtggcctgtacaatttattgtgagtactcattcctcgcatattgccaac
gaagctgattttagcaaggttcgttatttcctatccaaaaatggaaatgagaccaaagttaaggatctgggcgttgctttccaaagtgccg
aagcagcgggcgacaaagagtttttgcataaatacctgacgctcacaaaatgcgacctgtattttgcagaccgagcgatcctcgttgaa
ggggcgactgaaaggatattgttgcctcaaatgattaaaaaagttgacgcagctttagggactaacttaagacagaaatacctgtcagtc
gtcgaaattggtggagcttatgcacaccacttctacaaattcatagatttccttgagcttaaaaccctctttattactgacttagacgcagttg
attcaaaacaacatcatgccgctgtaatggtgagccaaggggacaggtcgagtaacgtcggcatttcaaaatggtttggcgaggagg
gttattctgacttagctactattagagctaaagattcagattctaaaattctcggatatcggcgacttgcttttcaggtagatgaagatgttag
tggtttatgcggccgcagctttgaagatgcatttatcttggttaatagccagttattccagctaaataacttaactggttccgcacttgaggc
cgctgtttatgacaaggcaaaagatatcggcaaaaagagcaaagctgattttgctattgaatattgtataagtaatacggattggctcgtc
cccaagtatatccaagaagggtgcgcatggttagacgaagacccaacagttgctgttgaaggagtgcaatcatga

SEQ ID NO: 8 nucleic acid sequence of ST36cps locus

atgttcttgatcatgtctgcttcttttgttgatcaagagctacaatcagaatttgggaaaataccaccaagtttttaccttaggaaataaaag
gctttttcaacatcaattaaaactggctccgaaaaacagcattgtttacctgtcattaccagagtcatttttttatttcgaatacagatgaaaaat
ggctaataaatcagggagttagaatcttaaaaatataccagagaatttgagtttaggagcatcgttagtagcatcgcttaacctttctgaacat
aacctagattctccatttagtgttttgtttggcgatacattattcaatgagttaccagttggtgaaaacattatttgtgtttcagaaagtagtaat
agctacaactgggctgtagtgacagataatgaaatgcactggcttactgatagtgaaaataaaattgactctaatgtaagaaatattgtca
acggttatttccgttttagttcgccgagaaatattattcgctctattactcggagtaattgggagtttctagatgggcttaatgaatatcataag
ataataggtctcaccgcagtatattctaagcagtggttagactttggtcacgttaacacttactataactctaaagctcattttactactcaac
gtgcttttaatgaattgagaattacttccgattatgtagaaaaatcgtcatttaaaagcaataaaattgcggcggaagccaattggttttcaa
ctcttccttattcactacgaaactatattcctcagtacctaggttctgagaaaaataatgagaaaataagatataagcttgaatacttatatctt
actgcattaaatgagctatatgtgtttagttcattaccagttaatacttgggagcaaattctaaggcaatgcatcgcatttattaaagattgcc
aacaagaaaaggcccctaatgattgtaattctaaccgtttagttgagttgtttggagataaaactcaagttcgattgaatgactattgtaata
gccagaatatatctatgtcgacacagtggatatttaatggagaagagaaagtatctttagaaggtattttgacggaaactgaacgttatctt
ccaaaatcggatgaaaagaaatggccattgtgtgtgttacatggtgatttttgctttagtaatgtgctttatgactttagatctaatcgtgttaa
aactattgatcctagaggtatgaccttaacaggtgaacaaacaatctatggcgacattcgatatgatattgcaaagttaagtcactcaatca
taggtatgtatgattggattattgctggttactatgaagttgatataaagggaaataatattaattttacaattgatgaaagtcatatacaaaaa

actattcaacagaagtttatcgagatggtgaaagtagaattcgatcttgaagcttctgaattaatagcaatgcagatccaattattcttgtcta
tgctaccacttcatcatgatgatattaatcggcaaaaagcactattcgctaatgcgtttaaattatacttttaatgaaaaggttgtaa

SEQ ID NO: 11 nucleic acid sequence of tlh locus

ttagaaacggtactcggctaagttgttgctactttctagcattttctctgcaacatagcggtgagttgctgttgttggatgcgtgacatccca
gaacacaaacttctcagcaccagacgctgcacactcagagcgcaatgcgtgggtgtacatgtaatcgacagacgatgagcggttgat
gtccaaacaaggatcgctcgcgttcacgaaaccgtgctcttctggcgcagaagttagcgtctcgaacaaggcgtgagtatcaaacaac
gtgatgttgtaaccttgcgctttgtagtacatcgcttgtgccttgatgaactcgttcatctcaagcactttcgcacgaattttgtcgatctcttct
tgtgttgagtacttaaactgaggcgctttcgtcgcgtctggcagtgtcatcaacatgaagttcttcgcacctgcgtccgtcaaacgaatca
gtgcttctgcataatctgctttcacttctggaacgccacggttgtagttcatgaagtcattcaaaccaaactcaagcgtaaacaaggtgttt
gctggtttgtagttcttcgccagttttgcgtaggttaagtacgaagaaacttgatcaccaacccctgttagcgcgatgtattggttctcacca
gccgcgccgccaactgcccagttgtagagcggaaggttcttcgctttggcaatgtattctgtccacacaaaaccgttggagaagtgacc
taagaaccagctgttcgggttagggaagcgccattgtgatgcgttaaagatgttgcctgtatcagacaagctgtcaccgagtgcaacca
ctttgttgatttgatctggctgcattgctgcgtcgttgctccagatcgtgtggttgtatgagaagcgattgtcagcggcgaagaacgtaatg
tctgcgttctcgttcgccaaatctaatgttgcttcacaacgctgacggataacgttttgcgacgtgttggtgtagaacatgtttttaagtgaa
acggagctccaccagtagccgtcaatggtgaagtagctaccatcttcgtttttgcccattcccaatcggtcgccggatcatctttcgagt
agctggtgcgataccaacagcgaacataggtataggtttggttttcttgcgtgctgatcacttcagacgctgaaaccatttctggtgataa
ggttggctcttcggcaactgcagaagcaagcgggagtaatgcagttaatagtgtgattgttttttcat

SEQ ID NO: 15 nucleic acid sequence of tdh locus

atgaaagaaagaatcattaatctgttctggttgtgtttagagcaggggggggagaattcttagttcattcttaattacgattctgatcgttaagc
acctaggggttgaacagtttggatcttttagtttagctcttgctatattgacagctttgggacctttgctggtcttggatttgactcgattttatt
taaaaaatttatcagcaatgaaggagatgaaaagactttactagggataagctgcttttcaagattatttattgcactatcgatcattagttta
actactttgataaatttaaacagtaatgctgtttatgttaatgtattaaacatactcgttctaggttttctatttgactcatttctagcatttaaagat
tattttcttgctaatcttaagaataagttttacacattttctacttttgtttcttcagtgattcaattggctcttgtttatactttagtacaaaaaaatg
ctagcatagaatacttcgcatggagctatgtattaacaaagttcatccaagcgttagtattaacttgttcatattataaaattaggcaatcatt
aatatttccaatttggaacaaagagttatcaagaaaattagttatagaatcttatccaatgatgctggccgcgtctattggtttactttatagtc
ttcaagaccaatttttattaaatactttcttggtgaatacgaacttgggttatactccgtaggcattaagtttattctcattcttatcgtattacca
acactaatctctaacgtattttacccaagcttagttaaaaaatttcattcaaataatattgaaatttataacaatcaattgcaggcgatatattta
ttatttttcattttaggtttgttactatttgcgttaatgtatttttcgtcggaaattgtaattgaaaaattatttgggactgattttgaacgatcaagtt
ctatcatggaaatatattctatattactggtggtatcattctttcaatctctgaataataaaatattaatattaaataatttacaatcagttatttta
aaagagcagtctttgcattaataacaaatgcaatcttgaatctcttccttatacctaaatttggtattaaaggtgcagcttatagtactgtgtta
tcagagatgttagtattaattagttatagcttcagaaaagatacaagatttatttttaaccatcaaatgagagctatatttttttgttaatttgttca
aaattgaaattataagaagtattaaaagatga

SEQ ID NO: 18 nucleic acid sequence of trh locus

atgaaactaaaactctactttgccttcagtttgctattggcttcgatattttcagtatctaaatcattcgcgattgacctgccatccatacctttc
cttctccaggttcggatgagctactatttgtcgttagaaatacaacaataaaaactgaatcaccagttaacgcaatcgttgatgactactgg
acaaaccgaaacataaaacgaaaaccatataaaagcgttcacggtcaatctattttcacgacttcaggctcaaaatggttaagcgcctat
atgacggtaaatattaatggaaataactacacaatggctgctctttctggctataaagatggcctttcaacggtcttcacaaaatcagaaaa
aacaagcctaaatcagaactattcttctgttagtgatttcgttggtgagaatgaagaatcattgccaagtgtaacgtatttggatgaaacgc
cagaatatttcgtcaatgtcgaagcatatgagagcggaaatgggcatatgtttgttatgtgcatttccaataaatcatcatttgatgaatgtat
gtcacaaaattaa

SEQ ID NO: 38 nucleic acid sequence of ST36flp locus

atgaaagaaagaatcattaatctgttctggttgtgtttagagcaggggggggagaattcttagttcattcttaattacgattctgatcgttaagc
acctaggggttgaacagtttggatcttttagtttagctcttgctatattgacagctttgggacctttgctggtcttggatttgactcgattttatt
taaaaaatttatcagcaatgaaggagatgaaaagactttactagggataagctgcttttcaagattatttattgcactatcgatcattagttta
actactttgataaatttaaacagtaatgctgtttatgttaatgtattaaacatactcgttctaggttttctatttgactcatttctagcatttaaagat
tattttcttgctaatcttaagaataagttttacacattttctacttttgtttcttcagtgattcaattggctcttgtttatactttagtacaaaaaaatg

ctagcatagaatacttcgcatggagctatgtattaacaaagttcatccaagcgttagtattaacttgttcatattataaaattaggcaatcatt
aatatttccaatttggaacaaagagttatcaagaaaattagttatagaatcttatccaatgatgctggccgcgtctattggtttactttatagtc
ttcaagaccaatttttattaaatactttcttggtgaatacgaacttgggttatactccgtaggcattaagtttattctcattcttatcgtattacca
acactaatctctaacgtattttacccaagcttagttaaaaaatttcattcaaataatattgaaatttataacaatcaattgcaggcgatatattta
ttatttttcattttaggtttgttactatttgcgttaatgtatttttcgtcggaaattgtaattgaaaaattatttgggactgattttgaacgatcaagtt
ctatcatggaaatatattctatattactggtggtatcattctttcaatctctgaataataaaatattaatattaaataatttacaatcagttatttta
aaagagcagtctttgcattaataacaaatgcaatcttgaatctcttccttatacctaaatttggtattaaaggtgcagcttatagtactgtgtta
tcagagatgttagtattaattagttatagcttcagaaaagatacaagatttatttttaaccatcaaatgagagctatatttttgttaatttgttca
aaattgaaattataagaagtattaaaagatga

SEQ ID NO: 41 is amino acid sequence of ST36prp, encoded by SEQ ID NO: 1
MKIKNVRIKNYRLLKDVSFSIDEKTTIIVGRNNTGKTSFAEAFRSFLNHAGPKVRYED
FNQSCLSGFEDALNAHQGGAEDDVVRPMLPTIELELLINYKENADEYGVLGDFIIDFN
DQLFETIILISYQLKDGKIGDFFSGLDTIKRKQYFSDLRARIEHYYEATVYAVEPTNQSN
KARLEFSSFKKLLLSGLINAQRGLDDETHNERDVLGKSLGNIFKSASSVGAPEAFKAQ
SAEIHNVVEGLQQVVDTDFQARVKALLPTLNIFGYPGLHDPNLSAATELNVKSLLESH
TRVFYQRDDHFTLPETYNGLGMRNLIFILFRIYEYFREFQSHQTPPKGHVIFIEEPEAHL
HPQMQEVFIRQLEQIVEQFQRELNNQQVWPVQFIVSTHSSHIANEADFSKVRYFLSKN
GNETKVKDLGVAFQSAEAAGDKEFLHKYLTLTKCDLYFADRAILVEGATERILLPQMI
KKVDAALGTNLRQKYLSVVEIGGAYAHHFYKFIDFLELKTLFITDLDAVDSKQHHAA
VMVSQGDRSSNVGISKWFGEEGYSDLATIRAKDSDSKILGYRRLAFQVDEDVSGLCG
RSFEDAFILVNSQLFQLNNLTGSALEAAVYDKAKDIGKKSKADFAIEYCISNTDWLVP
KYIQEGCAWLDEDPTVAVEGVQS.

SEQ ID NO: 42 is amino acid sequence of ST36cps, encoded by SEQ ID NO: 8
MFLIMSASFVDQELQSEFGKIPPSFLPLGNKRLFQHQLKLAPKNSIVYLSLPESFFISNT
DEKWLINQGVRILKIPENLSLGASLVASLNLSEHNLDSPFSVLFGDTLFNELPVGENIIC
VSESSNSYNWAVVTDNEMHWLTDSENKIDSNVRNIVNGYFRFSSPRNIIRSITRSNWE
FLDGLNEYHKIIGLTAVYSKQWLDFGHVNTYYNSKAHFTTQRAFNELRITSDYVEKSS
FKSNKIAAEANWFSTLPYSLRNYIPQYLGSEKNNEKIRYKLEYLYLTALNELYVFSSLP
VNTWEQILRQCIAFIKDCQQEKAPNDCNSNRLVELFGDKTQVRLNDYCNSQNISMST
QWIFNGEEKVSLEGILTETERYLPKSDEKKWPLCVLHGDFCFSNVLYDFRSNRVKTID
PRGMTLTGEQTIYGDIRYDIAKLSHSIIGMYDWIIAGYYEVDIKGNNINFTIDESHIQKT
IQQKFIEMVKVEFDLEASELIAMQIQLFLSMLPLHHDDINRQKALFANAFKLYFLMKR
L.

SEQ ID NO: 43 is amino acid sequence of tlh, encoded by SEQ ID NO:11
MKKTITLLTALLPLASAVAEEPTLSPEMVSASEVISTQENQTYTYVRCWYRTSYSKDD
PATDWEWAKNEDGSYFTIDGYWWSSVSLKNMFYTNTSQNVIRQRCEATLDLANENA
DITFFAADNRFSYNHTIWSNDAAMQPDQINKVVALGDSLSDTGNIFNASQWRFPNPN
SWFLGHFSNGFVWTEYIAKAKNLPLYNWAVGGAAGENQYIALTGVGDQVSSYLTYA
KLAKNYKPANTLFTLEFGLNDFMNYNRGVPEVKADYAEALIRLTDAGAKNFMLMTL
PDATKAPQFKYSTQEEIDKIRAKVLEMNEFIKAQAMYYKAQGYNITLFDTHALFETLT
SAPEEHGFVNASDPCLDINRSSSVDYMYTHALRSECAASGAEKFVFWDVTHPTTATH
RYVAEKMLESSNNLAEYRF.

SEQ ID NO: 44 is amino acid sequence of tdh, encoded by SEQ ID NO: 15
MLAAFKTFAFELPSVPFPAPGSDEILFVVRDATFNTNAPVNVKVSDFWTNRNVKRKP
YKDVYGQSVFTTSGTKWLTSYMTVNINDKDYTMAAVSGYKRGHSAVFVKSDQVQL

QHSYNSVANFVGEDEDSIPSKMYLDETPEYFVNVEAYESGSGNILVMCISNKESFFEC
EHQK.

SEQ ID NO: 45 is amino acid sequence of trh, encoded by SEQ ID NO: 18
MKLKLYFAFSLLLASIFSVSKSFAIDLPSIPFPSPGSDELLFVVRNTTIKTESPVNAIVDD
YWTNRNIKRKPYKSVHGQSIFTTSGSKWLSAYMTVNINGNNYTMAALSGYKDGLST
VFTKSEKTSLNQNYSSVSDFVGENEESLPSVTYLDETPEYFVNVEAYESGNGHMFVM
CISNKSSFDECMSQN

SEQ ID NO: 46 is amino acid sequence of ST36flp, encoded by SEQ ID NO:38
MKERIINLFWLCLEQGGRILSSFLITILIVKHLGVEQFGSFSLALAILTALGPFAGLGFDS
ILFKKFISNEGDEKTLLGISCFSRLFIALSIISLTTLINLNSNAVYVNVLNILVLGFLFDSF
LAFKDYFLANLKNKFYTFSTFVSSVIQLALVYTLVQKNASIEYFAWSYVLTKFIQALVL
TCSYYKIRQSLIFPIWNKELSRKLVIESYPMMLAASIGLLYSLQDQFFIKYFLGEYELGL
YSVGIKFILILIVLPTLISNVFYPSLVKKFHSNNIEIYNNQLQAIYLLFFILGLLLFALMYF
SSEIVIEKLFGTDFERSSSIMEIYSILLVVSFFQSLNNKILILNNLQSVIFKRAVFALITNAI
LNLFLIPKFGIKGAAYSTVLSEMLVLISYSFRKDTRFIFNHQMRAIFFVNLFKIEIIRSIKR
.

SEQ ID NO: 57 is nucleic acid sequence of ST631 locus
Agttcatcaggtagagagttagaggaatatgtacagagtacgtatcgctttcttctaaatatgaaagatgaaggtgtaactgtagagcga
aatatatacttatctggaaagtctggagccaaacatcaaattgatgtcttttatgagtttaaaacagctggtattacacatagggtagctattg
aatgtaaagaccactcacgcccagtggaaaaagggaaggttcaagagtttgcatataaactgcaagatatcggtgggatttctggtgtc
atggtatctcaagcagggtatcaatcaggcgctgaattgattgctaagcaggcagatatcctgctgaaaactactgatgaacttcctccta
ctccttggttaatggctgagaggttggaaagtgtagctcttccaacagaaaattacagggggaaccattttgggtgatcatggagcata
gtgaaggtaaagttaatggctcatactatggtaacgaagataatggtcgtaagttcataccttgtttttctccaaatatcatgctcaactaaa
ttttgatgaaggtggacttgatgaatcttgttggtgtgttcgtggtttaccaaggcacgcttttagggcatttttgctgttactagaattatttga
gcgccaaaaggtcgagcctatgatctgttttagacctcctggcgatactagtgaaataggctgggcagggttagttactacacgtgactt
acttgtaaaagagtactattgtgaagatctgcctagagtgctaaacaaatccgcttaa.

SEQ ID NO: 58 is ST631 amino acid sequence
MSSSGRELEEYVQSTYRFLLNMKDEGVTVERNIYLSGKSGAKHQIDVFYEFKTAGIT
HRVAIECKDHSRPVEKGKVQEFAYKLQDIGGISGVMVSQAGYQSGAELIAKQADILL
KTTDELPPTPWLMAERLESVALPTENYRGEPFWVIMEHSEGKVNGSYYGNEDNGRK
FIPLFFSKYHAQLNFDEGGLDESCWCVRGLPRHAFRAFLLLLELFERQKVEPMICFRP
PGDTSEIGWAGLVTTRDLLVKEYYCEDLPRVLNKSA.

SEQ ID NO:67 is ST34reg nucleic acid sequence
atggatttgaacttaatcaatactttccttgttgtcgttgaatatcagtcgtataccaaagcggcggaacacttaggggtaacacaacccgc
aattagtgcatcgatgaaacgactagaacaactatctaacaaaaatctttttgttagaaaagggcgaaatattgagttgacctcaaccgca
caccactgggttccattattcagacgagcattaagcataattaatgatgctgtcattgagcaagccacatttcaagtctgctgtaccgaac
cctctttttcgagactgaccgcgtctccaagcttttcgttgcgttgcgctcctgtttcgtccctctccttacttgacgatcttcgtctacacaaa
gtggacttggttattgacaacctccccactatagaaacctcatttgtatgtgagttggtatacgaagagccaattgtggtgatttgccgtca
aggacatccacgtataactggcagtacatttaactcatcaatgttctatgccgaacaacactgtgtattggttgatacggaatatcgggca
gtgaatcttggtggggcgctcctcgaccctacacaacatttacacataggcatgacaaccgtatctctatctggaatggtactgaacgtat
ctaagctagactaccttggaatcttgccactctcttttgctagagaatggcaggattctcttaaactacaaatattgccatgtcccataaaaa
gtcaatctattggttataatatgatttatcacaaaagagatgaacataatgttgcccaccaaaaactgagaaggcaaattcgtcacgacct
cgttcaaaacttgatggttaggaatttctga.

SEQ ID NO:68 is ST34 amino acid sequence

MDLNLINTFLVVVEYQSYTKAAEHLGVTQPAISASMKRLEQLSNKNLFVRKGRNIEL
TSTAHHWVPLFRRALSIINDAVIEQATFQVCCTEPSFSRLTASPSFSLRCAPVSSLSLLD
DLRLHKVDLVIDNLPTIETSFVCELVYEEPIVVICRQGHPRITGSTFNSSMFYAEQHCV
LVDTEYRAVNLGGALLDPTQHLHIGMTTVSLSGMVLNVSKLDYLGILPLSFAREWQ
DSLKLQILPCPIKSQSIGYNMIYHKRDEHNVAHQKLRRQIRHDLVQNLMVRNF.

SEQ ID NO:73 is ST674 nucleic acid sequence
gaagatggtccaagagggaagcatagtctccctctttcaaaaagggataatttagaaaatctggttttactttgtaaaacgcatcataaatt
agttgacgatcacgtagatgagttttcggttagtgatttgacaacactgagagaggagcactttaaatgggtgtcaaataagttaaatgag
cctaggcaatgggagtgtaacttaagtcagcttacttatattaatgtacctcgattatcgatgttatcttctaggttgggctatgaagtagattt
agatgaatacggtaagtttgaaacattatattcactgcgttggtcgcttaataaattgatgaggcaatttctatcaacattaaacaagataaa
cgtgaatacattggacttttcctctgttaactatcccgatatccggttggttgga gctacttgctcaatctctgat.

SEQ ID NO: 74 is ST674 amino acid sequence:
MAVLEKTRNMLWALSAGRCAYCKNKLVVESKKKNFSLVGEVAHIVAQKEDGPRG
KHSLPLSKRDNLENLVLLCKTHHKLVDDHVDEFSVSDLTTLREEHFKWVSNKLNEP
RQWECNLSQLTYINVPRLSMLSSRLGYEVDLDEYGKFETLYSLRWSLNKLMRQFLST
LNKINVNTLDFSSVNYPDIRLVGATCSISDSFRTKNVPMIGRDDKDPVTFCGDLKKDP
HIYKKYPNFKLVMRIQPSWITTSTAFLAFRPSGGVSTFSGLITVSEVDVENSVIYAIPLV
LGLPVSDFELMMKEPKLFREEDSSVVGKKINKKTSLIEFEDLEKAIEQDTKYVNPPDC
CDVCRASLENQTYFVDGAIKSSSAWAFLCEYCFEKDGVGIGWGLGQLFKKNKHNE
WLLVGGFAPESDDDYYI.

SEQ ID NO: 79 is nucleic acid sequence of ST1127 locus
cgtaaagtaaaagagcctggtcttgcatccaccggaattaatctatcttttaggttaatagtacgaacaaactcaccgtatagggatgacc
aaccatctagcactttagaaatgctgtttagttgcatagtccctcgacgactaggacgagaaatgtgaatttcgtgatttacagatagaatt
gattcttcgatttcttccgactcaatgtcttctattctgaagtcgcatggtggaagatcctcttctgggatatcttctgcagctaagatctcaat
attatggttattaccatcgaatgcagtaaagattttgaaaattgaattttcttcagggtaaagaaaagcatctctcaaaaggcgagaacctg
tttcgagagattctaatcgtttggcagacataggaacgtaaagccagacgtctccatcttctagttcatctatccaataaactaagaaatag
aggcctagggcattcactacactaaatagtttaggcccatcgtattcttcgtagaccttatttatctctatagtgcctaagcttttctcttgaac
taaagggtacattacatcattcttctatacattcgaatcgcgtgtgtggttcagagtccgttttaaaccatacagttatatggtggtggcctcc
ggtttctttcatcactccgtcgtcttcaataagcgcgcccgttgacaatagcttttgtcgaagaggtttgaacaacttcttctttttgagcgctt
ctttttttctgagtgaataacgatgagccataaagcatttctctatcagttggctttttcgcttttcttccagggttttcttcaagtgtagaactaaa
acaggtttttacaggtggattactattcaccaatcggaaagctgttccagatgcagcgatcgcatcttttggaggcatatcatcgggatag
taatctgggtaatcactcatacatatgcacaggttcgtaaaaaaacattgttgcatattagttaggttctatttgttaatcaacggcctaatac
gtgtttgtgtgatagaaatcgcttcctctgatagatcattaccccaaaatttgcggtcatgactcaaagctgaaacacctacagcgccagc
ccccataaatgggtcacaaacaagttcgccttctgtcgagctttgttttattagaatatccatcagttcaacgggtttttctgtaggataacct
ctatgaatgcggggaactgaa

SEQ ID NO:80 is nucleic acid sequence of ST631 hypothetical protein locus, which is expressed in pathogenic and non-pathogenic *V. parahaemolyticus* bacteria
ttttctggagataatgaattgagcagagaaatatacgatatggaacgacgctgtgttaataggcacaatgatttagcgggtcacattaatc
ggatgagtcagcagcaagctcgtatcatgacaaacgaaaatatagcggcgacaaaagaagtagttgctcacgtatacggaaaagcta
actcttacagcaatgtaattattgctgctggttatgtcggtttctttacttttatggtcaagtctaaaaagtgatttacctcaatgggctattttaag
ctcaggcgcacttatactgatatcactgatgacatttataggatttgaactatataaaatgatcagtgtttcagtacaaatgcacagagtctc
aaaacggcttcaaaagcccgacatgtcgtctctaagtgaaatacagcgtattgagcaaaaaagcgcactgattaacgctagggtctgg
gtttttacagttattccaacagttttgtcgggatttggagccggtcttatattgttatactgctttctggttgactttatagcaccatatttgcaac
aaacctaacaaacaatttaagagtgattcagcacgcttggcattttcggtttgggttgagttcggtgtttacggtggtcaaattgatagtcgt
ggttgcgtgcttcacaccttaattgggcgttaggctttgcgtacaaaatcgaataggtattaaaattgaaaggtacagaaatacaagctca
aaatgtaaccaaaccgattcagcttttagctgcttggcttgttggtttaatcgcaattaacggctcttttcttggagctgctaatgttatttcaa

ctccaacttgggctgccggattactagttattgcctcggtatttaatgtccctctgtttctaggcttaatttttttcttacaaactaagtttcgtgc
agaactacaagaagattcgttctatgctaagcacctagataaagtaacaggggcaatgcagactagtgatacatcgataaaagcggaa
gtcaaagctgatgttcaaaaaatacaaaaggaaaatgcggaaaactatgaagcaatacagaaagaacttcgggctgtaatgaattcctt
aaatggtattgcatatagttctggaaatagtgaacttgagcagaaagtcgtagaagcccagcaacgagtagcaaaattagagtctaaca
ataaaaagttctcaattaagcttgccttgaataaagtactagatcaatttaatgctatttcgcatgaacttacagcttcaggttatattatcgcg
gacgtttttgggagcgacaaacttaactctaaagctatctcttacattgatggtgtcgataaagttttattgcttgacatagttaatactttaaa
gccttttggatttgatcgaattgactatgatcctgaagataagggctttgatagagatgcaaaagtttacattggctcctatatagatgatttc
cctgaagctagaaggtcagttttaatttctcatgaagttatggatattctagaagatgatactaagtcactgagagatttaaataatttcattc
gagatcaacaagcctaacaaggcgtttaagacagattccca.

SEQ ID NO: 85 nucleic acid sequence of MAVP-36 and MAVP-46 (Orf3 WP_005477619 to hyp WP_047715660) and (any f237-like phage, ORF 3-5)

atgtctgtatgcgtcaccgtcgttaaccagtatggcaatttgaaagcaacgaaaacgcctgttgcggattgccaagaatacgtgctgatttt
cggcggtggactaccaagaatataaggaaccagtcctcttcaacggtgacttgttcctgtatgtcagtggcgtgctcttgatcaacatggt
cgttggtcactgggtgggtcgtgttgttcgccttatgagtaaaaggtaaatcttatgaaaaaactagaacttgttgtaactaacgtaaaaca
cgcagtcgtaaacaaaaaaaccgcagctggcgctgctcttatggtcgcgtctgtctctccggccttcgctgaagtcgatatcacgggcg
caatcaactctgcggtatccggtggtcaagctaacgtatcactggttgtggcgggtctaattggtatggctgcactgggctttggtgtgac
catggttgttggcttcttacgtcgctaacggttcacctctatgcctcctttatcgggtaatttacttggagatgttctcgctatcgttctaggtgt
tgcctttgcgggggcattcctccacggctttgtgagtggcatcaatactcactaatcaacggataaaggggcttcggctccctttttttatt
ggttttatacaatgaatcactatctccgttttttattgtccttgttattctatgcgctagtcatcatacgtatgctttagaagcacgtattagtcat
atgcaaatgaggggttgtggctctcaaggtgattgggttgacccttacaaggtgaatacttgttttttggatactgggtatttcgactcatgc
acatttgagaagacatcatatgctaatgctcgcgatccctatcaaacagtttgtgataatgggctcggtctttcttattctgaggttcgttgtc
cagaaaatagcgaatttgacccttcaaccttacgttgtaaatcggtttgtgaatatggcaagaaccctgacggcacctgcatggatgcttg
ccagttcaaacagtccattggtgatacggtgaaattgcattggcaccctgccatatacggcgaactggtgacaggcgcgtgctacgga
gactacggtgccactcgatgcgaagtgaccaaaaacgaatccaccattatttgtactggcgttcctgatggacagtacacgcccgactc
tcaatgctctctgcgctttgcttacactggacgtcagtgtgacggtggcacacttttctggggcgtgaatgggccagacgagccaatcat
tccaccggatacgccagaagacccaactcatgaccctgatgacccaacaggcgagattgaagacccaagtgtcctacccgacgattc
aaccaacacggttaatcccggtgtcgttgatgataaaccggatgtagaagaccctgacacggatgaatcgacagacacggcagtcctt
tctgctattaaagggcttaacgtggatgtgaacaaaggcattcatgatcttaacgtcgatatcaaccagtcacacgctgacatcaccaac
gcggtgattgatgtgaaaggctctttggtcgataacacccaagccattcaagaacagcaaatcaatgacaacaagatttataacaacac
caaggcactcatccaacaggccaacggcgatatcactacggcggtgaacaacaataccaacgccaccattggtattcgtaacgattta
aaagggcttggtgattcaatgggcgaactcgatagcagcttaaatgcgattgagggtctattgactggctcagagtttggtacacctacg
ggcaccgctatcactggcgaaatcttcacggcagaagactttgccaacctgcaaaccacgatagatgaaaaagccgaatccatccaa
ggctatgtggacgatattaaaggcttaatcactatcggcaccaacttcaacaacggcacattaagcgacaagtcttttaacatcaaaggc
gcaaccgttgaatcaggactacagcgttttgatgcggtatcgggctacgtgcgccctgtcgtgctgttcatttgtgccttaatcgccctttg
ggttctgtttggtaatcggagtaaataa.

SEQ ID NO: 92 nucleic acid sequence of MAVP-26 Orf10-to Hypothetical A, (North East Atlantic phage only, not 10290 phage)

taatttaacataatatacataatgcgcactgatatagggttcctgtggacttgcaatcactaaggccaatccagcacaagccattgaaat
gcttgataatccaagtccgttaaacttttttcctatgttctgccatagtgtctgcgcttcgtgcgtttttgctttatccatcgccaagccaatcag
tgccttttctttgtcttcaccaatagtttcagcaagcataagtatctgattttcattgagataacttcgacctttttcttacttctgtgagcatttgag
ggcttacacctaggtcatgagcaatctgcttgtattgaatgtagttcatttgctctttataagcatcaatgagcttgtttgtgtacatttctgcttt
tcctctaatcacgctattggactgattttagtcttttagtacagattttgctgtgttgacggtacagaaatatctgtatttaatcgctacagaattt
actgtatcagaccgccttagctttgggcgtttgcccttgacgctttcgcgcttggctttggcggtcactctctcaactagtcaggtggttgta
atgatcgtattagaaactgaagttcaaaacgtcaatgttaaaacgttaacgctccgtcacttcgcagtgtctcacgttccagctttcaaactt
tgttacatcactacaactgacatctttgaagaagtcgtcgttcctttcaatcactttggctactgcattagcacattcgaaaacaaccaagag
tcctttggttatctctcggttggcgattacgagtttcgctttgaatctgacgagcatgaagttctatgtcgtttcttaggcatgacaccttcaaa
agcgacggctttagaggctcagtaa.

SEQ ID NO: 93 nucleic acid sequence of MAVP-26 Orf10-to Hypothetical A, (North East Atlantic phage only, not 10290 phage)

taatttaacataatatacataatgcgcactgatatagtggttcctgtggacttgcaatcactaaggccaatccagcacaagccattgaaatg
cttgataatccaagtccgttaaacttttttcctatgttctgccacagtgtctgcgcttcgtgcgtttttgctttatccatagccaagccaatcagt
gccttttctttgtcttcaccaatagtttcagcaagcataagtatctgattttcattgagataacttcgacctttttcttacttctgtgagcatttgag
ggcttacacctaggtcatgagcaatctgcttgtattgaatgtagttcatttgctctttataagcatcaatgagcttgtttgtgtacatttctgcttt
tcctctaatcacgctattggactgattttagtcttttagtacagattttgctgtgttgacggtacagaaatatctgtatttaatcactacagaattt
actgtatcagaccgccttagctttgggcgtttgcccttgacgctttcgcgcttggctttggcggtcactctctcaactagtcaggtggttgta
atgatcgtattagaaactgaagttcaaaacgtcaatgttaaaacgttaacgctccgtcacttcgcagtgtctcacgttccagctttcaaactt
tgttacatcactacaactgacatctttgaagaagtcgtcgttcctttcaatcactttggctactgcattagcacattcgaaaacaaccaagag
tcctttagttacctctcggttggcgattacgagtttcgctttgaatctgacgagcatgaagttctatgtcgtttcttaggcatgacaccttcaa
aagcgacggctttagaggctcagtaa.

SEQ ID NO: 98 is nucleic acid sequence of MAVP-26, also referred to as Vipa 26. Region from HypD (WP_047724020) to Orf9 (WP_047724015).

cgctaggttttatatgtcaaggatatgaaatggctaagtttttaaatacaagtgctacaaactactatctcgaagagcttattaagaatgcttc
cgaaagactgattctaattagccctttctcaagcttaatgatcgcattcgagagcttttggaagacaaggaccgattaaaaatcgatatta
gaattgtctatggcaaaagcgaactacaaccggatgagattaactggcttaaaagcctctcctttgtgcgtactagtttttgtaaaaacctc
catgcaaagtgctacatgaatgaaagtgcttgtatcattacaagcttaaatctctatgagttcagccaagtaaacaataatgaaatgggtat
cttcattgaccgtgacgaagaccccaatgtctacaaagattcctacgaggaagctcaacgcattattcgtattgtctatggcaaaagcga
actacaaccgaaagttcgagctgctaatttagatacggaacttactgaaaagcctgttacagataatgaactaattaaactcagctcctct
aagttagccaaaaagcacaaacttaaaacagatgagtttctcagcttgtgtgttaataaaggatacttaacgttagatgacggaaagcact
cattaaccgatgaaggtaaatcttcaggtggtgagtttaaatacagcaaacgtttcggtccatattttgtctggccagagtccttggaagta
tcgtaactattgggttcaatattttaaagtgagtacagcaaacatgaagattaaaataagagctatgttaagtttagctgccgcccttagttt
gataggttgcacctccactaacttcagtgaatataagggtctacctggccataagtccatagctgttggctctaatggcgttatcgcatatt
cctcatcgaagcccaatgccgaaacagccattacaacagccattgataaatgctcttcaatcggtggcaaagattgtaagtttatcgatgt
tgatggttatagtcctatcggtaataaaacatatatgtatgatgatcgcgcagctagtgttgattacctagaaaaaggttcatacgcagtaa
ggtataaatgtgagcaattaacgaaagcaatggcacaatcattatttcgttcaggccatacttacctagatggagactcagatgggaagc
cttgtgagtcaaacctttggagctcatattattcgacagataccagtagcagaaaagcaaaaggtactaactgtcattatgttcgtgggtat
cgaagaaaaaatggtacttatgtcagtggttatacacgctgtcgttgatttacaaggctccattcggagccttttttcatataatttttctttaata
ctcttgcatacttgagaatctgatgcgcaatagcaatgtcatttgacgcacctaactcaagtaaagcaaccccaattaatacttgctgcgc
agtaaccaactgtcctgttggaagttctaaccgatcatgcctcattacgaagttttcccaatcttcacagctgctaagttccctacccttattc
attcgcattaggcgcttacactccggaggaatggatttccccttgtcccattccttgatcgtcctcacagtttttaaacaaagtttggcagctt
cttcgacggttaaaccacattcaaattcacgaaaaatatagtttttagtcatttcgtgatacttcattgaattgtccctcaaaagagagacattt
tataggatacgcatatgcaatcgcattcaacataagcgcccataatgcgcaccagg

SEQ ID NO: 99 is nucleic acid sequence of MAVP-36, also referred to as Vipa 36. Region from HypD (WP_047724020) to Orf9 (WP_047724015).

Atggctaagtttttaaacacaagtgcaacaaactactacctcgaagaactaattaagaacgcttctgaaaggctgatcctcatcagccct
tttctcaagcttaatgatcgcattcgagagcttttggaagacaaagaccgattaaagatagacattcgaattgtctatggcaaaagcgagc
tacaacctgatgagattaactggcttaaaagcctctcctttgtgcgtaccagtttttgcaaaaacctccatgcaaagtgctacatgaatgaa
agtgcttgtatcattacaagtttaaacctctacgagtttagccaagtaaacaataacgaaatgggtatcttcattgaccgtgacgaagacc
ccaatgtctacaaagattcctacgaggaagcgcaacgcattattcgtattagtgatgaagttagaatctcgttagagaaagttgaagctgc
taaattagatacggaatctactcaaaagcctgttacagagaatgaactaattaaacttagttcctctaagttagctaaaaagcataaactta
aaacagatgacttccttcagatgtgtgtaagcaagggctacttatctttcgaagatggaaaacattctttaaccgaagaagggaaatcgtt
gggtggtgagttcaagtacagtaaacgttttggtccttactttatctggccagagtcattagaggttgaatagaaaaataaggctcctgttg
gagccttcaatcacactatttttctttagtattcttgcatacttcaatatttggtgagccacctttatatcattcgatgcacctaactcaagtaaag
caaccccaatcaatacttgctgcgcagtaaccaactgtcctgttggaagctctaatcgatcatgcctcattacgaagttttcccaatcttca
caagagctcaattccctacccgtattcatcctcatcaagcgcttacactctggaggaatggattttcccttgtcccattctttgaccgttctca

cagtttttaaacaaagtttggcagcttcttcgacggttaaaccacattcaaattcacgaaaaatatagtttttagtcatttcgtgatacttcattg aattgtccctcaaaagagagacattttataggatacgcatatgcaatcgcattcaacataagcgcc.

SEQ ID NO: 21 gccgaactcaaaagcagtaa.
SEQ ID NO: 22 ggtgtggacgaccaataaatcaag.
SEQ ID NO: 23 gggctccgtgtagaagtgg.
SEQ ID NO: 24 atttggtgattgtagaagtagatgg.
SEQ ID NO: 25 acgccttagactagaaataggag.
SEQ ID NO: 26 ttgacaacccagaagcagattgg.
SEQ ID NO: 27 aactcttaatgaaggctaatacatct.
SEQ ID NO: 28 atttgcgtaatcgcttatcacg.
SEQ ID NO:29 tcttgaagtagaactactgtgacg.
SEQ ID NO:30 tttcgcctaatacgtcttcgtgac.
SEQ ID NO:31 cgatgaaagatcgtaaaagagacg.
SEQ ID NO:32 catttccagagactgaaaacctg.
SEQ ID NO:33 accttatcttggaagtagaagtgg.
SEQ ID NO:34 gtattgtttgtatacgggtaaaggc.
SEQ ID NO:35 aaccgaagctataaaagtcataga.
SEQ ID NO:39 accaacacaaatctcgtccc.
SEQ ID NO:40 acaaccattatgctattcttactct.
SEQ ID NO: 49 tcctgcgtcaatgcgaactac
SEQ ID NO: 50 gattgcgtaacagggaaacatc
SEQ ID NO: 51 ttgcgtaacagggaaacatc
SEQ ID NO: 52 gattgcgtaacagggaaaca
SEQ ID NO: 55 tcaagtagtccatctctcaatctcct
SEQ ID NO: 56 agaagcaatggtatcatactcggt
SEQ ID NO: 61 acccgcaatccgaaacg
SEQ ID NO: 62 cccgaagatgctgaaagacga
SEQ ID NO: 65 aggaacaacagcaacttatagtcag
SEQ ID NO: 66 ctatgccaacagtacggatacaca
SEQ ID NO: 71 cttctaccaggttctcccttcg
SEQ ID NO: 72 gatagtctctaactcgttcatcg
SEQ ID NO: 77 gcatttcattttctcggaccag
SEQ ID NO: 78 aagtcaaggggcgtaagta
SEQ ID NO: 83 tcgttgcttttgcggaca
SEQ ID NO: 84 tggcatagtggttacctgaca
SEQ ID NO: 86 gcatagtggtacctgaca
SEQ ID NO: 89 aaagaatgaagacactcgtaaact
SEQ ID NO: 90 ctaatgactcggagatttcggcag
SEQ ID NO: 91 aatgactcggagatttcggcag
SEQ ID NO: 96 ttcacgatgtacttactttcacga
SEQ ID NO: 97 agttacttcatagtgctttactgat
SEQ ID NO: 102 cttacgacggttgtacctatattta
SEQ ID NO: 103 ctgtttccatacgacggttttcac
SEQ ID NO: 107 tgagttgtgttcttctctagctgtt
SEQ ID NO: 108 ctactcgccaactacaggtt
SEQ ID NO: 109 gcgagcgcaagtgctttggca
SEQ ID NO: 113 agggaaaaggacggggg
SEQ ID NO: 114 gcgacggtaacatatcagaaatag
SEQ ID NO: 115 actgtaggatgtactgacac

SEQ ID NO: 119 aacgaaagtcaaacgataaccga
SEQ ID NO: 120 acaaatggcagtatatccgcgaa
SEQ ID NO: 121 tctttatgttgttagttttgact
SEQ ID NO: 131 actgcgtcgaaatccctgattgaat
SEQ ID NO: 132 gcgatttggtgattgtagaagtagatg
SEQ ID NO: 133 cgctcctcccaataagactgaatc
SEQ ID NO: 134 gattcagtcttattgggaggagcg
SEQ ID NO: 135 tttagatacttaaacatcttcaccacac
SEQ ID NO: 136 tatggacagtcagcagctttaaccacctc
SEQ ID NO: 137 tatggacagtcagcagctttaaccacct
SEQ ID NO: 138 tatggacagtcagcagctttaaccacct
SEQ ID NO: 139 gttgtagtacggcgacattaccactc
SEQ ID NO: 145 tgtagtacggcgacattaccactc
SEQ ID NO: 146 ttagacaagaccaacacaaatctcg
SEQ ID NO: 147 tttccagggtttcgacagttatatc
SEQ ID NO: 148 ttcagaaaagtagtggaagtaacg
SEQ ID NO: 149 gatccccaacttgtcaaacctagaaaatcaaat
SEQ ID NO: 150 gatccccaacttgtcaaacct
SEQ ID NO: 157 cgactctccaacctttcacatcg
SEQ ID NO: 158 agaagcaatggtatcatactcggt
SEQ ID NO: 159 atagccaccctaaagaccaca
SEQ ID NO: 160 gtaattggttcctcatcctccttca
SEQ ID NO: 161 cccactagtacctcgtatcacttccattt
SEQ ID NO: 162 agagttcgtcccatagttagtccgc
SEQ ID NO: 163 agagttcgtcccatagttagtccgc
SEQ ID NO: 167 ctgtagctatacccacggc
SEQ ID NO: 168 gctctgctacgtcggtaag
SEQ ID NO: 169 agagtacgcagagggaccacttacac
SEQ ID NO: 173 aacccgcaatccgaaacgcatgt
SEQ ID NO: 174 actatcagcaccaacgcacg
SEQ ID NO: 175 tcgacttagccaaaccaatgt.

Brief Description of the Drawings

Figure 1 provides a diagram showing identification of ST36 12 clade strains from among northern New England clinical isolates of *V. parahaemolyticus*. A consensus neighbor-joining tree was constructed from four concatenated housekeeping gene loci including *dnaE, dtdS, pntA,* and *tnaA* sequences (1868bp) by using a Jukes-Cantor model, with statistical support assessed by 1,000 bootstrap re-assemblies to identify all ST36 complex strains. The bar indicates 0.2% divergences, and branches with less than 70% bootstrap support are unlabeled.

Figure 2 provides a diagram showing the distribution of putative diagnostic loci in ST36 and related strains. Multiple genome reference-sequence alignment based phylogenies using three high quality genomes including ST36 isolate 10290 (NZ_AVOH01000000), BB22OP,

and RIMD 2210633 as references were reconstructed using REALPHY v1.09 with a representative sub-set of sequenced isolates where the merged alignment represents 75% coverage of sites of the largest reference genome (10290). The distribution of each of three potentially diagnostic genes is represented by (+) for gene present, and (-) for gene absent. The distribution of these loci in all available draft genomes is indicated in Figure 4. Note that in Figure 4, isolate VP-2007-007 was identified as ST 306 using the SRST2 program [Inouye M, et al., *(2012).BMC Genomics 13:338.*].

Figure 3 shows a photographic image of an agarose gel demonstrating results of a multiplex PCR assay for identification of ST36 *V. parahaemolyticus*. The presence of virulence or strain-associated amplicons including *tdh*, *trh*, *prp* (using ST36*prp*-3 primers) and *cps*, and the species specific marker *tlh* on a subset of ST36 clade members, and isolates identified from adjacent, related clades using published and newly designed primers (Table 2) were visualized on a 1.2% SeaKem LE agarose (Lonza, Rockland, ME USA) gel, with 1x Gelred (Phenix Research Products, Candler, NC USA) in TAE buffer, compared with 1 Kb Plus DNA Ladder (Invitrogen, Grand Island, NY USA).

Figure 4 provides a listing of diagnostic loci in all available draft genomes of *V. parahaemolyticus*. Note "a": For all high quality draft genomes which had no sequence type identified in www.pubMLST.org, the sequence type was identified using the SRST2 program (Inouye M, et al., *(2012).BMC Genomics 13:338.*); Unk: sequence type is not known due to new sequence type or incomplete sequences at the 7 loci, n/a: Information was unavailable. Note "b": Location of reported infection or isolation by US state; Note "c": source identified as clinical (C) or environmental (E); Note "d": year of isolation; Note "e" only partial coding sequence for *cps* identified from this genome.

Figure 5 provides a photographic image of a gel showing results from three PCR reactions, identified in the image as: a PCR reaction I, PCR reaction II, and PCR reaction III. The reactions were used identify pathogen sequence types using PCR. Results shown in the left lane of the PCR Reaction I gel lanes, identify ST36(I) products obtained using multiplex PCR using the *prp*3 and *cps* primers, with the expected 5 amplicons. The right lane of PCR Reaction I gel lanes shows results that provide secondary confirmation obtained using multiplex prp2 and flp primers. Results shown in PCR reaction II gel lanes identify ST631 (left) and ST1127 (right) obtained using multiplex, including the *tlh* control PCR (tlhF2/R).

Results shown in the PCR reaction III gel lanes identify ST674 (left lane) and ST34 (right lane) (including a secondary confirmation of ST631 with a less specific amplicon ENV). See Table 2 for primer sequence information.

Figure 6 provides a photographic image of a gel showing results from two PCR reactions, labeled in the image as PCR reaction I and PCR reaction II. In the experiments, PCR was used in the identification of *V. parahaemolyticus* sequences. The results for PCR reaction I show: bands at: tlh (tlhF2/tlhR primers), NE Phage (ST36NEOrf10F/ST36NEHypR primers), any f237-like phage (ST36PhageF2/R2 primers). The results for PCR reaction II show: tlh (tlhF2/tlhR primers), Vipa 26 and Vipa 36 size difference (ST36PhHypDF3/ST36PhOrf9R1 primers). See Table 2 for sequence information for primers.

Figure 7 is a photographic image of a gel showing results of amplifications demonstrating a shared pathogenicity island architecture among tdh/trh positive pathogens of several sequence types include in the eight lanes from left to right: ST36, ST631, ST110, ST674, ST34, ST324, ST674 and ST749 and is not present in the non-pathogenic strains and non-regional pathogenic sequence types that lack the island shown in the four right-hand lanes of the gel.

Figure 8 provides a table of ST loci in the left column, additional identifying information in the center column and associated GenBank® Accession numbers in the right column.

Detailed Description

Definitive markers of virulence have now been identified for *V. parahaemolyticus*, and assays have been developed that allow identification pathogenic *V. parahaemolyticus* known to cause infections in subjects and to contaminate substrates. In certain aspects of the invention, assays are provided that can be used to identify the status of infection of a subject with, or contamination of a substrate with *V. parahaemolyticus* bacteria. In some aspects of the invention, methods include determining one or more of the presence and level of a pathogenic *V. parahaemolyticus* bacterium in a sample. A sample assayed using methods and compositions of the invention may be obtained from a variety of different sources such as, but not limited to: cells, tissues, subjects, and substrates. In certain aspects of the invention, the identification of status of a pathogenic *V. parahaemolyticus* polynucleotide in a sample may be extrapolated to identify the presence of the pathogenic *V. parahaemolyticus* bacterium in the cell, tissue, or subject, or substrate from which the sample was obtained.

The presence of pathogenic *V. parahaemolyticus* bacterium may indicate infection or contamination of the source of the sample such as the cell, tissue, subject or the substrate by the pathogenic *V. parahaemolyticus* bacterium. Thus, some embodiments of the invention include methods of determining whether a cell, tissue or subject has, or is at risk of having, a pathogenic *V. parahaemolyticus* infection and certain embodiments of the invention include methods of determining whether a substrate is contaminated by, or is a risk of contamination by a pathogenic *V. parahaemolyticus* bacterium.

The invention in some aspects also includes methods to monitor treatment of a pathogenic *V. parahaemolyticus* infection in a cell, tissue, or subject and to monitor decontamination of a substrate contaminated by a pathogenic *V. parahaemolyticus* bacterium. Thus, some aspects of the invention include methods such as assays to identify changes in a level of a pathogenic *V. parahaemolyticus* bacterium in a sample. In instances in which a sample is obtained from a cell, tissue, or subject and/or from a substrate, methods of the invention may be used to monitor treatment of infection or contamination of the cell, tissue, or subject or a substrate, respectively, by pathogenic *V. parahaemolyticus* bacterium. The invention also provides in some aspects, methods useful to characterize one or more compounds to determine whether or not they may be useful to treat a pathogenic *V. parahaemolyticus* infection or to reduce or eliminate contamination by a pathogenic *V. parahaemolyticus* bacterium. Certain methods of the invention may be used to determine efficacy, cell toxicity, and other characteristics of compounds that may be used to treat a pathogenic *V. parahaemolyticus* infection or to reduce or eliminate pathogenic *V. parahaemolyticus* contamination of a substrate.

Pathogenic *V. parahaemolyticus* bacteria are rod-shaped, gram-negative, motile bacteria that are found in brackish water and saltwater. They can be found in organisms such as seafood, shellfish that when eaten by a subject, can cause gastrointestinal illness. Entry of the bacteria into a subject may also occur through contact with water that contains the pathogenic *V. parahaemolyticus* bacteria and thus can cause infections via contact with open wounds and handling of contaminated substrates such as shellfish, materials, liquids, etc. Outbreaks of *V. parahaemolyticus* bacterial infections generally occur along coastal regions during times when the water temperature supports high bacterial levels. Infection in a subject may be indicated by severe diarrhea, nausea, vomiting, abdominal pain, and fever and can last from 2 to 10 days. Subjects or substrates that have infection by or contamination with pathogenic *V. parahaemolyticus* bacteria may spread the disease, via contact by another subject, for example, transmission from body fluids, diarrhea, or by contact with wounds,

food, beverages, etc. Although most infections are not life-threatening, deaths do result from pathogenic *V. parahaemolyticus* infection.

Pathogenic *V. parahaemolyticus* bacteria are a cause of a growing number of cases of food-borne illnesses. Methods and tools of the invention that can be used to determine (also referred to herein as "to identify") the status of a pathogenic *V. parahaemolyticus* infection or contamination may, in some embodiments include determining the presence, absence, and/or level or quantity of the pathogenic *V. parahaemolyticus* bacteria in or on a cell, tissue, subject, and/or substrate have now been identified. Such methods of the invention may be used in identify disease or risk of disease or contamination or risk of contamination by pathogenic *V. parahaemolyticus*. Disease or contamination by a pathogenic *V. parahaemolyticus* bacteria are conditions characterized by abnormal (e.g., increased) levels of the pathogenic *V. parahaemolyticus* bacteria. With respect to a pathogenic *V. parahaemolyticus* level in a sample, as used herein a "normal" level would be no detectable pathogenic *V. parahaemolyticus* bacteria as determined using a method of the invention, and an "abnormal" level would be a detectable level of pathogenic *V. parahaemolyticus* bacteria determined using a method of the invention. As used herein, with respect to the level of a pathogenic *V. parahaemolyticus* bacteria, the terms: "increased", "elevated", and "higher" are used interchangeably, and the terms "decrease", "reduced", and "lower" are used interchangeably.

Methods and assays set forth in certain embodiments of the invention can be used to identify the status of, and evaluate and compare levels of pathogenic *V. parahaemolyticus* bacteria in samples, or in some embodiments of the invention, in or on subjects and substrates. Methods in some aspects of the invention may include detecting in a sample the presence and/or level of one or more specific polynucleotides and/or polypeptides that have now been identified as biomarkers for pathogenic *V. parahaemolyticus* bacteria. Information obtained using methods of the invention can be used to identify the status of infection of a subject or contamination of a substrate by a pathogenic *V. parahaemolyticus* bacterium and may be used to select a treatment for a subject and to select an agent with which to reduce contamination of a substrate.

As used herein the term "status" with respect to a pathogenic *V. parahaemolyticus* in a sample, may mean presence, absence, or level of the pathogenic *V. parahaemolyticus* in the sample. In certain embodiments of the invention the status of a pathogenic *V. parahaemolyticus* in a sample is binary, and is determined to be either present or absent using a method of the invention. In certain embodiments of the invention, a status may be a level

of a pathogenic *V. parahaemolyticus* bacteria in a sample, and may be a difference in a level between two or more samples, and or between a sample and a control. In some aspects of the invention, a determination of a level greater than zero of a pathogenic *V. parahaemolyticus* polynucleotide in a sample indicates the presence of a pathogenic *V. parahaemolyticus* in the sample and in the source from which the sample was obtained. In some embodiments the source from which a sample is obtained is a subject and in certain embodiments of the invention, the source is a substrate. In certain aspects of the invention, a status of a pathogenic *V. parahaemolyticus* in a sample may be used to determine a status of a pathogenic *V. parahaemolyticus* bacterial infection in a subject from whom the sample was obtained, or a pathogenic *V. parahaemolyticus* bacterial contamination of a substrate from which the sample was obtained..

As used herein the term "status" with respect to infection or contamination may mean presence, absence, onset, end, recurrence, progression, regression, increase, decrease, or other indication of the state of the infection or contamination by the pathogenic *V. parahaemolyticus* bacteria. For example, in certain embodiments of the invention the status of infection or contamination is binary, and is determined to be either present or absent using a method of the invention. In certain embodiments of the invention, a status may be the state of progression of infection or contamination, which can be determined by testing a first sample obtained from a subject or substrate and testing a second sample obtained from the subject or substrate, respectively, wherein the second sample is obtained from the subject or substrate at a later point in time. An increase in the level of a pathogenic *V. parahaemolyticus* bacteria determined in the second sample compared to the level determined in the first sample indicates the status of progression or increase of the infection or contamination of the subject or substrate by the pathogenic *V. parahaemolyticus* bacteria. A determination of a level of zero of a pathogenic *V. parahaemolyticus* bacteria in a first sample obtained from a subject or substrate followed by a determination of a level greater than zero of the pathogenic *V. parahaemolyticus* bacteria in a sample subsequently obtained from the subject or substrate, respectively, may indicate a status of onset of the infection or contamination of the subject or substrate, respectively, by the pathogenic *V. parahaemolyticus* bacteria. Similarly, a decrease in the level of a pathogenic *V. parahaemolyticus* bacteria determined in a sample obtained from a subject or substrate at a later time point compared to that determined in a sample obtained from the subject or substrate, respectively, at an earlier time point would indicate a status of regression or decrease of the infection or contamination with the pathogenic *V. parahaemolyticus* bacteria.

A determination of a level of a pathogenic *V. parahaemolyticus* bacteria that is greater than zero in a first sample obtained from a subject or substrate followed by determination of a zero level of the pathogenic *V. parahaemolyticus* bacteria in a sample subsequently obtained from the subject or substrate, respectively, indicate a status of the infection or contamination of the subject or substrate, respectively, by the pathogenic *V. parahaemolyticus* bacteria as ended or resolved.

Some embodiments of the invention provide methods that may be diagnostic and/or prognostic for infection with or contamination by pathogenic *V. parahaemolyticus* bacteria. Methods of the invention, in some aspects, utilize detection of the identified nucleic acid and/or polypeptide biomarkers to determine the presence and/or level of the bacteria in biological samples obtained from subjects, culture, and/or substrates. It has now been identified that increased presence of signature pathogenic *V. parahaemolyticus* molecules in or on a cell, tissue, subject, or substrate is correlated with infection or contamination of the cell, tissue, subject, or substrate with the pathogenic *V. parahaemolyticus* bacteria.

As used herein, the term "*V. parahaemolyticus* molecule" means a polynucleotide or polypeptide naturally expressed by a *V. parahaemolyticus* bacterium and the term "pathogenic *V. parahaemolyticus* molecule" means a polynucleotide or polypeptide naturally expressed by a pathogenic *V. parahaemolyticus* bacterium. As used herein, a "signature" pathogenic *V. parahaemolyticus* molecule refers to a polynucleotide or polypeptide that is naturally expressed by a pathogenic *V. parahaemolyticus* bacterium and is not naturally expressed by a non-pathogenic *V. parahaemolyticus* bacterium.

In an infection of a cell, tissue, or subject with, or contamination of a substrate by, pathogenic *V. parahaemolyticus* bacteria, the level of the pathogenic bacteria is statistically significantly higher in that cell, tissue, subject, or substrate as compared to the level of the pathogenic bacteria in cells, tissues, subjects, or substrates that do not have the infection or contamination. It has now been identified that the presence and/or level of a signature pathogenic *V. parahaemolyticus* polynucleotide or polypeptide of the invention in a sample obtained from a subject or substrate correlates with the presence and/or level of the pathogenic *V. parahaemolyticus* bacteria in the subject or substrate, respectively. In an infection of a cell, tissue, or subject with a pathogenic *V. parahaemolyticus* bacteria, and in contamination of a substrate with a pathogenic *V. parahaemolyticus* bacteria, the level of a signature pathogenic *V. parahaemolyticus* polynucleotide or polypeptide of the invention is statistically significantly higher in or on the cell, tissue, subject, or substrate as compared to the level of the signature pathogenic *V. parahaemolyticus* polynucleotide or polypeptide in a

cell, tissue, subject, or substrate not infected with or contaminated by the pathogenic *V. parahaemolyticus* bacteria. Thus, detecting a level of a signature pathogenic *V. parahaemolyticus* polynucleotide or polypeptide of the invention in a sample obtained from a subject or substrate permits identification of the status of the infection of the subject with the pathogenic *V. parahaemolyticus* bacteria, and permits identification of the status of the contamination of the substrate with the pathogenic *V. parahaemolyticus* bacteria, respectively.

In some embodiments of the invention detection of an amount greater than zero and/or the presence of a signature pathogenic *V. parahaemolyticus* polynucleotide or polypeptide in a sample identifies the status of the source (for example, the subject or substrate) of the sample as infected and/or contaminated with the pathogenic *V. parahaemolyticus* bacteria. In some embodiments of the invention, the level of a signature pathogenic *V. parahaemolyticus* polynucleotide or polypeptide in a control sample is zero, which means the absence of the pathogenic *V. parahaemolyticus* bacterium. In certain embodiments of the invention, any detected level of a signature pathogenic *V. parahaemolyticus* polynucleotide or polypeptide of the invention that is greater than zero in a sample need not be compared to a control because a level greater than zero confirms the presence of the pathogenic *V. parahaemolyticus* bacterium in the sample.

Methods of the invention, in some embodiments, may include detecting a level of a pathogenic *V. parahaemolyticus* polynucleotide (and hence bacteria) in a sample and comparing the detected level to one or more of: a level of zero, a control level, and a prior level in a biological sample from the same subject or substrate. For example in some aspects of the invention, a method may include detecting a level of a pathogenic *V. parahaemolyticus* bacterium in a sample and the sample was obtained from a subject that has, is suspected of having, is susceptible to having, or is at risk of having an infection by the pathogenic *V. parahaemolyticus* bacterium.

As used herein, the term "level" when used in reference to detecting a level of a *V. parahaemolyticus* polynucleotide, polypeptide, or bacterium means: (a) a quantitative value such as an amount, weight, or concentration of the polynucleotide, polypeptide, or bacterium; or (b) the "presence" or "absence" of the polynucleotide, polypeptide or bacterium. It will be understood that in certain embodiments of the invention, the presence or absence of a *V. parahaemolyticus* polynucleotide, polypeptide, or bacterium is detected as a binary determination, without quantification. Thus, detecting may include one or more of determining the presence or absence of a pathogenic *V. parahaemolyticus* bacterium

polynucleotide or polypeptide molecule; and determining a quantity of a pathogenic *V. parahaemolyticus* bacterium polynucleotide or polypeptide molecule. Means for detecting levels of polypeptides and polynucleotides and for determining amino acid and nucleic acid sequences are known in the art and non-limiting examples of detection means are provided herein.

In certain aspects of the invention, a level of a pathogenic *V. parahaemolyticus* bacterium in/on a subject or substrate can be detected using methods of the invention to measure the pathogenic *V. parahaemolyticus* bacterium polynucleotide and/or polypeptide molecule (for example, one or more biomarker polynucleotides or polypeptides) in an *in vitro* assay of a biological sample obtained from the subject or substrate, respectively. As used herein, the term "measure" used in reference to measuring a level of a *V. parahaemolyticus* polynucleotide comprising a nucleic acid sequence means: (a) determining a quantitative value, such as an amount, weight, concentration, relative amount, relative concentration, etc., of a polynucleotide comprising the nucleic acid sequence; or (b) determining the "presence" or "absence" of a polynucleotide comprising the nucleic acid sequence. It will be understood that in certain embodiments of the invention, the presence or absence of a *V. parahaemolyticus* polynucleotide is measured as a binary determination, without quantitation. Thus, measuring may include one or more of determining the presence or absence of a pathogenic *V. parahaemolyticus* bacterium polynucleotide or polypeptide molecule; and determining a quantity of a pathogenic *V. parahaemolyticus* bacterium polynucleotide or polypeptide molecule. Means for measuring levels of polypeptides and polynucleotides and for determining amino acid and nucleic acid sequences are known in the art. Non-limiting examples of measuring means are provided herein.

Detection methods suitable for use in certain embodiments of methods of the present invention can be used to detect and measure levels of pathogenic *V. parahaemolyticus* bacterium polypeptide or nucleic acid molecules in samples (also interchangeably herein as biological samples) *in vitro* as well as *in vivo*. For example, in certain embodiments of the invention *in vitro* techniques for detecting and/or measuring levels of pathogenic *V. parahaemolyticus* polynucleotides (genes) include but are not limited to: polymerase chain reaction methods, reverse transcriptase quantitative polymerase chain reaction (RT-qPCR), Northern hybridizations, *in situ* hybridizations, DNA or oligonucleotide array, sequencing, etc. In certain aspects of the invention *in vitro* techniques for detecting and/or measuring levels of pathogenic *V. parahaemolyticus* bacterium DNA include, but are not limited to: polymerase chain reaction (PCR) and Southern hybridizations. *In vitro* techniques for

detecting and/or measuring levels of pathogenic *V. parahaemolyticus* bacterium polypeptides in certain embodiments of the invention include, but are not limited to enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, immunofluorescence, and other known suitable techniques.

In some embodiments of the invention, methods may include comparing a level of a pathogenic *V. parahaemolyticus* polynucleotide or polypeptide molecule in a sample to a control value or level for the pathogenic *V. parahaemolyticus* polynucleotide or polypeptide molecule. As used herein a "control" may be a normal control or an infection or contamination control. Selection and use of appropriate controls in comparative, diagnostic, and assay methods are well known in the art. In some embodiments of the invention, a normal control level may zero and represent the amount in a subject or substrate free of the pathogenic *V. parahaemolyticus* bacterium. A control level of a pathogenic *V. parahaemolyticus* molecule can readily be determined by measuring a level of the pathogenic *V. parahaemolyticus* polynucleotide or polypeptide molecule using a method of the invention, as described herein. In some embodiments of the invention, an infection or contamination control level may be obtained from a sample from a subject or substrate known have infection with or contamination by, respectively, the pathogenic *V. parahaemolyticus* bacterium. In some embodiments, a disease control pathogenic *V. parahaemolyticus* bacterium level may be based on levels obtained from one or more subjects known to have the infection, or one or more surfaces known to have contamination with the pathogenic *V. parahaemolyticus* bacterium. In certain embodiments of the invention, the disease control may be a sample from a subject diagnosed with a pathogenic *V. parahaemolyticus* infection and the subject's disease control may be compared to another sample obtained from the subject at a different time. In certain embodiments of the invention, a disease control level of a pathogenic *V. parahaemolyticus* molecule can readily be determined by measuring levels of the pathogenic *V. parahaemolyticus* molecule in a sample obtained from a subject known to have a pathogenic *V. parahaemolyticus* infection, or obtained from a substrate known to have a pathogenic *V. parahaemolyticus* contamination.

In some embodiments of the invention, a control level of pathogenic *V. parahaemolyticus* is a level determined from samples, subjects, substrates, tissues, cells, etc. that do not have contamination by or infection by the pathogenic *V. parahaemolyticus* bacteria that is being tested for in the sample. For example, in some embodiments, a control level of pathogenic *V. parahaemolyticus* molecule is a level determined in a normal sample that does not have contamination by the pathogenic *V. parahaemolyticus* bacteria that is

suspected to be in the sample obtained from the subject or substrate. In such a case, the presence of infection and/or contamination by the pathogenic *V. parahaemolyticus* can be determined based on an increase in the level of the pathogenic *V. parahaemolyticus* polynucleotide or polypeptide in the subject's or substrate's sample as compared to the control that is free of the pathogenic *V. parahaemolyticus*. In another embodiment of the invention, the control is from a normal subject or substrate and the test sample is from a subject or substrate that is suspected of having infection with or contamination by a pathogenic *V. parahaemolyticus* bacteria.

In some aspects of the invention methods are provided that include comparing a level of a pathogenic *V. parahaemolyticus* molecule measured a biological sample obtained from a subject, or substrate to a control value for determining infection or contamination status, stage, prognosis, etc. Changes over time in the presence and/or level of a pathogenic *V. parahaemolyticus* molecule in a subject or substrate can be assessed by determining levels in two or more samples obtained from a subject or substrate at different times. Values obtained from a sample obtained at one time can be compared to values obtained from samples obtained at other times. For example, a first level obtained from a subject may serve as a baseline level or control level for that subject, thus allowing comparison of the pathogenic *V. parahaemolyticus* molecule level and the determination of change or stability of the pathogenic *V. parahaemolyticus* infection over time. Levels of a pathogenic *V. parahaemolyticus* molecule may also be measured after a specific course of treatment of the pathogenic *V. parahaemolyticus* infection or contamination has been initiated, with the intent of determining the efficacy of that treatment of the subject or the decontamination of the substrate, or just the natural change of the level of the pathogenic *V. parahaemolyticus* bacteria in or on a subject or substrate over time. For example, though not intended to be limiting, samples may be obtained from a substrate such as a body of water, at different times of the year and the levels of a signature pathogenic *V. parahaemolyticus* polynucleotide or polypeptide determined to assess changes in the contamination of the substrate over time. Thus, the status of infection with or contamination by a pathogenic *V. parahaemolyticus* bacterium in a subject or substrate can be monitored over time using aspects of methods of the invention to assess change in levels of one or more pathogenic *V. parahaemolyticus* molecules.

In some aspects of the invention, assessing a change in at least one of the presence, absence, and level of a pathogenic *V. parahaemolyticus* molecule in a subject or substrate may be desirable, and methods of the invention may be used to monitor the level of the

pathogenic *V. parahaemolyticus* bacteria in biological samples obtained from the subject or substrate over time to assess changes. Some aspects of the invention include methods to monitor the level of a pathogenic *V. parahaemolyticus* molecule over time to assess changes in the level. A decrease in a level of the pathogenic *V. parahaemolyticus* bacteria over time may be indicated by a level of the pathogenic *V. parahaemolyticus* molecule that, in a second biological sample is less than 100%, 99%, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 50%, 45%, 40%, 35%, 30%, 25%, 20%, 15%, 10%, 5%, or 1% of the level identified in a first biological sample obtained from the subject or substrate at a prior time point than the time at which the second sample was obtained. An increase in a level of a pathogenic *V. parahaemolyticus* bacterium over time may be indicated by a level of the pathogenic *V. parahaemolyticus* molecule that, in second biological sample is more than 100%, 101%, 105%, 110%, 115%, 120%, 125%, 130%, 140%, 150%, 160%, 170%, 180%, 190%, 200%, 500%, 1000%, 10,000% of the level identified in a first biological sample obtained from the subject or substrate at a prior time point, including all percentages in between those stated. A higher level of increase, for example 200%, 400%, 600%, 1000% or higher increase may indicate the severity of the infection or contamination with a higher percent increase corresponding to a more severe infection or contamination. An increase in a level of a pathogenic *V. parahaemolyticus* molecule between a first and second sample may also indicate the progression of a pathogenic *V. parahaemolyticus* infection or contamination. In embodiments where the level of a pathogenic *V. parahaemolyticus* molecule was zero in a first biological sample, any level over zero in the second biological sample indicates an increase in a level of the pathogenic *V. parahaemolyticus* bacteria and in some embodiments may indicate the onset of infection or contamination in a subject or substrate, respectively. It will be understood that if zero pathogenic *V. parahaemolyticus* molecules are detected in a sample, any level of detection in a subsequent sample that is greater than zero indicates an increase. Similarly, if any level of a pathogenic *V. parahaemolyticus* molecule is detected in an initial sample and no pathogenic *V. parahaemolyticus* molecules are detected in a subsequent sample, it indicates a decrease.

Samples (also referred to herein as biological samples) may be obtained from a subject or substrate at any desired interval such as at least once per day, every other day, once per week, every other week, once per month, every other month, every third month, every year, etc. It will be understood that interval testing using methods of the invention may be performed to assess the level of a pathogenic *V. parahaemolyticus* bacteria to determine information including, but not limited to: the likelihood of infection from a substrate,

reinfection of a substrate, reinfection of a subject, decreased contamination of a substrate, carrier status of a subject, etc.

Assessment of efficacy of candidate agents to treat subjects or de-contaminate substrates that have been contacted with a pathogenic *V. parahaemolyticus* bacterium may also be done according to some embodiments of the invention, such as, in a non-limiting example, screening assays to assess candidate agents or compounds to reduce or eliminate pathogenic *V. parahaemolyticus* bacteria from a subject or substrate. It will be understood that a therapeutic regimen or decontamination regime may be either prophylactic or a treatment of an existing pathogenic *V. parahaemolyticus* infection or contamination of a subject or substrate, respectively. The invention in some aspects provides methods that may be used to monitor a subject's or substrate's response to prophylactic treatment provided to a subject or substrate. Some embodiments of the invention may include methods to assess treatments intended to reduce, prevent, or eliminate pathogenic *V. parahaemolyticus* bacteria and/or to provide an indicator of a status of a pathogenic *V. parahaemolyticus* infection or contamination, and in some embodiments of the invention, can be used to select one or more therapies for the subject, or one or more decontamination treatments for the substrate, for example, to select a drug therapy, disinfection protocol, etc. In some aspects, methods of the invention may include one or more of selecting a treatment for a subject or substrate and treating a subject or substrate for a pathogenic *V. parahaemolyticus* infection or contamination based at least in part on the results of an assay of the invention.

Assays for assessing pathogenic *V. parahaemolyticus* bacteria levels in certain embodiments of the invention may include determining one or more pathogenic *V. parahaemolyticus* molecule levels, including but not limited to determining levels of signature polynucleotides that encode pathogenic *V. parahaemolyticus* polypeptides and/ determining levels of pathogenic *V. parahaemolyticus* polypeptides in cells, tissues, subjects, and substrates. Levels of pathogenic *V. parahaemolyticus* polynucleotides and polypeptides can be determined in a number of ways when carrying out various embodiments of methods of the invention. In some embodiments of the invention, a level of a pathogenic *V. parahaemolyticus* polynucleotide acid or polypeptide is measured in relation to a control level of the polynucleotide or polypeptide, respectively, in a biological sample from at least one of a cell, tissue, subject, and substrate. One possible measurement of the level of pathogenic *V. parahaemolyticus* polynucleotide or polypeptide is a measurement of an absolute level of the polynucleotide or polypeptide. This could be expressed, for example, in the level of the polynucleotide or polypeptide per unit of cells or tissue. Another

measurement of a level of a pathogenic *V. parahaemolyticus* polynucleotide or polypeptide is a measurement of a change in the level of the polynucleotide or polypeptide over time. This may be expressed in an absolute amount or may be expressed in terms of a percentage increase or decrease over time. Methods disclosed herein may be used in embodiments of the invention to assess pathogenic *V. parahaemolyticus* polypeptide and polynucleotide molecules to assess the status of infection in a subject and contamination of a substrate and/or the efficacy of treatments for such infection and/or contamination.

As used herein a "subject" refers to an animal, such as, but not limited to a human, a non-human primate, a rodent, a dog, cat, a fish, a bird, or other animal. . Thus, in addition to human medical application, some aspects of the invention include veterinary application of methods described herein. In certain aspects of the invention a subject is an invertebrate, including but not limited to a shellfish, including but not limited to a clam, oyster, or mussel. A subject may be suspected of being infected with a pathogenic *V. parahaemolyticus* bacterium and methods of the invention can be used to measure the pathogenic *V. parahaemolyticus* molecules in a biological sample from the subject, thus the subject may be diagnosed with the infection by the pathogenic *V. parahaemolyticus* bacteria. In some embodiments of the invention, a subject may not have been previously or be currently diagnosed with infection by a pathogenic *V. parahaemolyticus* bacteria, but may be considered at risk of having the infection, for example, a subject who may be free of symptoms, but who may have been exposed to, or be considered to be at risk of having been exposed to a pathogenic *V. parahaemolyticus* bacterium.

Similarly, a substrate may be suspected of having contamination with a pathogenic *V. parahaemolyticus* bacterium and methods of the invention can be used to measure the pathogenic *V. parahaemolyticus* in a biological sample from the substrate, thus the substrate may be determined to be contaminated with the pathogenic *V. parahaemolyticus* bacteria. In some embodiments, a substrate may not have been previously, or be currently identified as being contaminated with the pathogenic *V. parahaemolyticus*, but may be considered to be at risk of contamination by a pathogenic *V. parahaemolyticus* bacterium. In some embodiments of the invention, a substrate may have previously identified as having contamination by a pathogenic *V. parahaemolyticus* bacteria, but conditions such as temperature or environmental status may have changed and methods of the invention may be used to identify an increase, or decrease in contamination of the substrate.

In some embodiments of the invention, a sample comprises a cell or tissue or extracellular material from a subject or substrate. In some embodiments a sample comprises

one or more of: a skin sample, fluid sample, tissue sample, stool sample, mucus sample, sweat sample, tear sample, urine sample, pus sample, tissue scraping, gastric sample, emesis sample, inflammatory exudate sample, blood sample, or lymph sample. It will be understood that additional types of samples may also be used in methods of the invention. In certain embodiments of the invention, a sample is a cultured sample.

In certain embodiments of the invention a sample is a sample obtained from a substrate. In some embodiments of the invention, a substrate comprises a liquid. A liquid component of a sample may, in some embodiments comprise water. In some embodiments of the invention a substrate is at least a portion of a tool, work surface, a medical device, body of water, clothing, skin, tissue, an edible substance (e.g., a beverage or food), etc. In certain embodiments, a substrate may comprise a solid, such as, but not limited to: metal, wood, dirt, rock, plastic, glass, cork, fiber, a polymer, stone, leather, plant matter, fiberglass, nylon, or a fabric. In some embodiments of the invention a substrate may comprise a food or beverage substance. Non-limiting examples of food substances that may be assessed using methods of the invention include shellfish, fish, water organisms, plants, dairy products, etc. In certain embodiments, a sample may be obtained from one or more of a plurality of substrates, for example from one or more individual shellfish that make up part of a plurality of shellfish, and results of an detection method of the invention may be used to identify whether or not the plurality of substrates is contaminated with a pathogenic *V. parahaemolyticus* bacteria.

As used herein, the term "isolated", when used in the context of a sample, is intended to indicate that the sample has been removed from a subject, culture, or substrate. In some embodiments of the invention, a sample is isolated from a subject, culture, or substrate and is subjected to a method of the present invention without further processing or manipulation subsequent to its isolation. In some embodiments of the invention, a sample can be processed or manipulated subsequent to being isolated and prior to being subjected to a method of the invention. For example, a sample can be refrigerated (e.g., stored at 4°C), frozen (e.g., stored at -20°C, stored at -135°C, frozen in liquid nitrogen, or cryopreserved using any one of many standard cryopreservation techniques known in the art). Furthermore, a sample may be purified subsequent to isolation from a subject, culture, or a substrate and prior to subjecting it to a method of the present invention.

As used herein, the term "purified" when used in the context of a sample, is intended to indicate that at least one component of the isolated biological sample has been removed from the sample such that fewer components, and consequently, purer components, remain following purification. For example, a serum sample can be separated into one or more

components using centrifugation techniques known in the art to obtain partially-purified sample preparation. Furthermore, it is possible to purify a biological sample such that substantially only one component remains. For example, a sample can be purified such that substantially only the polypeptide or polynucleotide component of the sample remains. Thus it will be understood that a sample may originally comprise one or more of a pathogenic *V. parahaemolyticus* bacterium may be purified and processed prior to assay and thus may no longer contain the *V. parahaemolyticus* bacteria, but containing components of the pathogenic *V. parahaemolyticus* bacteria such as one or more pathogenic *V. parahaemolyticus* bacteria polynucleotides (also referred to herein as pathogenic *V. parahaemolyticus* polynucleotides). Therefore, identifying a status of a pathogenic *V. parahaemolyticus* polynucleotide in a sample that has been isolated or processed identifies that status in the original sample.

Furthermore, it may be desirable to amplify a component of a sample such that detection of the component is facilitated. For example, the mRNA component of a biological sample can be amplified (e.g., by RT-PCR) such that detection of *V. parahaemolyticus* mRNA is facilitated. As used herein, the term "RT-PCR" (an abbreviation for reverse transcriptase-polymerase chain reaction) involves subjecting mRNA to the reverse transcriptase enzyme, resulting in the production of a cDNA which is complementary to the base sequences of the mRNA. Large amounts of selected cDNA can then be produced by means of the polymerase chain reaction which relies on the action of heat-stable DNA polymerase for its amplification action. Alternative amplification methods include: self-sustained sequence replication (Guatelli, J. C. et al., 1990, Proc. Natl. Acad. Sci. USA 87: 1874-1878), transcriptional amplification system (Kwoh, D. Y. et al., 1989, Proc. Natl. Acad. Sci. USA 86: 1173-1177), Q-Beta Replicase (Lizardi, P. M. et all, 1988, Bio/Technology 6: 1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

V. parahaemolyticus Detection Techniques

V. parahaemolyticus molecules (for example, *V. parahaemolyticus* polypeptides and polynucleotides that encode *V. parahaemolyticus* polypeptides), can be detected and measured using any suitable means known in the art. In certain aspects of the invention a *V. parahaemolyticus* polypeptide and/or polynucleotide that encodes a *V. parahaemolyticus*

polypeptide are molecules that are present in pathogenic *V. parahaemolyticus* bacteria, and are not present in non-pathogenic bacteria. Thus, in certain embodiments the polypeptide molecules and nucleic acid molecules that encode the polypeptides are signature molecules. In some embodiments of the invention, a detection or measurement means for *V. parahaemolyticus* molecules includes an immunological assay, nucleotide determination (mRNA or DNA), etc. Examples of immunological assays suitable for use in methods of the invention may include, but are not limited to ELISA assays, assays that utilize an anti-*V. parahaemolyticus* polypeptide antibody (or FV derivative) to which is conjugated a detectable label. A detectable label may be conjugated (also referred to herein as "linked") either directly or indirectly. In some aspects of the invention, pathogenic *V. parahaemolyticus* presence and/or levels may be measured by a PCR reaction, in situ hybridization, DNA microarray hybridization, or other suitable method to identify the presence of a sequence present in a pathogenic *V. parahaemolyticus* bacterium. Methods of measuring levels of polynucleotides encoding pathogenic *V. parahaemolyticus* bacteria (i.e. pathogenic *V. parahaemolyticus* bacteria mRNA) may include, but are not limited to, real-time polymerase chain reaction (qRT-PCR), DNA array, and next generation sequencing methods. Application and optimization of such methods are known and practiced in the art.

The present invention features agents that are capable of detecting and/or quantitating a pathogenic *V. parahaemolyticus* bacteria-encoding polynucleotide such that the presence and/or level of the pathogenic *V. parahaemolyticus* bacteria are determined. As defined herein, an "agent" refers to a substance that is capable of identifying or detecting a pathogenic *V. parahaemolyticus* bacteria in a sample (e.g., identifies or detects pathogenic *V. parahaemolyticus* bacteria mRNA, DNA, polypeptide, activity, etc.). As used herein, the terms "labeled" or "labelable" refers to the attaching or including of a label (e.g., a marker or indicator) or ability to attach or include a label (e.g., a marker or indicator). A detectable label may be conjugated to (also referred to herein as "linked to") a polynucleotide either directly or indirectly. For example, though not intended to be limiting: a direct link may be a be when a detectable label and a polynucleotide are physically attached to each other, and an indirect link may be when a detectable label is physically attached to an entity and the entity and not the detectable label is physically attached to a polynucleotide. Detectable labels, markers, or indicators useful in methods of the invention may include, but are not limited to, for example, radioactive molecules, colorimetric molecules, and enzymatic molecules that produce detectable changes in a substrate. A detectable label for use in methods, compositions, and kits of the invention may be a radioactive molecule, a luminescent

molecule, a fluorescent molecule, a chemiluminescent molecule, biotin, an enzyme, a His tag, or an exogenous nucleic acid sequence, or other suitable detectable label known in the art. In some aspects of the invention, a detectable label comprises: Hexachloro-Fluorescein (HEX), VIC fluorescent dye, 4-5-Dichloro carboxy fluorescein (JOE), Cy3 fluorescent dye, or TexasRed (TxRed).

Methods of the invention for detecting the presence and/or quantity of a pathogenic *V. parahaemolyticus* bacterium may also include procedures such as an immunological assay, a polymerase chain reaction, real-time polymerase chain reaction (qRT-PCR), mass spectroscopy, etc. In addition, embodiments of the invention include may include nucleic "aptamers", i.e. nucleic acids [DNA, RNA or peptide nucleic acids (PNAs)] that possess high affinity for polypeptides now identified as present in pathogenic *V. parahaemolyticus* bacteria (which may not be present in non-pathogenic *V. parahaemolyticus* bacteria) and can be readily labeled for high throughput binding assays. Aptamers can be produced by standard molecular biological techniques by those skilled in the art by repeated rounds of binding, selection, and affinity, and amplification [Hamaguchi, et al. *Anal. Biochem.* (2001) 294; pt 2, pages 126-131].

In some embodiments of the invention an agent is a labeled or labelable polynucleotide probe capable of hybridizing to a pathogenic *V. parahaemolyticus* polynucleotide, (e.g., a pathogenic *V. parahaemolyticus* RNA or DNA). For example, the agent can be an oligonucleotide primer for the polymerase chain reaction that flanks or lies within the nucleotide sequence encoding pathogenic *V. parahaemolyticus* polypeptides that may be present in pathogenic *V. parahaemolyticus* and may not be present in non-pathogenic *V. parahaemolyticus* bacteria. In some embodiments of the invention, the biological sample being tested is an isolate, for example, RNA. In yet another embodiment, the isolate (e.g., the RNA) is subjected to an amplification process that results in amplification of a pathogenic *V. parahaemolyticus* polynucleotide. As defined herein, an "amplification process" is designed to strengthen, increase, or augment a molecule within the isolate. For example, where the isolate is mRNA, an amplification process such as RT-PCR can be utilized to amplify the mRNA, such that a signal is detectable or detection is enhanced. Such an amplification process is beneficial particularly when the biological, tissue, or tumor sample is of a small size or volume.

Pathogenic *V. parahaemolyticus* Nucleic Acid Binding Agents

Types of agents that can be used to determine levels of pathogenic *V. parahaemolyticus*-encoding polynucleotides may include, but are not limited to cDNA, riboprobes, synthetic oligonucleotides, and genomic probes. Probes of the invention may also be used in real-time PCR (also referred to as qPCR) methods. It will be understood that primers listed herein are also probes in the sense that they bind, hybridize, anneal to a sequence. The type of agent (e.g. probe) used will generally be dictated by the particular situation, such as riboprobes for in situ hybridization, and cDNA for Northern blotting, for example. In certain aspects of the invention, the probe is directed to nucleotide regions unique to the polypeptide. Detection of the pathogenic *V. parahaemolyticus*-encoding gene, per se, will be useful for diagnostic methods of the invention and for screening for compounds to treat infection and/or contamination by the pathogenic *V. parahaemolyticus* bacteria. Other forms of assays to detect targets more readily associated with levels of expression – transcripts and other expression products – will generally be useful as well. A probe may be as short as is required to differentially recognize pathogenic *V. parahaemolyticus* mRNA transcripts, and may be as short as, for example, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 bases, or more may be used. Non-limiting examples of polynucleotides that may be considered probes in certain instances are set forth herein as: SEQ ID NOs:2, 3, 4, 5, 6, 7, 9, 10, 12, 13, 14, 16, 17, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 39, 40, and polynucleotides identified as primers and probes in Tables 2 and 3. One or more probes may also be reverse engineered by one skilled in the art, for example using the nucleic acid sequence of a pathogenic *V. parahaemolyticus* provided herein, such as SEQ ID NO:1, SEQ ID NO:8, SEQ ID NO:38 or other provided sequence or variant thereof. The form of labeling of a probe used in an embodiment of the invention may be any that is appropriate, such as the use of radioisotopes, for example, ^{32}P and ^{35}S, etc. Labeling with radioisotopes may be achieved, whether the probe is synthesized chemically or biologically, by the use of suitably labeled bases using methods well known in the art. For use in some aspects of the invention, probes may be detectably labeled with a fluorescent (also referred to herein as a chemiluminescent or luminescent) molecule, examples of which include, but are not limited to: Hexachloro-Fluorescein (HEX), VIC fluorescent dye, 4-5-Dichloro carboxy fluorescein (JOE), Cy3 fluorescent dye, and TexasRed (TxRed). Additional types of detectable labels, known in the art and/or disclosed herein may also be used in conjunction with polynucleotide probes of the invention. In some embodiments of the invention a *V. parahaemolyticus* molecule may be

detected as part of a complex with one or more additional molecules such as polynucleotides or labeling agents.

Pathogenic *V. parahaemolyticus* RNA Detection Techniques

Detection of RNA transcripts may be achieved by Northern blotting, for example, wherein a preparation of RNA is run on a denaturing agarose gel, and transferred to a suitable support, such as activated cellulose, nitrocellulose or glass or nylon membranes. Radiolabeled cDNA or RNA is then hybridized to the preparation, washed and analyzed by autoradiography.

Detection of RNA transcripts can further be accomplished using known amplification methods. For example, it is within the scope of the present invention to reverse transcribe mRNA into cDNA followed by real-time polymerase chain reaction (RT-PCR); or, to use a single enzyme for both steps as described in U.S. Pat. No. 5,322,770, or reverse transcribe mRNA into cDNA followed by symmetric gap ligase chain reaction (RT-AGLCR). Each of these methods is well known and routinely used in the art. Other known amplification methods can also be utilized in methods of the invention. In situ hybridization visualization may also be employed, wherein a radioactively labeled antisense RNA probe is hybridized with a thin section of a biopsy sample, washed, cleaved with RNase and exposed to a sensitive emulsion for autoradiography. Biological samples may be stained with haematoxylin to demonstrate the histological composition of the sample, and dark field imaging with a suitable light filter shows the developed emulsion. Non-radioactive labels such as digoxigenin, or others referred to herein and/or known in the art may also be used.

Pathogenic *V. parahaemolyticus* Antibodies and Additional Binding Agents

In addition to using PCR and RNA detection methods for detection of a pathogenic *V. parahaemolyticus* molecule in diagnostic methods of the invention, it may also be possible to use other molecules or compounds that bind to a pathogenic *V. parahaemolyticus* molecule identified as being expressed in pathogenic but not non-pathogenic *V. parahaemolyticus* bacteria. For example, it may be possible to identify compounds, and/or molecules such as polypeptides that specifically bind to a pathogenic *V. parahaemolyticus* molecule. In addition, it may also be possible to use an antibody or other compound or molecule that binds to, and permits detection of a pathogenic *V. parahaemolyticus* molecule in a sample. In some embodiments a pathogenic *V. parahaemolyticus* molecule may be detected as part of a complex with one or more additional polypeptides.

An isolated pathogenic *V. parahaemolyticus* polypeptide, or fragment thereof, can be used as an immunogen to generate antibodies that bind pathogenic *V. parahaemolyticus* polypeptides using standard techniques for polyclonal and monoclonal antibody preparation. The full-length pathogenic *V. parahaemolyticus* polypeptide can be used or, alternatively, the invention provides antigenic peptide fragments of a pathogenic *V. parahaemolyticus* for use as immunogens. The antigenic peptide of pathogenic *V. parahaemolyticus* may comprise at least 8 amino acid residues of the amino acid sequence set forth herein and encompasses an epitope of a polypeptide from a pathogenic *V. parahaemolyticus* that is not present in a non-pathogenic *V. parahaemolyticus* polypeptide.

Pathogenic *V. parahaemolyticus* Protein Detection Techniques

In certain embodiments of the invention, methods include the use of diagnostic molecules (e.g., antibodies, antibody equivalents, binding molecules, etc.) to detect pathogenic *V. parahaemolyticus* polypeptides. Amino acid sequences of pathogenic *V. parahaemolyticus* polypeptides prp, cps, and flp are provided herein as SEQ ID NOs: 41, 42, and 46, respectively. Methods for the detection of polypeptides are well known to those skilled in the art, and include ELISA (enzyme linked immunosorbent assay), RIA (radioimmunoassay), Western blotting, and immunohistochemistry. Methods for immunoassays are routinely used and are well known in the art.

ELISA and RIA procedures may be conducted such that a pathogenic *V. parahaemolyticus* standard is labeled (with a radioisotope such as ^{125}I or ^{35}S, or an assayable enzyme, such as horseradish peroxidase or alkaline phosphatase), and, together with the unlabelled sample, brought into contact with the corresponding antibody, whereon a second antibody is used to bind the first, and radioactivity or the immobilized enzyme assayed (competitive assay). Alternatively, pathogenic *V. parahaemolyticus* in the sample is allowed to react with the corresponding immobilized antibody, radioisotope- or enzyme-labeled anti-pathogenic *V. parahaemolyticus* antibody is allowed to react with the system, and radioactivity or the enzyme assayed (ELISA-sandwich assay). Other conventional methods may also be employed as suitable. Enzymatic and radiolabeling of a detection agent (e.g., antibodies, binding molecules, etc.) may be carried out by conventional means.

Other techniques may be used to detect pathogenic *V. parahaemolyticus* molecules according to a practitioner's preference based upon the present disclosure. One such technique is Western blotting (Towbin et at., 1979 Proc. Nat. Acad. Sci. 76:4350), wherein a suitably treated sample is run on an SDS-PAGE gel before being transferred to a solid

support, such as a nitrocellulose filter. Anti-pathogenic *V. parahaemolyticus* antibodies (unlabeled) are then brought into contact with the support and assayed by a secondary immunological reagent, such as labeled protein A or anti-immunoglobulin (suitable labels including but not limited to ^{125}I, horseradish peroxidase and alkaline phosphatase). Chromatographic detection may also be used.

Immunohistochemistry may be used to detect expression of pathogenic *V. parahaemolyticus* in a biological sample. A suitable antibody is brought into contact with, for example, cells, washed, and then contacted with a second, labeled antibody. Labeling may be by fluorescent markers, enzymes, such as peroxidase, avidin, or radiolabelling. The assay may be scored visually, using microscopy, or using any other suitable methods.

Pathogenic *V. parahaemolyticus* Detection Kit

The invention also encompasses kits for detecting the presence of pathogenic *V. parahaemolyticus* in a sample (e.g., a cell sample, tissue sample, water sample, seafood sample, fish sample, wound sample, etc.). For example, a kit can comprise a labeled or labelable agent capable of detecting a pathogenic *V. parahaemolyticus* polypeptide or polynucleotide (e.g., RNA, DNA, etc.) in a sample and a means for determining a status (e.g., at least one of the presence and level/amount) of a pathogenic *V. parahaemolyticus* molecule in the sample. The agent can be packaged in a suitable container. The kit can further comprise a means for comparing the level of the pathogenic *V. parahaemolyticus* molecule in the sample with a standard or control and/or can further comprise instructions for using the kit to detect the pathogenic *V. parahaemolyticus* polynucleotide or polypeptide. A kit may also include a detectably labeled probe.

This invention in some aspects also provides a kit for measuring pathogenic *V. parahaemolyticus* molecules. Such a kit may include a diagnostic agent (e.g., an antibody or antibody fragments, or binding molecule, etc.) that selectively bind a pathogenic *V. parahaemolyticus* molecule or a set of DNA oligonucleotide primers that allows synthesis of cDNA encoding the polypeptide or a DNA probe that detects expression of a pathogenic *V. parahaemolyticus* mRNA, etc. In some embodiments of the invention, the primers and probes may comprise at least 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or more nucleotides and hybridize under stringent conditions to a DNA fragment having the nucleic acid sequence set forth herein, for example, the nucleic acid sequence of an ST36*prp* polynucleotide, a ST36*cps* polynucleotide, a *tlh* polynucleotide, a *tdh* polynucleotide, a *trh* polynucleotide, an ST36*flp* polynucleotide, an ST631*end* polynucleotide, ST631-ENV

polynucleotide, ST34*lys* polynucleotide, an ST34*reg* polynucleotide, an ST1127*hyp* polynucleotide, an ST36Phage or other polynucleotide set forth in Table 2, Table 3, or elsewhere herein, or a variant thereof. As herein used, the term "stringent conditions" means hybridization will occur only if there is at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or at least 99%, identity between the sequences. Table 1 provides additional information in reference to the prp, cps, flp, tdh, trh, and tlh sequences and Fig. 8 provides additional information in reference to loci such as the ST631, ST34, ST674, ST1127, and ST36 loci. Table 2 provides non-limiting examples of primers utilized in PCR amplifications. Table 3 provides oligonucleotide primers, which in some aspects of the invention may be used in qPCR detection. QPCR is also referred to as real-time PCR.

Table 1 Sequence Identification Information for prp, cps, flp, tdh, trh, and tlh

	Protein accession #	contig accession #	location in contig	locus tag
flp	EGF42675	AFBW01000018.1	220810..222099	VP10329_01545
tdh	EGF40527	AFBW01000025.1	7483.. 8007	VP10329_22563
trh	EGF40542	AFBW01000025.1	22125.. 22694	VP10329_22638
tlh	EGF40542	AFBW01000024.1	230297.. 231550	VP10329_04677
prp	EGF42613	AFBW01000018.1	160407..162398	VP10329_01235
cps	EGF42671	AFBW01000018.1	216766..218355	VP10329_01525

Table 2: Oligonucleotide primers used for isolate identification by PCR and examples of amplification conditions

Gene /locus	Primer sequence	SEQ ID NO	Amplicon size (bp)	Temp Anneal	Source
Tlh	F2: AGAACTTCATCTTGATGACACTGC	12	401	55	1
	F: AAAGCGGATTATGCAGAAGCACTG	13	450		2
	R: GCTACTTTCTAGCATTTTCTCTGC	14			
Tdh	F: GTAAAGGTCTCTGACTTTTGGAC	16	269	55	*1*
	R: TGGAATAGAACCTTCATCTTCACC	17			
Trh	F: CATAACAAACATATGCCCATTTCCG	19	500	55	*1*
	R: TTGGCTTCGATATTTTCAGTATCT	20			
ORF8	F-O3MM824: AGGACGCAGTTACGCTTGATG	47	369	55	1
	R-O3MM1192: CTAACGCATTGTCCCTTTGTAG	48			
ST36*prp*	F: CGGCTTGAGTTTTCGTCATT	2	609	55	2
	R: CCACACCTGCTGGTTATTTAGTTC	3			
ST36*prp*	F2: TGCGGAATCTGATCTTTATCCTC	6	1028	55	2
	R2: AACTGTTGGGTCTTCGTCTAACC	7			
ST36*prp*	F3: CCCGAGGCACATCTTCACC	4	699	55	2
	R3: TAAACCACTAACATCTTCATCTACC	5			
ST36*cps*	F1: TTGAGAATTACTTCCGATTATGTAGA	9	889	55	2
	R1: TAAACGCATTAGCGAATAGTGC	10			
ST36*flp*	F1: TGGTTGTGTTTAGAGCAGGG	36	747	55	2
	R1: TGTTGGTAATACGATAAGAATGAGA	37			
ST631*end*	F1:AGTTCATCAGGTAGAGAGTTAGAGGA	53	494 (1.5 min)	57	3
	R1: TCTTCGTTACCATAGTATGAGCCA	54			
ST631-ENV	F1: TGGGCGTTAGGCTTTGC	59	496 (1.5 min)	57	3

	F2: GGGCTTCTACGACTTTCTGCT	60			
ST34*reg*	F1:TCCTTGTTGTCGTTGAATATCAGTC R1: GATACGGTTGTCATGCCTATGTGT	63 64	592 (1.5 min)	60	3
ST674*hyp*	F1: GAAGATGGTCCAAGAGGGAAGC R1:CTATCAGAGATTGAGCAAGTAGC	69 70	449 (1.5 min)	56	3
ST1127*hyp*	F1:CGTAAAGTAAAAGAGCCTGGTC R1:TTCAGTTCCCCGCATTCAT	75 76	1234 (1.5 min)	56	3
ST36Phage	F2: AGCAACGAAAACGCCTGT R2: ACCGTATCACCAATGGACTGT	81 82	1000 (1.5 min)	55	3
ST36NEOrf10-Hyp	F: TTTCTTACTTCTGTGAGCATTTGA R: GATTACTGAGCCTCTAAAGCCGTC	87 88	618 (1.5 min)	55	3
ST36PhHypD-Orf9	F3: AAGTGCTACATGAATGAAAGTGCT R1: TCAATGAAGTATCACGAAATGACTA	94 95	854 (MA-36) 1440 (MA-26)	55	3
TdhUreG	fwd4: GAATGCTGCCAACATGGATATAAAT rev5: GACAAAGGTATGCTGCCAAAAGTG	100 101	2638 (3 min)	55	3

Sources: (1) Panicker et al., (2004) Appl Environ Microbiol. Dec; 70(12): 7436–7444; (2) Whistler et al. 2015 *J. Clin. Microbiol. June 2015 vol. 53 no. 6 1864-1872*; (3) unpublished

Table 3: Oligonucleotide primers, examples of labels, and Tm conditions.

Locus	Primer sequence	SEQ ID NO	Modification	Tm	Source
Tlh	F: ACTCAACACAAGAAGAGATCGACAA	104		61	(1)
	R: GATGAGCGGTTGATGTCCAA	105		63	
	Probe: CGCTCGCGTTCACGAAACCGT	106	5′ TxRED to 3′ BHQ2	72	
tdh	F: TCCCTTTTCCTGCCCCC	110		63	(1)
	R: CGCTGCCATTGTATAGTCTTTATC	111		61	
	Probe: TGACATCCTACATGACTGTG	112	5′ FAM to 3′ MGBNFQ	47	
trh	F: TTGCTTTCAGTTTGCTATTGGCT	116		64	(1)
	R: TGTTTACCGTCATATAGGCGCTT	117		64	
	Probe: AGAAATACAACAATCAAAACTGA	118	5′ TET to 3′ MGBNFQ	55	
*prp** 242bp	F1: TGACGCAGCTTTAGGGACTAACTTA[1]	122		64	(2)
	R1: CGCTAAACCACTAACATCTTCATCTAC	123		65	
	F2: GCGAGGAGGGTTATTCTGACTTAG	124		64	
	R2:CTAAGTCAGAATAACCCTCCTCGC[1]	125		64	
	R3: AAATCTATGAATTTGTAGAAGTGGTGTG	126		63	
	Probe1: ATACCTGTCAGTCGTCGAAATTGGTGGAG[1]	127	5′ Cy3 to 3′ BHQ2	74	
	Probe2: ATACCTGTCAGTCGTCGAAATTGGTGGA	128	5′ TxRED to 3′ BHQ2	73	
	Probe3: ATACCTGTCAGTCGTCGAAATTGGTGGA	129	5' VIC to 3'BHQ1	73	
	Probe4: CAACATCATGCCGCTGTAATGGTGAG	130	5′ Cy3 to 3′ BHQ2	72	
flp 143bp 211bp	F: AATCTGTTCTGGTTGTGTTTAGAGC	140		62	(2)
	R: AAAGGTCCCAAAGCTGTCAATATAG	141		62	
	R2: AAGTCTTTTCATCACCTTCATTGC	142		62	
	Probe: CTAGGGGTTGAACAGTTTGGATCTTTTAGTTTA	143	5′ FAM to 3′ BHQ1	73	
	Probe2: CTAGGGGTTGAACAGTTTGGA	144	5′ FAM to 3′ MGBNFQ	60	
end 162bp	F: GCTGAGAGGTTGGAAAGTGTAGC	151		62	(2)
	R: TCTTCGTTACCATAGTATGAGCCA	152		62	
	F2: TATCGGTGGGATTTCTGGTGT	153		59	
	R2: CATTAACCAAGGAGTAGGAGGAAGT	154		60	
	Probe: GGGTGATCATGGAGCATAGTGAAGGTAAA	155	5′ TxRED to 3′ BHQ2	72	
	Probe2: TCTCAAGCAGGGTATCAATCAGGCG	156	5′ TxRED to 3′ BHQ2	72	
631-ENV	F TGATAGTCGTGGTTGCGTGC	170		63	(2)
	R AGCTGAATCGGTTTGGTTACA	171		62	
	Probe: TTGGGCGTTAGGCTTTGCGTACA	172	TBD	72	
IAC	F: GACATCGATATGGGTGCCG	164		62	(1)
	R: CGAGACGATGCAGCCATTC	165	5′ Cy5 to 3′ BHQ2	62	
	Probe: TCTCATGCGTCTCCCTGGTGAATGTG	166		74	

*HEX, VIC, or JOE would be substituted for Cy3 if using Agilent device instead of Cephid Smartcycler, and TxRed if prp is combined with tdh and trh. prpF1/R1 have 330 bp amplicon, and [1]prpF1 and prpR2 give 242 bp amplicon- use with probe 1 on smartcycer. Primer R3 will give ~100bp band. Can us prpF2/R1 for SYBR assay. F1 and R3. Sources: (1) Nordstrom, et al., (2007) Appl Environ Microbiol. Sep;73(18):5840-7. Epub 2007 Jul 20 and (2) Unpublished.

Methods of Detection

The invention in some aspects provides methods for detecting the presence of a pathogenic *V. parahaemolyticus* molecule in a sample. The method may comprise contacting the sample with an agent capable of detecting pathogenic *V. parahaemolyticus* polypeptide or polynucleotide molecules (e.g., pathogenic *V. parahaemolyticus* RNA, cDNA, mRNA, or DNA, etc.) such that the presence of pathogenic *V. parahaemolyticus* is detected in the sample. An agent for detecting pathogenic *V. parahaemolyticus* polynucleotide using methods of the invention may be a labeled or labelable nucleic acid probe capable of hybridizing to *V. parahaemolyticus* mRNA. The nucleic acid probe may be, for example, the full-length *V. parahaemolyticus* sequences set forth herein, including but not limited to: SEQ ID NO:1 (*prp* locus), SEQ ID NO:8 (*cps* locus), SEQ ID No. 11 (*tlh* locus), SEQ ID NO:15, (*tdh* locus), SEQ ID NO:18 (*trh* locus) and SEQ ID NO:38 (*flp* locus), or a portion thereof, such as an oligonucleotide of at least 3, 4, 5, 10, 15, 30, 50, 100 or more nucleotides in length and sufficient to specifically hybridize under stringent conditions to pathogenic *V. parahaemolyticus* mRNA. Polypeptides that may be detected using methods of the invention, and whose presence may indicate the presence of a pathogenic *V. parahaemolyticus* bacteria include, but are not limited to the polypeptides set forth herein as SEQ ID NO:41, 42, and 46. Polynucleotides that may be used as probes in certain embodiments of methods and kits of the invention include, but are not limited to primers and probes set forth in Tables 2 and 3.

The term "labeled or labelable", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the diagnostic molecule (e.g., probe, antibody, binding molecule, etc.), as well as indirect labeling of the molecule by reactivity with another reagent that is directly labeled. Some non-limiting examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with a fluorescent molecule, with biotin such that it can be detected with fluorescently labeled streptavidin, or other detectable label.

Detection methods useful in certain embodiments of methods of the invention, including but not limited to those described above herein, can be used as the basis for a

method of detecting the presence or absence of a pathogenic *V. parahaemolyticus* molecule, can be used as the basis for a method of monitoring the progression of an infection or contamination by the pathogenic *V. parahaemolyticus* bacterium in a subject or substrate, or can be used as the basis for a method of determining a prognosis for a subject at risk for infection with or a subject or substrate at risk for contamination by a pathogenic *V. parahaemolyticus* bacterium.

Some aspects of the invention provide a prognostic method for determining whether a subject is at risk for developing an infection by pathogenic *V. parahaemolyticus* bacteria, or whether a substrate of subject is at risk for contamination by a pathogenic *V. parahaemolyticus* bacterium. Such methods may include contacting a sample obtained from the subject or substrate (or isolate of the sample) with an agent capable of detecting a pathogenic *V. parahaemolyticus* polypeptide or polynucleotide such that the presence and/or level of the pathogenic *V. parahaemolyticus* polypeptide or polynucleotide is detected in the sample or isolate, thereby determining whether the subject or substrate has an infection and/or contamination and/or is at risk for developing an infection or contamination by the pathogenic *V. parahaemolyticus* bacterium.

In some aspects and embodiments, diagnostic methods of the present invention may include determining the level of a pathogenic *V. parahaemolyticus* polynucleotide in a sample or isolate from a sample. In certain embodiments, methods of the present invention may include comparing the level of a pathogenic *V. parahaemolyticus* polypeptide or polynucleotide in a sample or isolate with the level of the pathogenic *V. parahaemolyticus* polypeptide or polynucleotide in a control sample. In yet another embodiment, a diagnostic or prognostic method of the invention may also include a step of forming a prognosis or forming a diagnosis, and may also include a step of determining an appropriate treatment for a subject or substrate by a health-care provider, environmental scientist, scientist, public health personnel, or other appropriate party, based at least in part on the determination of the presence, absence, and/or amount of the pathogenic *V. parahaemolyticus* molecule in a sample from the subject or substrate.

In certain aspects of the invention, determining the presence and/or amount of a pathogenic *V. parahaemolyticus* molecule can be carried out in conjunction with the determination of one or more other biomarkers of the pathogenic *V. parahaemolyticus* bacteria. For example, the level and/or presence of a pathogenic *V. parahaemolyticus* molecule (e.g., a polynucleotide that encodes a pathogenic *V. parahaemolyticus* polypeptide, a pathogenic *V. parahaemolyticus* polypeptide or signature fragment thereof), may be

detected using a panel that also permits detection at least one additional molecule whose expression and/or level is may be useful to characterize an infection or contamination with the pathogenic *V. parahaemolyticus* bacteria. For example, a panel of the invention may comprise a means to determine at least one of the level or presence one of more of a *prp* locus, a *cps* locus, and an *ftp* locus, etc. to generate an aggregate biomarker panel. In certain embodiments, a panel of the invention may comprise a means to determine at least one of the level or presence one of more of a *prp* locus, a *cps* locus, a *tlh* locus, a *tdh* locus, a *trh* locus, and an *ftp* locus, etc. to generate a biomarker panel that may be used in certain embodiments of the invention. A biomarker panel of the invention may include detection means for at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 30, 50, 100 or more biomarkers (including each integer in between the listed integers) in addition to a detection means for a pathogenic *V. parahaemolyticus* bacteria as set forth herein. As used herein, in certain aspects of the invention the term biomarker means a polynucleotide or polypeptide that is specific to a pathogenic *V. parahaemolyticus* bacteria

Pathogenic *V. parahaemolyticus* Sequences and Variants

One aspect of the invention involves isolated polynucleotide molecules that encode polypeptides of pathogenic *V. parahaemolyticus* bacteria or biologically active portions thereof, as well as polynucleotide fragments sufficient for use as hybridization probes to identify a pathogenic *V. parahaemolyticus* bacteria polypeptide-encoding polynucleotide. A polynucleotide or polypeptide of the invention may be a "signature" polynucleotide or polypeptide, which, as used herein means a polynucleotide or polypeptide that is present in a pathogenic *V. parahaemolyticus* bacteria but not present in a non-pathogenic bacterium, including but not limited to a non-pathogenic parahaemolyticus bacterium. Thus, the presence of a signature nucleic acid or polypeptide of a pathogenic *V. parahaemolyticus* bacterium in a biological sample indicates the presence of the pathogenic *V. parahaemolyticus* bacterium in the sample. Examples of signature polynucleotide sequences provided herein, though not intended to be limiting are *V. parahaemolyticus* sequences set forth as SEQ ID NO:1 (*prp* locus), SEQ ID NO:8 (*cps* locus), and SEQ ID NO:38 (*flp* locus), or fragments thereof that are sufficiently unique so as to identify the presence of the *prp* locus, *cps* locus, or *flp* locus, respectively. Examples of some of the signature polypeptide sequences provided herein, though not intended to be limiting are *V. parahaemolyticus* sequences set forth as SEQ ID NO:41 (*prp* locus), SEQ ID NO:42 (*cps* locus), and SEQ ID

NO:46 (*flp* locus), or fragments thereof that are sufficiently unique so as to identify the presence of the *prp* locus, *cps* locus, or *flp* locus, respectively.

As used herein, the term "polynucleotide" is intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., RNA, mRNA). A polynucleotide may be single-stranded or double-stranded or may be a double-stranded DNA molecule. An "isolated" polynucleotide molecule is free of sequences that naturally flank the polynucleotide (i.e., sequences located at the 5' and 3' ends of the polynucleotide) in the genomic DNA of the organism from which the polynucleotide is derived. The terms: "polynucleotide", "oligonucleotide", may be used interchangeably. For example an oligonucleotide primer may also be referred to as a polynucleotide primer, etc. Moreover, an "isolated" polynucleotide such as a cDNA molecule may be free of other cellular material.

In some aspects of the invention, an isolated polynucleotide of the invention comprises the nucleic acid sequence set forth herein as SEQ ID NO:1 (*prp* locus), SEQ ID NO:8 (*cps* locus), SEQ ID NO. 11 (*tlh* locus), SEQ ID NO:15, (*tdh* locus), SEQ ID NO:18 (*trh* locus), SEQ ID NO:38 (*flp* locus), or another sequence set forth in Tables 2 and 3 or elsewhere herein, or a variant and/or fragment thereof. Some nucleic acid sequences provided herein are identified as sequences of polynucleotides found in pathogenic *V. parahaemolyticus* bacteria, for example SEQ ID NOs: 1, 8, and 38. Some nucleic acid sequences provided herein are identified as sequences of polynucleotides found in *V. parahaemolyticus* bacteria that need not be pathogenic, for example SEQ ID NOs:11, 15, and 18. Table 1 provides additional information in reference to the prp, cps, flp, tdh, trh, and tlh sequences.

The invention further encompasses polynucleotide molecules that differ from the sequences for the *V. parahaemolyticus* sequences set forth as SEQ ID NO:1 (*prp* locus), SEQ ID NO:8 (*cps* locus), SEQ ID No. 11 (*tlh* locus), SEQ ID NO:15, (*tdh* locus), SEQ ID NO:18 (*trh* locus), SEQ ID NO:38 (*flp* locus) or another sequence set forth in Tables 2 and 3 or elsewhere herein, or a portion thereof, due to degeneracy of the genetic code and thus encode the same *V. parahaemolyticus* polypeptide as that encoded by the sequences set forth as set forth as SEQ ID NO:1 (*prp* locus), SEQ ID NO:8 (*cps* locus), SEQ ID No. 11 (*tlh* locus), SEQ ID NO:15, (*tdh* locus), SEQ ID NO:18 (*trh* locus), SEQ ID NO:38 (*flp* locus), or another sequence set forth in Tables 2 and 3 or elsewhere herein, respectively. Encoded *V. parahaemolyticus* polypeptides include, but are not limited to sequences set forth herein as SEQ ID NOs:41, 42, 43, 44, 45, and 46, and additional amino acid sequences disclosed herein.

A polynucleotide having the nucleic acid sequence as set forth herein as SEQ ID NO:1 (*prp* locus), SEQ ID NO:8 (*cps* locus), SEQ ID No. 11 (*tlh* locus), SEQ ID NO:15, (*tdh* locus), SEQ ID NO:18 (*trh* locus), SEQ ID NO:38 (*flp* locus), or another sequence set forth in Tables 2 and 3 or elsewhere herein, or a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. For example, a cDNA library can be probed using all or portion of a sequence set forth as SEQ ID NO:1 (*prp* locus), SEQ ID NO:8 (*cps* locus), SEQ ID No. 11 (*tlh* locus), SEQ ID NO:15, (*tdh* locus), SEQ ID NO:18 (*trh* locus), SEQ ID NO:38 (*flp* locus) as a hybridization probe and standard hybridization techniques (e.g., as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. Molecular Cloning: A Laboratory Manual. 2nd. edition, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1989).

Moreover, a polynucleotide encompassing all or a portion of a sequence set forth as SEQ ID NO:1 (*prp* locus), SEQ ID NO:8 (*cps* locus), SEQ ID No. 11 (*tlh* locus), SEQ ID NO:15, (*tdh* locus), SEQ ID NO:18 (*trh* locus), SEQ ID NO:38 (*flp* locus) or other polynucleotide sequence disclosed herein including but not limited to a ST36*end* locus, a ST34*reg* locus, a ST674*hyp* locus, a ST1127*hyp* locus, etc. can be isolated using any suitable method, including as a non-limiting example, use of the polymerase chain reaction using oligonucleotide primers set forth herein and/or designed based upon the sequence set forth herein as SEQ ID NO:1 (*prp* locus), SEQ ID NO:8 (*cps* locus), SEQ ID No. 11 (*tlh* locus), SEQ ID NO:15, (*tdh* locus), SEQ ID NO:18 (*trh* locus), SEQ ID NO:38 (*flp* locus), ST36*end* locus, ST34*reg* locus, ST674*hyp* locus, ST1127*hyp* locus, etc. can be isolated. For example, *V. parahaemolyticus* mRNA can be isolated from samples using standard, art-known methods and cDNA can be prepared using reverse transcriptase and art-known methods. Synthetic oligonucleotide primers for PCR amplification, such as those provided herein, can be designed based upon one or more nucleic acid sequences set forth herein as SEQ ID NO:1 (*prp* locus), SEQ ID NO:8 (*cps* locus), SEQ ID No. 11 (*tlh* locus), SEQ ID NO:15, (*tdh* locus), SEQ ID NO:18 (*trh* locus), SEQ ID NO:38 (*flp* locus), a ST36*end* locus, a ST34*reg* locus, a ST674*hyp* locus, a ST1127*hyp* locus, etc. and polynucleotides of the invention can be amplified using cDNA or, alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The polynucleotide so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to pathogenic and non-pathogenic *V. parahaemolyticus* polypeptide-encoding nucleic acid sequences can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

In addition to the pathogenic and non-pathogenic *V. parahaemolyticus* bacterial nucleic acid sequences set forth herein as SEQ ID NO:1 (*prp* locus), SEQ ID NO:8 (*cps* locus), SEQ ID No. 11 (*tlh* locus), SEQ ID NO:15, (*tdh* locus), SEQ ID NO:18 (*trh* locus), and SEQ ID NO:38 (*flp* locus), ST36*end* locus, ST34*reg* locus, ST674*hyp* locus, ST1127*hyp* locus, etc., it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of pathogenic or non-pathogenic *V. parahaemolyticus* may exist within a population (e.g., the bacterial population). Such genetic polymorphism in the *V. parahaemolyticus* gene may exist among bacteria within a population due to natural allelic variation. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of a gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in *V. parahaemolyticus* that are the result of natural allelic variation and that do not alter the pathogenicity of the *V. parahaemolyticus* bacteria are intended to be within the scope of the invention. Polynucleotide molecules corresponding to natural allelic variants of a *V. parahaemolyticus* cDNA of the invention can be isolated based on their homology to the polynucleotides disclosed herein using the cDNA, or a portion thereof, as a hybridization probe - according to standard hybridization techniques under stringent hybridization conditions, which are recognized in the art.

In some aspects of the invention, an isolated polynucleotide of the invention that hybridizes under stringent conditions to a nucleic acid sequence set forth as SEQ ID NO:1 (*prp* locus), SEQ ID NO:8 (*cps* locus), SEQ ID No. 11 (*tlh* locus), SEQ ID NO:15, (*tdh* locus), SEQ ID NO:18 (*trh* locus), SEQ ID NO:38 (*flp* locus), ST36*end* locus, ST34*reg* locus, ST674*hyp* locus, a ST1127*hyp* locus, or other sequence provided herein, corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleic acid sequence that occurs in nature (e.g., encodes a natural protein). In one embodiment, the nucleic acid encodes a natural *V. parahaemolyticus* polypeptide.

In addition to naturally occurring allelic variants of the *V. parahaemolyticus* sequence that may exist in the bacterial population, the skilled artisan will further appreciate that changes may be introduced by mutation into a nucleic acid sequence set forth as SEQ ID NO:1 (*prp* locus), SEQ ID NO:8 (*cps* locus), SEQ ID No. 11 (*tlh* locus), SEQ ID NO:15, (*tdh* locus), SEQ ID NO:18 (*trh* locus), SEQ ID NO:38 (*flp* locus), ST36*end* locus, ST34*reg* locus, ST674*hyp* locus, ST1127*hyp* locus, or other sequence provided herein thereby leading to changes in the amino acid sequence of the encoded *V. parahaemolyticus* protein without

altering the functional ability of the *V. parahaemolyticus* protein. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues may be made in the sequences. A "non-essential" amino acid residue is a residue that can be altered from a *V. parahaemolyticus* polypeptide without altering the activity of the *V. parahaemolyticus* polypeptide, whereas an "essential" amino acid residue is required for the *V. parahaemolyticus* polypeptide activity.

The invention, in some embodiments, includes use of full-length sequences of fragments of sequences of polynucleotides that encode pathogenic and non-pathogenic *V. parahaemolyticus* polypeptides, but that differ in sequence from the nucleic acid sequences set forth as SEQ ID NO:1 (*prp* locus), SEQ ID NO:8 (*cps* locus), SEQ ID No. 11 (*tlh* locus), SEQ ID NO:15, (*tdh* locus), SEQ ID NO:18 (*trh* locus), SEQ ID NO:38 (*flp* locus), ST36*end* locus, ST34*reg* locus, ST674*hyp* locus, ST1127*hyp* locus, or other sequence provided herein and yet can still be used in method of the invention to detect one or more of the presence and level of a pathogenic *V. parahaemolyticus* bacteria in a sample. In certain embodiments of the invention, a variant of a *prp* polynucleotide, useful in a method of the invention, comprises a nucleic acid sequence that has at least 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity to the sequence set forth as SEQ ID NO:1 or fragment thereof; a variant of a *cps* polynucleotide useful in a method of the invention comprises a nucleic acid sequence that has at least 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity to the sequence set forth as SEQ ID NO:8 or fragment thereof; a variant of a *tlh* polynucleotide useful in a method of the invention, comprises a nucleic acid sequence that has at least 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity to the sequence set forth as SEQ ID NO:11 or fragment thereof; a variant of a *tdh* polynucleotide useful in a method of the invention, comprises a nucleic acid sequence that has at least 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity to the sequence set forth as SEQ ID NO:15 or fragment thereof; a variant of a *trh* polynucleotide useful in a method of the invention, comprises a nucleic acid sequence that has at least 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity to the sequence set forth as SEQ ID NO:18 or fragment thereof; a variant of a *flp* polynucleotide useful in a method of the invention, comprises a nucleic acid sequence that has at least 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity to the sequence set forth as SEQ ID NO:38 or fragment thereof; and a variant; or a ST36*end* locus, ST34*reg* locus, ST674*hyp* locus, ST1127*hyp* locus, or other polynucleotide sequence provided herein that is

useful in a method of the invention, comprises a nucleic acid sequence that has at least 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity to its corresponding polynucleotide sequence of the ST36*end* locus, ST34*reg* locus, ST674*hyp* locus, ST1127*hyp* locus, or other polynucleotide sequence provided herein. respectively.

It will be understood that a fragment of one or more polynucleotides comprising a nucleic acid sequence set forth herein as SEQ ID NO:1, SEQ ID NO:8, SEQ ID No. 11, SEQ ID NO:15, SEQ ID NO:18, or SEQ ID NO:38 or a polynucleotide sequence of the ST36*end* locus, ST34*reg* locus, ST674*hyp* locus, ST1127*hyp* locus, or other polynucleotide sequence provided herein can be used in methods of the invention, for example, though not intended to be limiting, as primers or probes. A fragment as used herein may be a portion of a sequence set forth herein, or variant thereof. As used herein, a portion of a polynucleotide having a length of "n" means a polynucleotide having a length of "n" minus at least one, two, three, four, five, or more nucleotides. For example if "n" equals the number of nucleotides in a sequence, a fragment may have a length of "n" minus at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30 ,40, 50, 60, 70, 80, 90, 100, or more nucleotides (including all intervening integers). It will be understood by those in the art that the number of nucleotides that can be subtracted from a sequence to result in a fragment will depend on the number of nucleotides in the sequence, for example, more fragment sizes can be obtained from a polynucleotide having nucleic acid sequence that is 2000 nucleotides in length than can be obtained from a polynucleotide having a nucleic acid sequence that is 100 nucleotides in length. In certain embodiments of the invention, a polynucleotide fragment includes between 4 and "n" minus one nucleotides of the nucleic acid sequence having "n" nucleotides set forth herein, or a variant thereof.

To determine the percent similarity/identity of two polynucleotides the nucleic acid sequences of the polynucleotides can be aligned for optimal comparison purposes (e.g., gaps may be introduced in the sequence of one polynucleotide for optimal alignment with the other polynucleotide). The nucleotide residues at corresponding positions are then compared. When a position in one sequence is occupied by the same nucleotide residue as the corresponding position in the other sequence (e.g., a variant form of the *V. parahaemolyticus* sequence), then the molecules have identity at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % similarity = number of identical positions/total number of positions x 100). Such an alignment can be performed using any one of a number of well-known computer algorithms designed and used in the art for such a purpose. In some aspects of the invention, a variant of

a polynucleotide comprises a fragment of the polynucleotide. In certain aspects of the invention, a variant of a polynucleotide is or comprises a portion of the polynucleotide's nucleic acid sequence that has at least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the corresponding region of the polynucleotide's nucleic acid sequence when the sequences are aligned. For example, a polynucleotide fragment that is 20 nucleotides in length may have 100% identity with the nucleic acid sequence of a polynucleotide that is 200 nucleotides in length if all 20 nucleotides have identical positions shared by the sequences (e.g., are identical across the corresponding region) when the two polynucleotide sequences are aligned. Similarly, the sequence of a polynucleotide fragment that is 20 nucleotides long has identity of at least 90% to a nucleic acid sequence of a polynucleotide that is 200 nucleotides long if 18, 19, or 20 of the 20 nucleotides of the fragment are the same in the corresponding region of the nucleic acid sequence of the 200 nucleotide polynucleotide when the nucleic acid sequences of the fragment and the 200 nucleotide polynucleotide are aligned.

Polynucleotide fragments that may be useful in some embodiments of the invention may be primers or probes. Non-limiting examples of primer or probe sequences that may be useful in some embodiments of the invention are set forth herein as SEQ ID NOs: 2-7, 9, 10, 12-14, 16, 17, 19, 20, 36, 37, and other sequences set for in Tables 2 and 3, or complements thereof, some of which are set forth herein as SEQ ID Nos: 49, 50, 55, 56, 61, 62, etc. Those skilled in the art will understand how to determine a complement sequence using standard methods. In certain embodiments a primer sequence may also include additional elements, for example, though not intended to be limiting: a tag, detectable label, identifying sequence for purification, visualization, additional nucleotides, etc.

An isolated polynucleotide useful in methods of the invention can be created by introducing one or more nucleotide substitutions, additions, or deletions into a nucleic acid sequence such as, but not limited to, SEQ ID NO: 1, 8, 11, 15, 18, and 38, sequences set forth herein of the ST36*end* locus, ST34*reg* locus, ST674*hyp* locus, ST1127*hyp* locus, other polynucleotide sequence provided herein; or fragments thereof. Mutations can be introduced into the sequences by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. The following examples are provided to illustrate specific instances of the practice of the present invention and are not intended to limit the scope of the invention. As will be apparent to one of ordinary skill in the art, the present invention will find application in a variety of compositions and methods.

Examples

Example 1

Identification of strains as ST36 currently requires costly and time-consuming DNA isolation, PCR amplification and sequencing of 7 housekeeping loci; however, a subset of the loci informs whether isolates are likely to be or are related to ST36. From among 94 clinical isolates from infections reported in Massachusetts, New Hampshire, and Maine between 2010-2013, during which time infections from the ST36 strain were first reported from Atlantic sources [Martinez-Urtaza J, et al., (2013) N Engl J Med 369:1573-1574], 43 isolates were identical to ST36 at four housekeeping loci (*dnaE, dtdS, pntA and tnaA*) [González-Escalona N, et al., (2008) J Bacteriol 190:2831-2840]. The relationships of these probable ST36 isolates to 47 of the remaining clinical isolates from the region was determined by constructing a neighbor-joining tree of the concatenated and aligned sequences [Tamura K, et al. (2013) Mol Biol Evol 30:2725-2729] (Fig. 1). This identified three additional isolates that were related to but distinct from the ST36 clade. The analysis of all seven housekeeping loci using Illumina whole genome short reads from four probable ST36 isolates, including MAVP-26 (SAMN03107383), MAVP-36 (SAMN03107385), MAVP-45 (SAMN03177810) and MAVP-V (SAMN03177809) obtained from the Hubbard Center for Genome Studies at the University of New Hampshire, confirmed that these isolates are ST36 [Inouye M, et al., (2012) BMC Genomics 13:338; Jolley K. (2010) Database:pubmlsthttp://pubmlst.org/vparahaemolyticus/info/protocol.shtml. Accessed 01 Feb 2013].

To identify genomic differences for use in a ST36-diagnostic assay, whole genome comparisons were performed between the published draft genome for serotype O4:K12 ST36 strain 10329 (the 33 contigs including NZ_AFBW01000001.1 - 33.1) [Gonzalez-Escalona N, et al., (2011) J Bacteriol 193:3405-3406] and the genomes of two other pathogenic strains, including the pandemic strain RIMD 2210633 (NC_004605.1, NC_004603.1) [Makino K, et al, (2003) The Lancet 361:743-749] and pre-pandemic strain BB22OP (NC_019955.1, NC_019971.1) [Jensen RV, et al., (2013). Genome Announce 1:e00002-00012]. The unique genome content was visualized [Alikhan N-F, et al., (2011) BMC Genomics 12:402 DOI:410.1186/1471-2164-1112-1402; Darling AC, et al., (2004) Genome Res 14:1394-1403] and re-annotated using a *Vibrio* sp. specific database in NCBI with Prokka 1.8 [Seemann T. (2014) Bioinformatics: DOI:10.1093/bioinformatics/btu1153]. Few coding regions (~ six) in three different contigs appeared unique to strain 10329. It was then determined the extent the ORFs within these regions were present in and potentially exclusive to all strains in the 10329 genome group based on comparisons with draft genomes of *V. parahaemolyticus*

available at the time of this analysis (289 total) (www.ncbi.nlm.nih.gov/genome/691), accessed 10/12/2014) [Altschul SF, et al., (1997). Nucl Acids Res 25:3389-3402].

ORFs that were absent from more than one high-quality draft genome of the 10329 genome group (not all of which are ST36), or that were present in many other genome groups were eliminated from further consideration thereby focusing further analysis on two different regions of contig 10329_28 that were also present in the four Northeast regional ST36 isolates. ST36 diagnostic loci were identified. Genomic were identified, including regions that were present in 10329_28 of strain 10329 and in Northeastern ST36 isolates, including MAVP-V, MAVP-26, MAVP-36 and MAVP-45, but that were absent in pandemic and pre-pandemic strains including the *prp*, *cps* and O-antigen flippase (*flp*) genes. From the first region in this contig, an ORF identified as a pathogenesis-related protein (locus *prp*) (EGF42613) based on its similarity to a single ORF in *Vibrio* sp. Ex25 (YP_003285914.1) was selected as a potential diagnostic locus due to its uniqueness, even outside the species.

The designation of *prp* was proposed for this locus even though the gene more likely encodes an endonuclease or DNA helicase based on similarity searches in NCBI [Altschul SF, et al., (1997) Nucl Acids Res 25:3389-3402] in the absence of compelling evidence for function and a different name. Two additional ORFs including one identified as encoding a capsular polysaccharide (locus *cps*) (EGF42671) and another as O-antigen flippase (locus *flp*) (EGF42675) in a second region of the same contig were also identified as diagnostic targets due to their potential role in conferring the O4 antigenic property of the strain, which is a useful diagnostic phenotype.

To determine the extent that one or a combination of these loci are present only in strains closely related to ST36, multiple reference whole genome phylogenies were constructed with REALPHY v. 1.09 for a few high quality genomes from each NCBI genome group lineage that harbored at least one of the three loci under evaluation, and for strains branching relatively closely with the 10329 genome group (www.ncbi.nlm.nih.gov/genome/?term=vibrio%20parahaemolyticus) [Bertels F, et al., (2014) Mol Biolo Evol 31:1077-1088] (see Fig. 2). This phylogeny is based on a subset of high quality genomes compared to NCBI BLAST phylogeny, thereby using a higher proportion of informative sites. The three identified genes only co-occur in ST36, closely related ST59 and one single other related strain (vpV223/04) of unknown sequence type (see Fig. 2, Fig. 4).

Only one high quality genome (MDVP13, ST678) in the 10329 genome group apparently lacks *prp* (see Fig. 2, Fig. 4). Five other genomes harbored one or two of the three loci. These include strains in four different NCBI genome groups (NIHCB0757, S159, S048,

and SNUVpS-1) but not every strain in these genome groups harbored these genes. Based on its distribution and uniqueness, *prp* is likely sufficient for strain identification. It was reasoned that due to the complexity of serotypes within the ST36 clade, potential selection upon the locus that could produce variation [Banerjee SK, et al., (2014) J Clin Microbiol 52:1081-1088], and the occurrence of *cps* and/or *flp* in four more distantly related strains (Fig. 4), these may be used alone or in combination with each other and/or with *prp*. A combination *prp* with just one other locus (i.e. *cps* or *flp*) may be especially accurate.

Oligonucleotide primers that produce *prp*-, *cps*- and *flp*-specific amplicons were developed for simultaneous detection along with both hemolysin-encoding genes (*tdh* and *trh*) and the species specific locus (*tlh*) to improve an existing multiplex PCR assay [see for example: Panicker G, et al., (2004) Appl Environ Microbiol 70:7436-7444 and J.L. Nordstrom, et al., App Environ Microbiol (2007) 73:5840-5847]. Table 2 shows non-limiting examples of primers utilized in PCR amplifications.

Alignments of 29 available *prp* sequences [Tamura K, et al., (2013) Mol Biol Evol 30:2725-2729] allowed identification of regions of high sequence identity, which informed primer design to produce optimal amplicon size separation with the existing multiplex PCR amplicons. Secondary structure and primer cross-dimerization were minimized by design with the aid of the NetPrimer program (PREMIER Biosoft, CA, US). A similar strategy was used to design the *cps* and *flp*-specific PCR. Amplification of the *prp* locus was evaluated in individual and multiplex assays using genomic DNA from positive (F11-3A), and negative (G61) [Mahoney JC, et al., (2010) Appl Environ Microbiol 76:7459-7465; Ellis, CN, et al., (2012) Appl Environ Microbiol 78: 3778-3782] control strains using AccuStart PCR Supermix (Quanta, MD, US) with an initial denaturation at 94° C for 3 minutes, followed by 30 cycles with a denaturation at 94° C for 1 minute, primer annealing at 55° C for 1 minute, and extension at 72° C for 1 minute, and with a final extension at 72° C at the completion of the cycling for 5 minutes.

Amplicons from the amplifications were evaluated using standard procedures, for example: by electrophoresis of 1.5 µl of sample on 1.2% SeaKem LE agarose (Lonza, Rockland, ME, USA) gel with 1× GelRed (Phenix Research Products, Candler, NC, USA) in Tris-acetate-EDTA (TAE) buffer compared against a 1-kb Plus DNA ladder (Invitrogen, Grand Island, NY, USA).

The developed assay was also applied to the collection of 94 clinical strains and several closely related environmental strains (i.e. G1350) from the Northeast. The *prp* amplicon was detected in all 43 strains that grouped within the ST36 clade and no other

clinical or environmental isolates from northern New England, even the three strains identified as most closely related to the ST36 clade (Fig. 3). A subset of *prp*-positive and -negative strains was subsequently analyzed for the presence of *cps*, which was also only detected in ST36 clade strains (Fig. 3).

These data indicate that the presence of the *prp* amplicon is diagnostic for ST36 complex strains, and that the *cps* (or *flp*) amplicon could further enhance accuracy of identification. Not only was the assay useful for clinical identification, but it should also aid in surveillance. It will help determine the extent of this strain's geographic expansion, establishment of stable local populations and the seasonal dynamics of these strains, thereby aiding in management of shellfish harvesting and reducing public health risk.

Example 2

Specific Identification of strains by Sequence Type using PCR

In addition to ST36, several other sequence types have contributed to the rise in infections, and have been recovered from shellfish harvest areas suggesting they are not transient but resident pathogens in harvest areas. ST631 has caused infections from Atlantic harvest areas from FL to PEI Canada, and currently accounts for ~10% of infections from Gulf of Maine sources. ST34 (and closely related ST324), ST674, and ST1127 are more rarely detected in harvest areas and cause fewer infections but may be emergent types. All five of these strain types carry the same pathogenicity island harboring both *tdh* and *trh*.

A similar approach to that described for identifying sequence-type (ST)-specific genetic loci for the invasive ST36 strain, (See Example 1; and Whistler et al, 2015), was used to develop a detection process for additional pathogenic lineages. Whole genome sequence comparisons were conducted, and using BLAST Ring Image Generator (BRIG) and Multiple Alignment of Conserved Genomic Sequence With Rearrangements (MAUVE), highly unique regions were visualized, located, and compared to various reference genomes. To ensure identified DNA was uniquely present only in specific lineages (or Sequence Types), BLAST and alignment comparisons were carried out using the NCBI database of all complete and draft genomes. Primers were designed (See Table 2 for invented strain identification primers) and evaluated using PCR amplification methods for specificity and accuracy on a collection of several hundred unique isolates, and the expected high degree of specificity/accuracy was confirmed.

Methods

Amplification

PCR amplifications were performed with AccuStart PCR Supermix (Quanta, Gaithersburg, MD, US) with an initial denaturation at 94° C for 3 minutes, followed by 30 cycles with a denaturation at 94° C for 1 minute, primer annealing at 55° C for 1 minute, and extension at 72° C for 1 minute, and with a final extension at 72° C at the completion of the cycling for 5 minutes. Altered extension times were also used as appropriate. For example, Table 2 provides PCR annealing temps and indicates certain primer pairs for which a 1.5 min extension time was used. For example, in multiplex assays, a 1.5 min extension time was used when multiple amplifications were performed in the same cycler. As noted in Table 2, the extension time used for the TdhUreG primer pair was 3 minutes. Standard procedures to optimize PCR amplifications were used.

For the locus PCR amplifications, the amplicon length determined the length in time of the PCR extension reaction, which was 1 Kb per minute. All of the reactions were 3 minutes or less and most were 1 minute. The annealing temperatures used were one degree below the melt temperature. If two primer pairs were used in multiplex, the lower annealing temperature of the primers was used. For example, if the annealing for one was 55^{o} and the other 57^{o}, the 55^{o} temperature was used.

Amplicon detection

Amplicons from the amplifications were evaluated using standard procedures, for example: by electrophoresis of 1.5 µl of sample on 1.2% SeaKem LE agarose (Lonza, Rockland, ME, USA) gel with 1× GelRed (Phenix Research Products, Candler, NC, USA) in Tris-acetate-EDTA (TAE) buffer compared against a 1-kb Plus DNA ladder (Invitrogen, Grand Island, NY, USA).

Additional studies have been performed using multiplex amplification methods including a tiered approach for strain identification in which the first round was the 5-amplicon (tlh/tdh/trh/prp3/cps) amplification, which was then double checked with prp2/flp amplification from the sample (see Fig. 5, PCR Reaction I). Then any isolate that was found to be not positive for ST36 in PCR Reaction I assay, was tested with other primers together for the other isolates, such as ST631 and ST1127 (see Fig. 5, PCR Reaction II and PCR Reaction III). Figure 5 shows results from the assays.

Example 3

Enumeration of ST36 and ST631 by quantitative PCR-based most probable number (MPN) methods

In order to use these specific loci for quantification of pathogenic *Vibrio parahaemolyticus*, qPCR assays were designed that are compatible with the currently employed FDA protocol designed by J. Jones (Nordstrom *et al.* 2007) and based on the Bacteriological Analytical Manual (see fda.gov/Food/FoodScienceResearch/LaboratoryMethods/ucm2006949.htm). Table 3 herein provides a list of oligonucleotide primers used for pathogen quantification.

Methods

Amplification Conditions

Amplification methods were designed and optimized as needed using routine optimization procedures. For example, cycling was based on the length of the amplicon and annealing was based on temperature of melting. Table 3 provides locus information, primer sequences, and melting temperatures for the primers used in qPCR procedures. Table 3 also provides sequences of probes and some examples of detectable labels that were used. The Quanta ToughMix (Quanta Biosciences Inc, Gaithersburg, MD) was used (according to manufacturer's instructions) for multiplex amplifications with primers for prp, flp, and end. The cycling parameters were as follows:

Cycle 1: 95°C hold for 60 s (this was increased as necessary depending on enzyme used)
Cycle 2 (45 times):
95°C for 5 s
59°C for 45 s.

Assay Steps

The MPN protocol of Nordstrom uses a two-step enumeration: In step one, enriched oyster homogenate lysates are evaluated for the presence of *V. parahaemolyticus* by detection of *tlh* with an internal amplification control IAC and in step two, lysates that are positive for *V. parahemoltyicus* are subsequently evaluated for the presence of two virulence associated markers, including *tdh* and *trh* (with the IAC control) allowing the determination of the most probable number of these bacteria per gram of oyster.

Studies that were performed identified that a major limitation of the Nordstrom test was that these markers are present in environmental strains that are likely not pathogenic, and

the markers are absent in some clinical strains so they are associated with virulence, but are not very useful for predicting the abundance of specific pathogen type strains.

Experiments were performed to develop a more specific detection strategy that can be added to this widely used platform and/or with other testing described herein, to improve risk assessment by specific quantification of the two strains that cause a predicted 85% of the cases from Atlantic harvest areas in the US. ST36 also causes more than 50% of the cases from the Pacific harvest areas of the US.

Experiments were carried out and a third assay step was developed. Studies were performed to evaluate lysates (samples) that were positive for either *tdh* or *trh,* and if found to be positive for one or both, the samples were further tested for the presence of ST36 by targeted amplification of both *prp* and *flp*, and for ST631 by the targeted amplification of *end* using probes and primers designed to these loci (Table 3), using qPCR amplification methods.

The probes were fluorescently labeled (see label information in Table 3) and they attached to the loci and when the primers amplified the sequence it released the quencher from the fluorofor, which was detected. These assays can also be performed using the detectable label: SYBR green or other detectable labels incorporated in the amplicon or linked to the probes, as indicated in Table 3 and elsewhere herein. Evaluation and quantitation of amplicons was performed using standard methods and manufacturer's instructions.

The developed PCR reactions were compatible with the FDA-required IAC internal control. Presence of these pathogenic bacteria in the lysates was further confirmed by use of the strain specific PCR primers by end-point binary PCR (Table 2) including the *cps* locus, and the *prp*2 primer pair for validation, and ST631-ENV *or end* primers.

Other types of qPCR assays that were carried out included a combined qPCR detection of prp, flp, and end polynucleotides. In the assays, an initial step included one cycle of tlh/IAC amplification/detection, and if results were positive a next step was to amplify and detect tdh/trh/IAC, which was followed by a third step of amplifying and detecting prp/flp/end/IAC.

Example 4

Identification of Sub-populations of ST36 that are localized to the Gulf of Maine (Vipa-26 marker) or Katama Bay of Martha's Vineyard (Vipa-36 marker) and invention of a population marker based PCR

Whole genome phylogenies revealed that there were distinctive lineages with diagnostic characteristics of the populations in the Gulf of Maine, for Katama Bay of Martha's Vineyard that would allow traceback of isolates. The unique attributes were the presence of phage Pathogenicity Islands: Vipa-26 in the Gulf of Maine, and Vipa-36 only in Katama Bay. Experiments were performed and a detection assay that included PCR amplification processes for specific identification of strains by population/harvest areas was developed. Following application of the multiplex PCR (see Example 1 and Whistler et al., 2015), isolates that were positive for the ST36 marker were subsequently analyzed for the presence of f237-like or Vipa-like phage, followed by a PCR reaction that distinguished the Gulf of Maine, and Katama Bay population by size difference of the amplicon (see Figure 6). PCR reactions were performed that included: PCR reaction I: to test for presence of tlh (tlhF2/tlhR primers, see Table 2), NE Phage (ST36NEOrf10F/ST36NEHypR primers, see Table 2), any f237-like phage (ST36PhageF2/R2 primers, see Table 2). Additional PCR was performed that included PCR reaction II: that tested for the presence of tlh (tlhF2/tlhR primers, see Table 2), Vipa 26 and Vipa 36 size difference (ST36PhHypDF3/ST36PhOrf9R1 primers, see Table 2). These primers were validated on the entire University of New Hampshire (UNH) collection of clinical strains, and these markers were identified as present only in ST36 clinical strains isolated from these two locations, and to be absent in all other clinical strains (including those in the NCBI database). For a list of primers, see Table 2. For amplification conditions see Examples 1 and 2).

Example 5

Identification of a shared pathogenicity island architecture among tdh/trh positive pathogens of several sequence types (ST36, ST631, ST34/324, ST674, ST1127, ST749, and ST110), and invention of PI-hybrid marker based PCR

Through the application of PacBio genome sequencing and highly accurate genome assembly, it was identified that the major emergent and invasive pathogenic strains of *V. parahaemolyticus* harboring both *tdh* and *trh* share a previously undiscovered hybrid pathogenicity island architecture. The unique feature of the island is the presence of a smaller tdh island flanked by transposases that have integrated within the major pathogenicity island containing the urease cluster, *trh*, and type three secretion beta genes that are implicated in defining virulence. It was identified that it is not simply the presence of the *tdh* and *trh* genes in the genomes that is characteristic, but the synteny or architecture of these elements within the same island. Therefore a strategy was developed whereby all strains with

this specific island architecture could be detected by PCR amplification through the juncture of these two islands from *tdh* to *ureG* and detection, for example, by electrophoresis.

Figure 7 shows results amplifications of the identified island region. The results demonstrated that there is a shared pathogenicity island architecture among tdh/trh positive pathogens of several sequence types including, ST36, ST631, ST1110, ST674, ST34, ST324, ST674 and ST749. The results showed that the pathogenicity island architecture was not present in non-pathogenic strains, and strains that lack the island, as shown by the lack of the product in the four right-hand lanes of the gel.

The primers used in the PCR amplification are the primers listed in Table 2 as TdhUreG. The forward primer, SEQ ID NO:100, anneals to *tdh* gene, the reverse primer, SEQ ID NO: 101, anneals to *ureG*, and this unique juncture is the pathogenicity island. The region amplified is the region is this unique island. The reaction conditions are also provided in Example 1 and 2 herein with additional detail in Table 2. PCR amplifications were performed with AccuStart PCR Supermix (Quanta, Gaithersburg, MD, US) with an initial denaturation at 94° C for 3 minutes, followed by 30 cycles with a denaturation at 94° C for 1 minute, primer annealing at 55° C for 1 minute, and extension at 72° C for 3 minutes, and with a final extension at 72° C at the completion of the cycling for 5 minutes.

Amplicons from the amplifications were evaluated using standard procedures, for example: by electrophoresis of 1.5 µl of sample on 1.2% SeaKem LE agarose (Lonza, Rockland, ME, USA) gel with 1× GelRed (Phenix Research Products, Candler, NC, USA) in Tris-acetate-EDTA (TAE) buffer compared against a 1-kb Plus DNA ladder (Invitrogen, Grand Island, NY, USA).

In the gel shown in Fig. 7, the actual strain number is listed first, and then the sequence type is in brackets. The presence of the "E" after the sequence type indicates that the sample was obtained from the environment such as an oyster. ST324 is a very close relative of ST34, and both are pathogens. ST674E is a pathogen type, ST749 is a clinical strain and also pathogen type; and ST110 is a pathogen. The last four lanes are strains that don't have the island and they are also negative. These last four lanes demonstrated that the strains that were not expected to amplify, did not amplify

Equivalents

Although several embodiments of the present invention have been described and illustrated herein, those of ordinary skill in the art will readily envision a variety of other means and/or structures for performing the functions and/or obtaining the results and/or one

or more of the advantages described herein, and each of such variations and/or modifications is deemed to be within the scope of the present invention. More generally, those skilled in the art will readily appreciate that all parameters, dimensions, materials, and configurations described herein are meant to be exemplary and that the actual parameters, dimensions, materials, and/or configurations will depend upon the specific application or applications for which the teachings of the present invention is/are used. Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. It is, therefore, to be understood that the foregoing embodiments are presented by way of example only and that, within the scope of the appended claims and equivalents thereto; the invention may be practiced otherwise than as specifically described and claimed. The present invention is directed to each individual feature, system, article, material, and/or method described herein. In addition, any combination of two or more such features, systems, articles, materials, and/or methods, if such features, systems, articles, materials, and/or methods are not mutually inconsistent, is included within the scope of the present invention.

All definitions, as defined and used herein, should be understood to control over dictionary definitions, definitions in documents incorporated by reference, and/or ordinary meanings of the defined terms.

The indefinite articles “a” and “an,” as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean “at least one.”

The phrase “and/or,” as used herein in the specification and in the claims, should be understood to mean “either or both” of the elements so conjoined, i.e., elements that are conjunctively present in some cases and disjunctively present in other cases. Other elements may optionally be present other than the elements specifically identified by the “and/or” clause, whether related or unrelated to those elements specifically identified, unless clearly indicated to the contrary.

All references, patents and patent applications and publications that are cited or referred to in this application are incorporated in their entirety herein by reference.

What is claimed is:

Claims

1. A method of identifying the status of a pathogenic *V. parahaemolyticus* in a sample, the method comprising:

(a) detecting in a sample, a level of at least one ST36*prp* polynucleotide comprising a nucleic acid sequence set forth as SEQ ID NO:1, 2, 3, 4, 5, 6, 7, 21, 22, 23, 24, 25 or 26, or a variant thereof;

(b) comparing the level of the detected ST36*prp* polynucleotide to a control level of the detected ST36*prp* polynucleotide; and

(c) identifying the status of the pathogenic *V. parahaemolyticus* in the sample, based at least in part on a difference between the level of the detected ST36*prp* polynucleotide and the control level of the detected ST36*prp* polynucleotide.

2. The method of claim 1, wherein a higher level of the detected ST36*prp* polynucleotide compared to the control level of the detected ST36*prp* polynucleotide identifies the status of the pathogenic *V. parahaemolyticus* as present in the sample.

3. The method of claim 1 or 2, wherein the control level of the detected ST36*prp* polynucleotide is zero.

4. The method of any one of the preceding claims, the method further comprising:

(d) detecting in the sample a level of at least one of a ST36*cps* polynucleotide comprising a nucleic acid sequence set forth as SEQ ID NO:8, 9, 10, 27 or 28, or a variant thereof;

(e) comparing the level of the detected ST36*cps* polynucleotide to a control level of the detected ST36*cps* polynucleotide; and

(f) identifying the status of the pathogenic *V. parahaemolyticus* in the sample, based at least in part on the difference between the level of the detected ST36*cps* polynucleotide and the control level of the detected ST36*cps* polynucleotide.

5. The method of claim 4, wherein a higher level of the detected ST36*cps* polynucleotide compared to the control level of the detected ST36*cps* polynucleotide identifies the status of the pathogenic *V. parahaemolyticus* as present in the sample.

6. The method of claim 4 or 5, wherein the control level of the detected ST36*cps* polynucleotide is zero.

7. The method of any one of the preceding claims, further comprising:

(g) detecting in the sample a level of at least one of a *tlh* polynucleotide comprising a nucleic acid sequence set forth as SEQ ID NO:11, 12, 13, 14, 29, 30 or 31, or a variant thereof; and

(h) comparing the level of the detected *tlh* polynucleotide to a control level of the detected *tlh* polynucleotide.

8. The method of claim 7, wherein a higher level of the detected *tlh* polynucleotide compared to the control level of the detected *tlh* polynucleotide aids in identifying the status of the pathogenic *V. parahaemolyticus* in the sample.

9. The method of claim 7 or 8, wherein the control level of the detected *tlh* polynucleotide is zero.

10. The method of any one of the preceding claims, further comprising:

(i) detecting in the sample a level of at least one of a *tdh* and *trh* polynucleotide comprising a nucleic acid sequence set forth as SEQ ID NO:15, 16, 17, 18, 19, 20, 32, 33, 34 or 35, or a variant thereof; and

(j) comparing the level of the detected polynucleotide to a control level of the detected polynucleotide.

11. The method of claim 10, wherein a higher level of the detected *tdh* or *trh* polynucleotide compared to the control level of the detected polynucleotide aids in identifying the status of the pathogenic *V. parahaemolyticus* in the sample.

12. The method of claim 10 or 11, wherein the control level of the detected *tdh* or *trh* polynucleotide is zero.

13. The method of any one of the preceding claims, further comprising:

(k) detecting in the sample a level of at least one of a ST36*flp* polynucleotide comprising a nucleic acid sequence set forth as SEQ ID NO:36, 37, 38, 39 or 40, or a variant thereof;

(l) comparing the level of the detected ST36*flp* polynucleotide to a control level of the detected ST36*flp* polynucleotide; and

(m) identifying the status of the pathogenic *V. parahaemolyticus* in the sample, based at least in part on difference between the level of the detected ST36*flp* polynucleotide and the control level of the detected ST36*flp* polynucleotide.

14. The method of claim 13, wherein a higher level of the detected ST36*flp* polynucleotide compared to the control level of the detected ST36*flp* polynucleotide aids in identifying the status of the pathogenic *V. parahaemolyticus* in the sample.

15. The method of claim 13 or 14, wherein the control level of the detected ST36*flp* polynucleotide is zero.

16. The method of any one of the preceding claims, wherein the variant of the SEQ ID NO:1, 2, 3, 4, 5, 6, 7, 21, 22, 23, 24, 25 or 26 nucleic acid sequence has at least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the nucleic acid sequence of SEQ ID NO:1, 2, 3, 4, 5, 6, 7, 21, 22, 23, 24, 25 or 26, respectively.

17. The method of claim 4, wherein the variant of the SEQ ID NO:8, 9, 10, 27 or 28 nucleic acid sequence has at least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the nucleic acid sequence of SEQ ID NO:8, 9, 10, 27 or 28, respectively.

18. The method of claim 7, wherein the variant of the SEQ ID NO:11, 12, 13, 14, 29, 30 or 31 nucleic acid sequence has at least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the nucleic acid sequence of SEQ ID NO:11, 12, 13, 14, 29, 30 or 31, respectively.

19. The method of claim 10, wherein the variant of the SEQ ID NO:15, 16, 17, 18, 19, 20, 32, 33, 34 or 35 nucleic acid sequence has at least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the nucleic acid sequence of SEQ ID NO:15, 16, 17, 18, 19, 20, 32, 33, 34 or 35, respectively.

20. The method of claim 13, wherein the variant of the SEQ ID NO:36, 37, 38, 39 or 40 nucleic acid sequence has at least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the nucleic acid sequence of SEQ ID NO:36, 37, 38, 39 or 40, respectively.

21. The method of any one of the preceding claims, wherein the variant of the nucleic acid sequence comprises a fragment of the nucleic acid sequence.

22. The method of any one of the preceding claims, wherein the variant of the nucleic acid sequence comprises a fragment of the nucleic acid sequence and the fragment has at least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the region of the nucleic acid sequence with which it aligns.

23. The method of any one of the preceding claims, wherein the sample is obtained from a subject.

24. The method of any one of claims 1-22, wherein the sample is obtained from a substrate.

25. A method of identifying the status of a pathogenic *V. parahaemolyticus* in a sample, the method comprising:

(a) detecting in a sample a level of at least one ST36*cps* polynucleotide comprising a nucleic acid sequence set forth as SEQ ID NO:8, 9, 10, 27 or 28, or a variant thereof;

(b) comparing the level of the detected ST36*cps* polynucleotide to a control level of the detected ST36*cps* polynucleotide; and

(c) identifying the status of a pathogenic *V. parahaemolyticus* in the sample, based at least in part on a difference in the level of the detected ST36*cps* polynucleotide and the control level of the detected ST36*cps* polynucleotide.

26. The method of claim 25, wherein a higher level of the detected ST36*cps* polynucleotide compared to the control level of the detected ST36*cps* polynucleotide identifies the status of the pathogenic *V. parahaemolyticus* as present in the sample.

27. The method of claim 25 or 26, wherein the control level of the detected ST36*cps* polynucleotide is zero.

28. The method of claim 25, further comprising detecting in the sample a level of at least one of a polynucleotide comprising a ST36*prp* nucleic acid sequence or variant thereof, a *tlh* nucleic acid sequence or variant thereof, a *tdh* nucleic acid sequence or variant thereof, a *trh* nucleic acid sequence or variant thereof, and an ST36*flp* nucleic acid sequence or variant thereof.

29. The method of any one of claims 25-28, wherein the variant of the SEQ ID NO: 8, 9, 10, 27, or 28 nucleic acid sequence has at least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the nucleic acid sequence of SEQ ID NO:8, 9, 10, 27 or 28, respectively.

30. The method of any one of claims 25-29, wherein the variant of the nucleic acid sequence comprises a fragment of the nucleic acid sequence.

31. The method of any one of claims 25-30, wherein the variant of the nucleic acid sequence comprises a fragment of the nucleic acid sequence and the fragment has at least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the region of the nucleic acid sequence with which it aligns.

32. The method of any one of claims 25-31, wherein the sample is obtained from a subject.

33. The method of any one of claims 25-31, wherein the sample is obtained from a substrate.

34. A method of identifying the presence of a pathogenic *V. parahaemolyticus* in a sample, the method comprising:

(a) detecting in a sample a level of at least one ST36*flp* polynucleotide comprising a nucleic acid sequence set forth as SEQ ID NO: 36, 37, 38, 39 or 40, or a variant thereof;

(b) comparing the level of the detected ST36*flp* polynucleotide to a control level of the detected ST36*flp* polynucleotide; and

(c) identifying the status of a pathogenic *V. parahaemolyticus* in the sample, based at least in part on the difference between the level of the detected ST36*flp* polynucleotide and the control level of the detected ST36*flp* polynucleotide.

35. The method of claim 34, wherein a higher level of the detected ST36*flp* polynucleotide compared to the control level of the detected ST36*flp* polynucleotide identifies the status of the pathogenic *V. parahaemolyticus* as present in the sample.

36. The method of claim 34 or 35, wherein the control level of the detected ST36*flp* polynucleotide is zero.

37. The method of claim 34, further comprising detecting in the sample a level of at least one of a polynucleotide comprising a ST36*prp* nucleic acid sequence or variant thereof, a *tlh* nucleic acid sequence or variant thereof, a *tdh* nucleic acid sequence or variant thereof, a *trh* nucleic acid sequence or variant thereof, and an ST36*cps* nucleic acid sequence or variant thereof.

38. The method of any one of claims 34-37, wherein the variant of the SEQ ID NO:36, 37, 38, 39 or 40 nucleic acid sequence has at least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the nucleic acid sequence of SEQ ID NO:36, 37, 38, 39 or 40, respectively.

39. The method of any one of claims 34-38, wherein the variant of the nucleic acid sequence comprises a fragment of the nucleic acid sequence.

40. The method of any one of claims 34-39, wherein the variant of the nucleic acid sequence comprises a fragment of the nucleic acid sequence and the fragment has at least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the region of the nucleic acid sequence with which it aligns.

41. The method of any one of claims 34-40, wherein the sample is obtained from a subject.

42. The method of any one of claims 34-40, wherein the sample is obtained from a substrate.

43. A method of assaying a sample, the method comprising:

(a) detecting in a sample determined to have at least one tdh and trh polynucleotide comprising a nucleic acid sequence set forth as SEQ ID NO: 15, 16, 17, 18, 19, 20, 32, 33, 34, 35, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120 or 121, or a variant thereof, and a level of at least one: (i) ST36*prp* polynucleotide comprising at least one of the nucleic acid sequences set forth as SEQ ID NO:1, 2, 3, 4, 5, 6, 7, 21, 22, 23, 24, 25, 26, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138 or 139, or a variant thereof; (ii) ST36*flp* polynucleotide comprising at least one of the nucleic acid sequences set forth as SEQ ID NO: 36, 37, 38, 39, 40, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149 or 150, or a variant thereof; (iii) ST631*end* polynucleotide comprising at least one of the nucleic acid sequences set forth as: SEQ ID NO: 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162 or 163, or a variant thereof; and (iv) 631-ENV polynucleotide comprising at least one of the nucleic acid sequences set forth as: SEQ ID NO: 59, 60, 61, 62, 170, 171, 172, 173, 174 or 175, or a variant thereof; and

(b) comparing the level of the detected polynucleotide in at least one of steps (i), (ii), (iii), and (iv) to a control level of the detected polynucleotide in steps (i), (ii), (iii), and (iv), respectively.

44. The method of claim 43, further comprising:

(c) determining the presence of a difference between the detected polynucleotide level and the control level of the detected polynucleotide as compared in (b); and

(d) identifying the presence or absence of the pathogenic *V. parahaemolyticus* in the sample, based at least in part on difference between the level of the detected polynucleotide of at least one of steps (i), (ii), (iii), and (iii) and the respective control level of the detected polynucleotide.

45. The method of claim 43 or 44, wherein assaying the sample comprises identifying the presence of a pathogenic *V. parahaemolyticus* in a sample.

46. The method of any one of claims 43-45, further comprising detecting an additional amplification control polynucleotide in the sample.

47. The method of any one of claims 43-46, wherein the nucleic acid sequence of the variant of the tdh, trh, ST36*prp*, ST36*fl,* ST631*end* and ST631-ENV polynucleotide has at

least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the nucleic acid sequence of the polynucleotide of which it is a variant thereof.

48. The method of any one of claims 43-47, wherein the variant of the nucleic acid sequence comprises a fragment of the nucleic acid sequence.

49. The method of any one of claims 43-48, wherein the variant of the nucleic acid sequence comprises a fragment of the nucleic acid sequence and the fragment has at least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the region of the nucleic acid sequence with which it aligns.

50. The method of any one of claims 43-49, wherein the sample is obtained from at subject.

51. The method of any one of claims 43-49, wherein the sample is obtained from a substrate.

52. A method of identifying the status of a pathogenic *V. parahaemolyticus* in a sample, the method comprising:

(a) detecting in a sample a level of at least one of (a) a ST36Phage polynucleotide comprising a nucleic acid sequence set forth as SEQ ID NO 81, 82, 83 or 84, or a or a variant thereof; (b) a ST36PhHypD-Orf9 polynucleotide comprising a nucleic acid sequence set forth as SEQ ID NO:94, 95, 96 or 97, or a variant thereof; (c) a ST36NEOrf10-Hyp polynucleotide comprising a nucleic acid sequence set forth as SEQ ID NO: 87, 88, 89 or 90, or a variant thereof; and (d) a tlh polynucleotide comprising a nucleic acid sequence set forth as SEQ ID NO: 12 or 14, or a variant thereof;

(b) comparing the level of the detected ST36Phage polynucleotide, ST36PhHypD-Orf9 polynucleotide, ST36NEOrf10-Hyp polynucleotide, and tlh polynucleotide to a control level of the detected ST36Phage polynucleotide, ST36PhHypD-Orf9 polynucleotide, ST36NEOrf10-Hyp polynucleotide, and tlh polynucleotide, respectively; and

(c) identifying the status of a pathogenic *V. parahaemolyticus* in the sample, based at least in part on the difference between the level of the detected ST36Phage polynucleotide, ST36PhHypD-Orf9 polynucleotide, ST36NEOrf10-Hyp polynucleotide, and tlh

polynucleotide, and the control level of ST36Phage polynucleotide, ST36PhHypD-Orf9 polynucleotide, ST36NEOrf10-Hyp polynucleotide, and tlh polynucleotide, respectively.

53. The method of claim 52, wherein a higher level of the detected ST36Phage polynucleotide compared to the control level of the detected ST36Phage polynucleotide and the presence of a higher level of one or more of the ST36PhHypD-Orf9 polynucleotide and the ST36NEOrf10-Hyp polynucleotide identifies the status of the pathogenic *V. parahaemolyticus* as present in the sample.

54. The method of claim 52 or 53, wherein the control level of one or more of the detected ST36Phage polynucleotide, ST36PhHypD-Orf9 polynucleotide, and the ST36NEOrf10-Hyp polynucleotide is zero.

55. The method of claim 52, further comprising detecting in the sample a level of at least one of a polynucleotide comprising a ST36*prp* nucleic acid sequence or variant thereof, a *tlh* nucleic acid sequence or variant thereof, a *tdh* nucleic acid sequence or variant thereof, a *trh* nucleic acid sequence or variant thereof, and an ST36*cps* nucleic acid sequence or variant thereof.

56. The method of any one of claims 52-55, wherein the nucleic acid sequence of the variant of the SEQ ID NO: 81, 82, 83, 84, 94, 95, 96, 97, 87, 88, 89, 90, 12 or 14 nucleic acid sequence has at least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the nucleic acid sequence of which it is a variant.

57. The method of any one of claims 52-56, wherein the variant of the nucleic acid sequence comprises a fragment of the nucleic acid sequence.

58. The method of any one of claims 52-57, wherein the variant of the nucleic acid sequence comprises a fragment of the nucleic acid sequence and the fragment has at least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the region of the nucleic acid sequence with which it aligns.

59. The method of any one of claims 52-58, wherein the sample is obtained from a subject.

60. The method of any one of claims 52-58, wherein the sample is obtained from a substrate.

61. A method of identifying the status of a pathogenic *V. parahaemolyticus* in a sample, the method comprising:

(a) detecting in the sample, the level of at least one ST631*end* polynucleotide comprising at least one nucleic acid sequence set forth as SEQ ID NO: 53, 54, 55, 56, 57, 58, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162 or 163, or a variant thereof;

(b) comparing the level of the detected ST631*end* polynucleotide to a control level of the detected ST631*end* polynucleotide; and

(c) identifying the presence of the pathogenic *V. parahaemolyticus* in the sample, based at least in part on a difference between the level of the detected ST631*end* polynucleotide and the control level of the detected ST631*end* polynucleotide.

62. The method of claim 61, wherein a higher level of the detected ST631*end* polynucleotide compared to the control level of the detected ST631*end* polynucleotide identifies the status of the pathogenic *V. parahaemolyticus* as present in the sample.

63. The method of claim 61 or 62, wherein the control level of the detected ST631*end* polynucleotide is zero.

64. The method of claim 61, further comprising detecting in the sample a level of at least one of a polynucleotide comprising an ST631-ENV nucleic acid sequence or variant thereof, an ST34*reg* nucleic acid sequence or variant thereof, a ST674*hyp* nucleic acid sequence or variant thereof, a ST1127*hyp* nucleic acid sequence or variant thereof, a ST36Phage nucleic acid sequence or variant thereof, an ST36NEOrf10-Hyp nucleic acid sequence or variant thereof, a ST36PhHypD-Orf9 nucleic acid sequence or variant thereof, a TdhUreG nucleic acid sequence or variant thereof, a ST36*cps* nucleic acid sequence or variant thereof, a ST36*prp* nucleic acid sequence or variant thereof, a *tlh* nucleic acid sequence or variant thereof, a *tdh* nucleic acid sequence or variant thereof, a *trh* nucleic acid sequence or variant thereof, a ST36*flp* nucleic acid sequence or variant thereof, and an ORF8 nucleic acid sequence or variant thereof.

65. The method of any one of claims 61-64, wherein the nucleic acid sequence of the variant of the ST631*end,* ST631-ENV, ST34*reg*, ST674*hyp*, ST1127*hyp*, ST36Phage, ST36NEOrf1O-Hyp, ST36PhHypD-Orf9, TdhUreG, ST36*cps*, ST36*prp*, *tlh, tdh*, *trh*, ST36*flp*, ORF8 polynucleotide has at least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the nucleic acid sequence of the polynucleotide of which it is a variant thereof.

66. The method of any one of claims 61-65, wherein the variant of the nucleic acid sequence comprises a fragment of the nucleic acid sequence.

67. The method of any one of claims 61-66, wherein the variant of the nucleic acid sequence comprises a fragment of the nucleic acid sequence and the fragment has at least 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the region of the nucleic acid sequence with which it aligns.

68. The method of any one of claims 61-67, wherein the sample is obtained from a subject.

69. The method of any one of claims 61-67, wherein the sample is obtained from a substrate.

70. The method of any one of claims 1-69, wherein detecting comprises performing a polymerase chain reaction (PCR) to detect the level of the least one polynucleotide comprising an ST36*prp* nucleic acid sequence or variant thereof, an ST36*cps* nucleic acid sequence or variant thereof, a ST36*flp* nucleic acid sequence or variant thereof, *ST631end* nucleic acid sequence or variant thereof, *ST631*-ENV nucleic acid sequence or variant thereof, a ST34*reg* nucleic acid sequence or variant thereof, a ST674*hyp* nucleic acid sequence or variant thereof, a ST1127*hyp* nucleic acid sequence or variant thereof, a ST36Phage nucleic acid sequence or variant thereof, a ST36NEOrf1O-Hyp nucleic acid sequence or variant thereof, a ST36PhHypD-Orf9 nucleic acid sequence or variant thereof, and a TdhUreG nucleic acid sequence or variant thereof.

71. The method of claim 70, wherein the PCR comprises qPCR.

72. The method of claim 71, wherein an oligonucleotide probe in the qPCR comprises a detectable label.

73. The method of claim 72, wherein the oligonucleotide probe comprises a nucleic acid sequence set forth herein as: SEQ ID NO: 127, 128, 129, 130, 143, 144, 155, 156 or 170, or a variant thereof, wherein the variant has an 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to the nucleic acid sequence of the polynucleotide of which it is a variant thereof.

74. The method of any one of claims 1-69, wherein detecting comprises performing in situ hybridization to measure the level of the at least one polynucleotide comprising an ST36*prp* nucleic acid sequence or variant thereof, an ST36*cps* nucleic acid sequence or variant thereof, a ST36*flp* nucleic acid sequence or variant thereof, *ST631end* nucleic acid sequence or variant thereof, *ST631*-ENV nucleic acid sequence or variant thereof, a ST34*reg* nucleic acid sequence or variant thereof, a ST674*hyp* nucleic acid sequence or variant thereof, a ST1127*hyp* nucleic acid sequence or variant thereof, a ST36Phage nucleic acid sequence or variant thereof, a ST36NEOrf1O-Hyp nucleic acid sequence or variant thereof, a ST36PhHypD-Orf9 nucleic acid sequence or variant thereof, and a TdhUreG nucleic acid sequence or variant thereof.

75. The method of any one of claims 1-69, wherein detecting comprises performing DNA hybridization to measure the level of the at least one polynucleotide comprising a ST36*prp* nucleic acid sequence or variant thereof, an ST36*cps* nucleic acid sequence or variant thereof, a ST36*flp* nucleic acid sequence or variant thereof, *ST631end* nucleic acid sequence or variant thereof, *ST631*-ENV nucleic acid sequence or variant thereof, a ST34*reg* nucleic acid sequence or variant thereof, a ST674*hyp* nucleic acid sequence or variant thereof, a ST1127*hyp* nucleic acid sequence or variant thereof, a ST36Phage nucleic acid sequence or variant thereof, a ST36NEOrf1O-Hyp nucleic acid sequence or variant thereof, a ST36PhHypD-Orf9 nucleic acid sequence or variant thereof, and a TdhUreG nucleic acid sequence or variant thereof.

76. The method of claim 75, wherein the DNA hybridization is DNA microarray hybridization.

77. The method of any one of claims 23, 32, 41, 50, 59 and 68, further comprising one or more of selecting a therapeutic agent to reduce a pathogenic *V. parahaemolyticus* infection in the subject; and administering a therapeutic agent to treat a pathogenic *V. parahaemolyticus* infection in the subject, based at least in part on the identified status of the pathogenic *V. parahaemolyticus* in the sample.

78. The method of any one of claims 24, 33, 42, 51, 60 and 69, further comprising one or more of selecting an agent to reduce a pathogenic *V. parahaemolyticus* contamination of the substrate; and contacting the substrate with an agent selected to reduce a pathogenic *V. parahaemolyticus* contamination of the substrate, based at least in part on the identified status of the pathogenic *V. parahaemolyticus* in the sample obtained from the substrate.

79. The method of any one of one of claims 1-69, wherein the presence of the pathogenic *V. parahaemolyticus* in the sample comprises the presence of one or more pathogenic *V. parahaemolyticus* bacteria in the sample.

80. The method of claim 77 or 78 wherein the agent is an anti-*V. parahaemolyticus* agent.

81. The method of any one of claims 23, 32, 41, 50, 59 and 68, wherein the subject is at least one of: suspected of having a pathogenic *V. parahaemolyticus* infection and diagnosed with a pathogenic *V. parahaemolyticus* infection.

82. The method of any one of claims 23, 32, 41, 50, 59 and 68, wherein the subject is a vertebrate, and optionally is a mammal.

83. The method of any one of claims 23, 32, 41, 50, 59 and 68, wherein the subject is an invertebrate.

84. The method of any one of claims 23, 32, 41, 50, 59 and 68, wherein the sample comprises one or more of a skin sample, fluid sample, tissue sample, stool sample, pus sample, gastric sample, emesis sample, inflammatory exudate sample, blood sample, or lymph sample.

85. The method of any one of claims 1-69, wherein the sample is a cultured sample.

86. The method of any one of claims 24, 33, 42, 51, 60 and 69, wherein the substrate is at least one of: known to have been exposed to a *V. parahaemolyticus* bacteria, suspected of being contaminated with a *V. parahaemolyticus* bacteria; and identified as being contaminated with a *V. parahaemolyticus* bacteria.

87. The method of claim 78, wherein reducing the pathogenic *V. parahaemolyticus* contamination of the substrate comprises reducing the level of the pathogenic *V. parahaemolyticus* bacteria on the substrate.

88. The method of any one of claims 24, 33, 42, 51, 60 and 69, wherein the substrate comprises a liquid and the liquid optionally comprises water.

89. The method of any one of claims 24, 33, 42, 51, 60 and 69, wherein the substrate comprises one or more of a metal, wood, plastic, glass, cork, fiber, a polymer, or a fabric.

90. The method of any one of claims 24, 33, 42, 51, 60 and 69, wherein the substrate comprises a food substance, and wherein the food substance optionally comprises shellfish.

91. The method of any one of claims 24, 33, 42, 51, 60 and 69, wherein the substrate comprises at least a portion of a tool, work surface, a medical device, body of water, clothing, skin, tissue, an edible substance, a beverage, or a food.

92. The method of any one of claims 24, 33, 42, 51, 60 and 69, wherein the substrate is a contaminated substrate and comprises one or more pathogenic *V. parahaemolyticus* bacteria.

93. The method of any one of claims 24, 33, 42, 51, 60 and 69 wherein the sample comprises a fluid sample, semisolid sample, aqueous sample, or tissue sample.

94. A kit for performing the method of any one of claims 1-69, wherein the kit comprises at least one a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to at least one of SEQ ID NO: 2-7, 9, 10, 36, 37, 52, 53, 59, 60, 63, 64, 75, 76, 81, 82, 87, 88, 94, 95, 122-126, 140-142, 151-154, 170,

171 and 172; wherein the kit additionally comprises instructions for detecting a status of a *V. parahaemolyticus* polynucleotide in a sample.

95. The kit of claim 94, wherein the *V. parahaemolyticus* polynucleotide is a pathogenic *V. parahaemolyticus* polynucleotide.

96. The kit of claim 94 or 95, wherein the status comprises one or more of: the presence or absence of the polynucleotide; and a level of the polynucleotide.

97. The kit of any one of claims 94-96, wherein the kit comprises one or more of:

(a) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:2 and a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:3;

(b) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:4 and a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:5;

(c) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:6 and a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:7;

(d) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:9 and a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:10;

(e) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:36 and a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:37;

(f) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:53 and a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:54;

(g) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:59 and a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:60;

(h) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:63 and a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:64;

(i) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:69 and a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:70;

(j) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:75 and a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:76;

(k) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:122 and a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:123;

(l) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:124 and a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:125;

(m) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:126;

(n) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:140 and one or more of a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:141, and a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:142;

(o) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:151 and a container

comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:152;

(p) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:153 and a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:154; and

(q) a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:171 and a container comprising an oligonucleotide primer the sequence of which has at least 90%, 95%, 99%, or 100% sequence identity to SEQ ID NO:172.

98. The kit of anyone of claims 94-97, further comprising a oligonucleotide probe.

99. The kit of any one of claims 94-98, wherein one or more of the oligonucleotide primers and probes is linked to one or more detectable labels.

100. The kit of claim 99, wherein the detectable label comprises one or more of a radioactive molecule, a luminescent molecule, a chemiluminescent molecule, biotin, an enzyme, a His tag, or an exogenous nucleic acid sequence.

101. The kit of claim 99, wherein the link is an indirect link.

102. The kit of any one of claims 99-101, wherein the detectable label comprises: Hexachloro-Fluorescein (HEX), VIC fluorescent dye, 4-5-Dichloro carboxy fluorescein (JOE), Cy3 fluorescent dye, or TexasRed (TxRed).

103. A composition comprising a polynucleotide having a sequence set forth as SEQ ID NO: 2-7, 9, 10, 12, 13, 21-30, 36, 37, 40, 53-56, 59, 60, 63-66, 69-72, 81-84, 87-91, 94-97, 100-103, 122-174 or 175, or a variant thereof.

104. The composition of claim 103, wherein the polynucleotide is linked to one or more detectable labels.

105. The composition of claim 103, wherein the detectable label comprises one or more of a radioactive molecule, a luminescent molecule, a chemiluminescent molecule, biotin, an enzyme, a His tag, or an exogenous nucleic acid sequence.

106. The composition of claim 105, wherein the link is an indirect link.

107. The composition of any one of claims 103-106, wherein the detectable label comprises: Hexachloro-Fluorescein (HEX), VIC fluorescent dye, 4-5-Dichloro carboxy fluorescein (JOE), Cy3 fluorescent dye, or TexasRed (TxRed).

108. A method of identifying a test compound as an agent to reduce one or both of pathogenic *V. parahaemolyticus* in subject and pathogenic *V. parahaemolyticus* contamination of a substrate, the method comprising,

(a) identifying the level of a pathogenic *V. parahaemolyticus* bacteria in a first portion of a sample using a method of any one of claims 1-69;

(b) contacting a second portion of the sample of (a) with a test compound;

(c) incubating the contacted second portion of the sample with the test compound;

(d) identifying the level of the pathogenic *V. parahaemolyticus* bacteria in the incubated second portion of the sample; and

(e) comparing the level of the pathogenic *V. parahaemolyticus* bacteria in the first portion of the sample to the level of the pathogenic *V. parahaemolyticus* bacteria in the incubated second portion of the sample, wherein a decrease in the level of the pathogenic *V. parahaemolyticus* bacteria in the incubated second portion of the sample compared to the control level identified in step (a) identifies the test compound as a candidate compound to treat a pathogenic *V. parahaemolyticus* infection in a subject or as a candidate compound to reduce a pathogenic *V. parahaemolyticus* contamination of a substrate.

109. The method of claim 108, wherein the incubation is at least 1 min, 5 min, 30 min, 1 hour, 6 hours, 12 hours, or 18 hours in duration.

110. The method of claim 108, further comprising confirming efficacy of the candidate compound for treating a subject having or at risk of having the pathogenic *V. parahaemolyticus* infection or for reducing contamination of a substrate by the pathogenic *V. parahaemolyticus* bacteria.

111. The method of claim 108, wherein the control level detected in step (a) is greater than zero.

112. The method of any one of claims 108-111, wherein detecting the level of the pathogenic *V. parahaemolyticus* bacteria in the sample comprises detecting in the sample, the level of a pathogenic *V. parahaemolyticus* polynucleotide of the pathogenic *V. parahaemolyticus* bacteria.

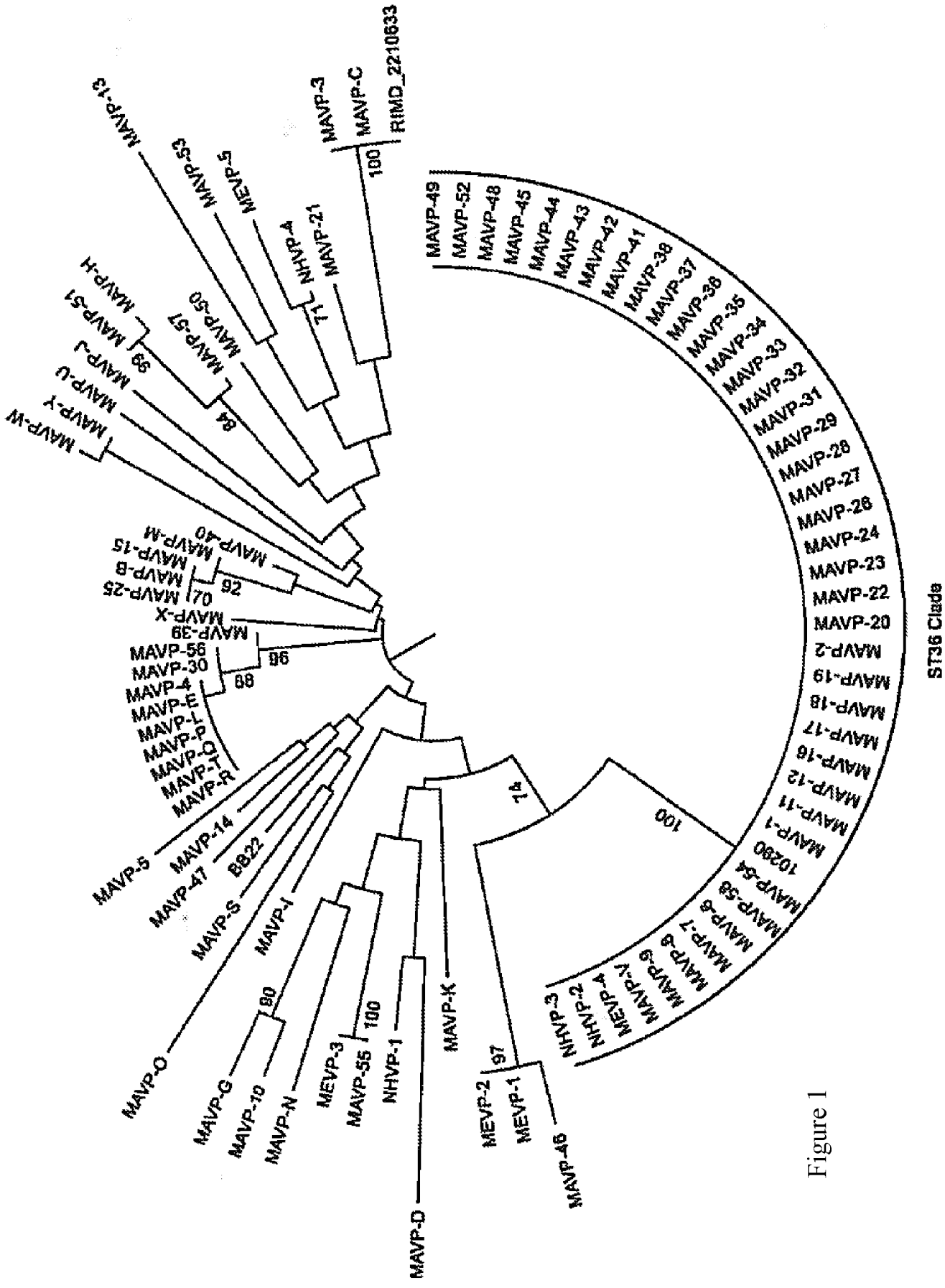

MAVP-49
MAVP-52
MAVP-48
MAVP-45
MAVP-44
MAVP-43
MAVP-42
MAVP-41
MAVP-38
MAVP-37
MAVP-36
MAVP-35
MAVP-34
MAVP-33
MAVP-32
MAVP-31
MAVP-29
MAVP-28
MAVP-27
MAVP-26
MAVP-24
MAVP-23
MAVP-22
MAVP-20
MAVP-2
MAVP-19
MAVP-18
MAVP-17
MAVP-16
MAVP-12
MAVP-11
MAVP-1
1029O
MAVP-54
MAVP-58
MAVP-6
MAVP-7
MAVP-8
MAVP-9
MAVP-V
MEVP-4
NHVP-2
NHVP-3
100
14
ST36 Clade
MEVP-2
MEVP-1
MAVP-46
97
MAVP-K
MAVP-D
NHVP-1
MEVP-3
MAVP-55
100
MAVP-I
MAVP-N
MAVP-10
MAVP-G
80
MAVP-O
MAVP-5
MAVP-14
MAVP-47
BB22
MAVP-S
MAVP-56
MAVP-30
MAVP-4
MAVP-E
MAVP-L
MAVP-P
MAVP-Q
MAVP-T
MAVP-R
88
96
MAVP-39
MAVP-X
MAVP-25
MAVP-B
MAVP-15
MAVP-40
MAVP-M
92
70
MAVP-W
MAVP-Y
MAVP-U
MAVP-J
MAVP-51
MAVP-H
99
MAVP-57
MAVP-50
84
MAVP-53
MEVP-5
MAVP-13
NHVP-4
MAVP-21
71
MAVP-3
MAVP-C
RIMD_2210633
100

Figure 1

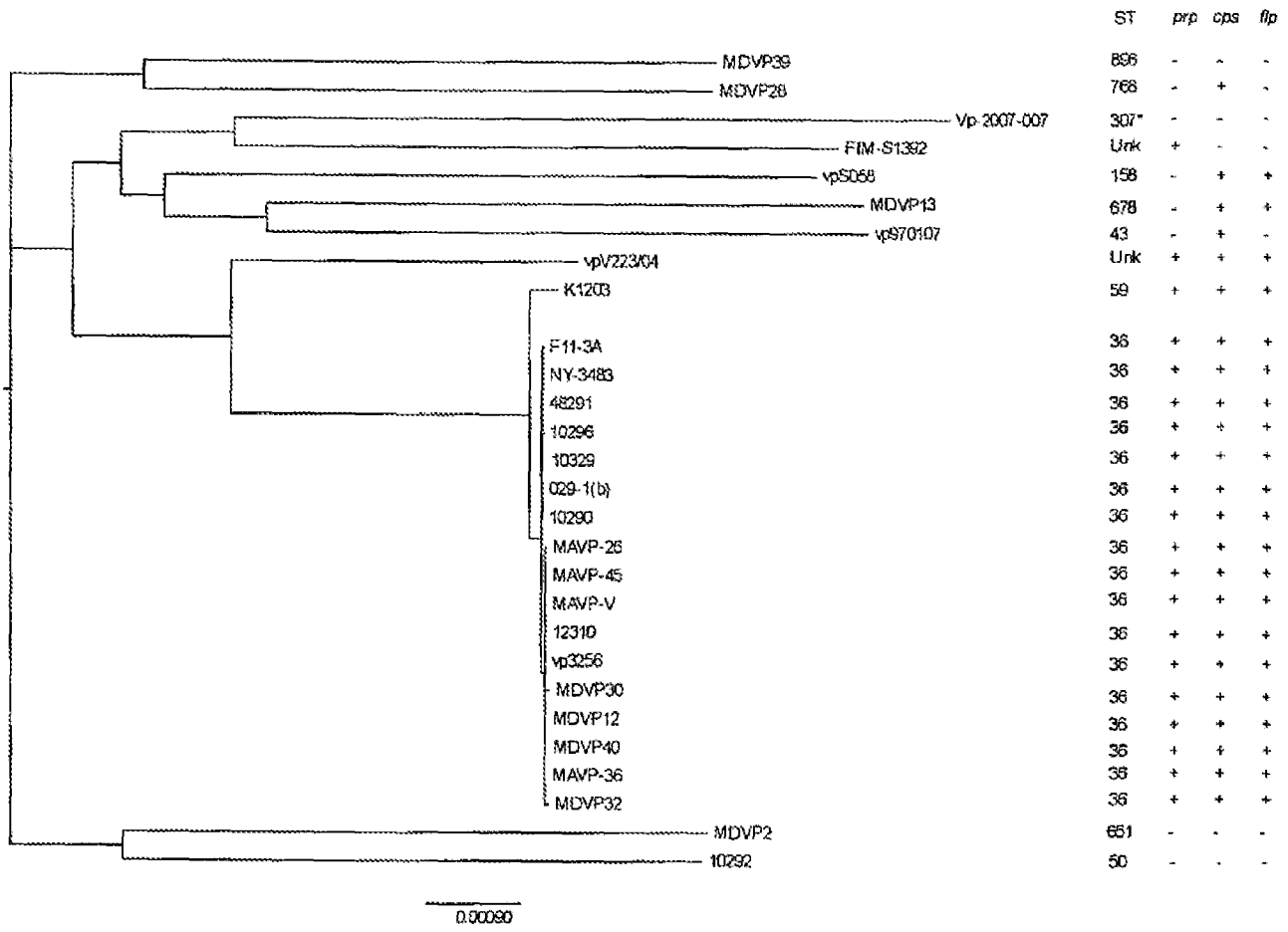


Figure 2

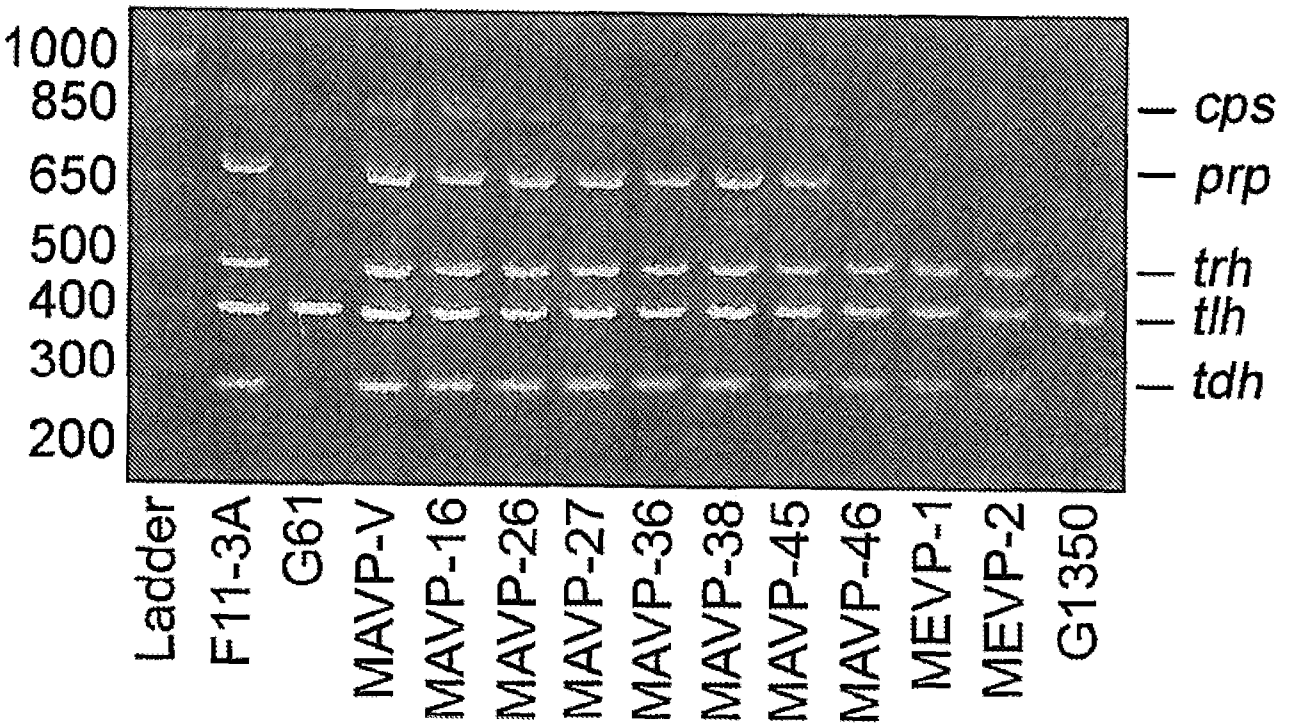


Figure 3

Distribution of diagnostic loci in draft genomes of *Vibrio parahaemolyticus*[a]

Strain	Genome Group	Sequence Type	*Prp*	*cps*	*flp*	Isolation Location[b]	Source[c]	Year[d]
vpV223/04	n/a	Unk	+	+	+	n/a	n/a	n/a
vpS038	10329	59	+	+	+	USA	E	1982
K1203	10329	59	+	+	+	AK	E	2004
K1198	10329	59	+	+	+	AK	E	2004
MDVP40	10329	36	+	+	+	MD	C	2013
MDVP30	10329	36	+	+	+	MD	C	2013
MDVP12	10329	36	+	+	+	MD	C	2012
MDVP32	10329	36	+	+	+	MD	C	2013
MDVP36	10329	36	+	+	+	MD	C	2013
MDVP38	10329	36	+	+	+	MD	C	2013
MDVP33	10329	36	+	+	+	MD	C	2013
MDVP42	10329	36	+	+	+	MD	C	2013
MDVP43	10329	36	+	+	+	MD	C	2013
MAVP-36	10329	36	+	+	+	MA	C	2013
MAVP-26	10329	36	+	+	+	MA	C	2013
MAVP-45	10329	36	+	+	+	MA	C	2013
MAVP-V	10329	36	+	+	+	MA	C	2011
12310	10329	36	+	+	+	WA	C	2016
vp3256	10329	36	+	+	+	USA	C	2007
F11-3A	10329	36	+	+	+	WA	E	1988
48291	10329	36	+	+	+	WA	C	1990
10296	10329	36	+	+	+	WA	C	1997
NY-3483	10329	36	+	+	+	NY	E	1998
029-1(b)	10329	36	+	+	+	OR	E	1997
10290	10329	36	+	+	+	WA	C	1997
48057	10329	36	+	+	+	WA	C	1990
10329	10329	36	+	+	+	WA	C	1998
CFSAN007462	10329	36	+	+	+	MD	C	2013
vpS037	10329	36	+	+	+	USA	C	1994
MDVP13	10329	678	-	+	+	MD	C	2012
vpS058	NIHCB0757	158	-	+	+	Japan	C	1970
Vp970107[e]	S159	43	-	+	-	USA	C	1997
MDVP28	S159	768	-	+	-	USA	E	2010
vpS048	S048	322	+	-	-	USA	E	1997
FIM-S1392	SNUVpS-1	Unk	+	-	-	Mexico	E	2014
10292	S129	50	-	-	-	WA	C	1997
MDVP2	S129	651	-	-	-	MD	C	2012
MDVP39	S129	896	-	-	-	MD	C	2013
VP2007-007	S100	307	-	-	-	USA	E	2007

Figure 4

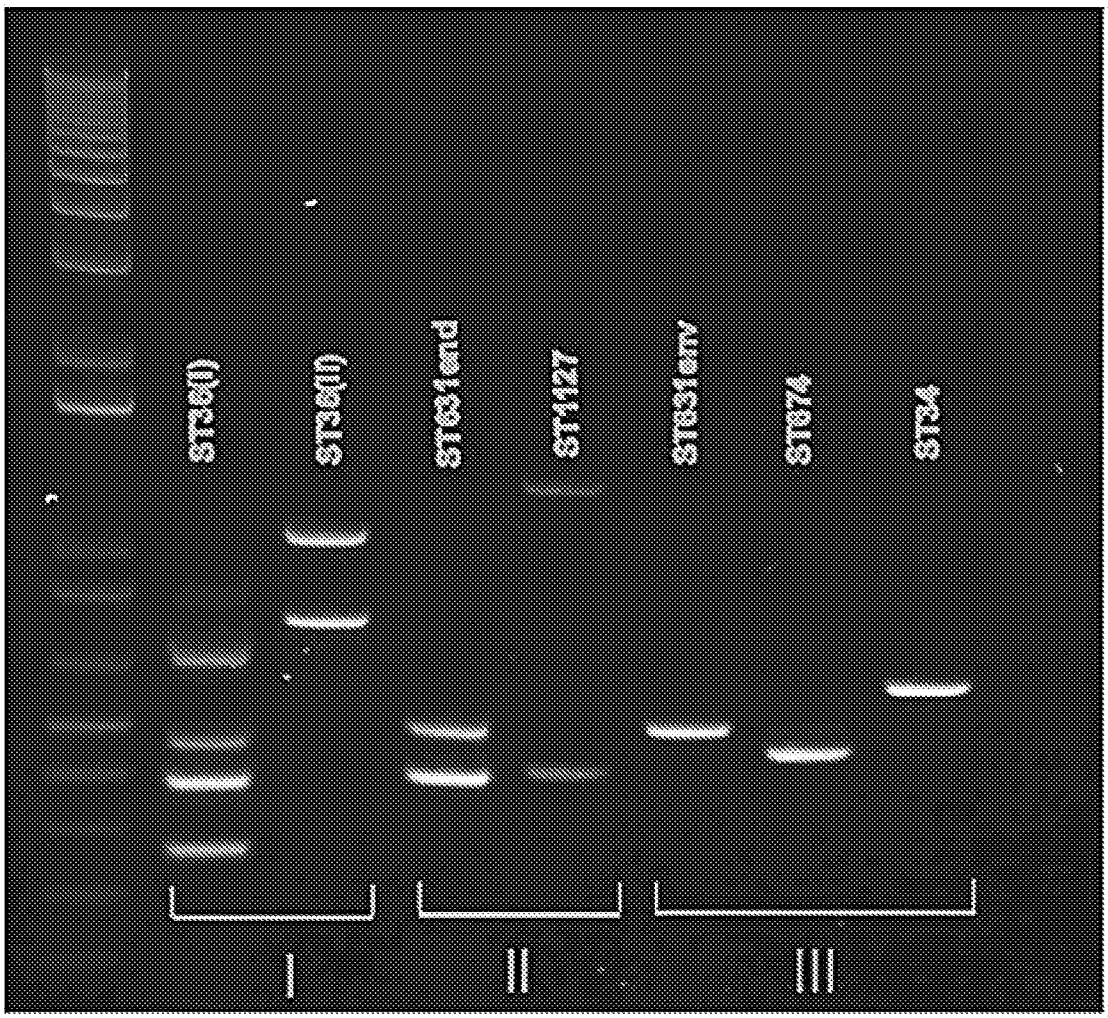
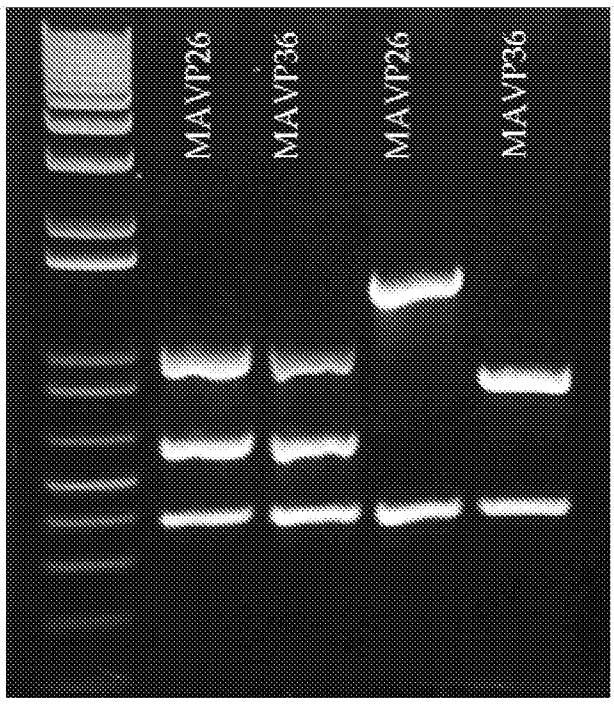


Figure 5



PCR reaction I — Any f237-like — NE phage — *tlh*

PCR reaction II — Vipa26 — Vipa36 — *tlh*

Figure 6

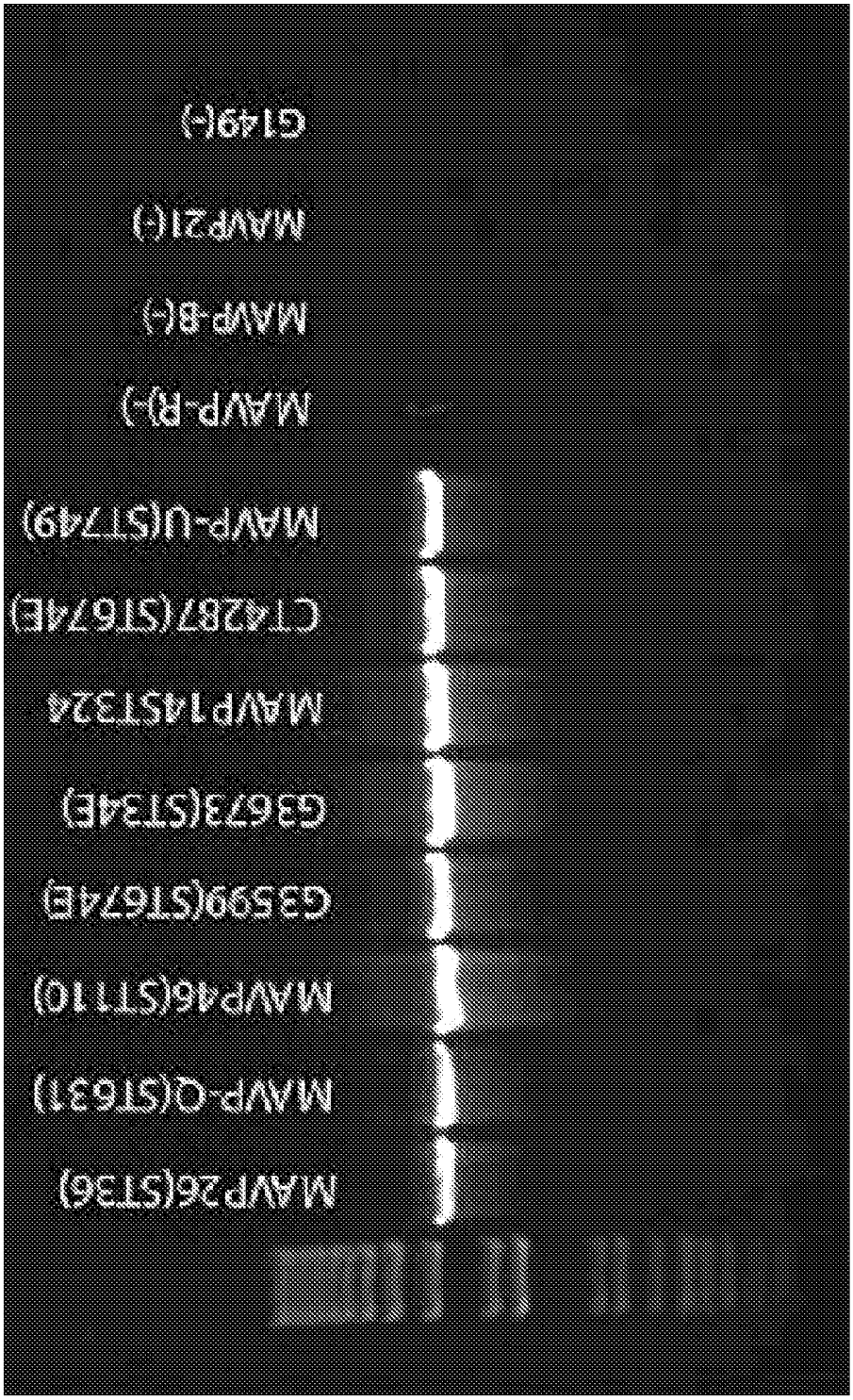

MAVP26(ST36)
MAVP-Q(ST631)
MAVP46(ST710)
G3599(ST674E)
G3673(ST34E)
MAVP145T324
CT4287(ST674E)
MAVP-U(ST749)
MAVP-R)-)
MAVP-8(-)
MAVP21(-)
G149(-)

Figure 7

ST	Annotation	Accession number
631	restriction endonuclease	EQM04338 (VP2007-095)
34	lysR family protein	ETT15213 (Vibrio parahaemolyticus 50, which is also ST34)
674	Hypothetical protein	WP_053056597
1127	Hypothetical protein	KU711831
36	Hypothetical protein	WP_047724020 Contig: NZ_LBHD01000006 locus tag YA91_RS18875
36	Orf9	WP_047724015 Contig: NZ_LBHD01000006 Locus tag YA91_RS18860
36	Orf10	WP_024701899 Contig: NZ_LBHD01000006 Locus tag YA91_RS18940
36	Hypothetical protein	WP_012842283 Contig: NZ_LBHD01000006 Locus tag YA91_RS18935
36	Orf3	WP_005477619 Contig: NZ_LBHD01000006 Locus tag YA91_RS18915
36	Orf5	WP_047715660 NZ_LBHE01000002.1 Locus tag YA90_RS06390
All above	UreG	EQL83620.1 (VP2007-095) AB831_22120 (MAVP-Q)
All above	tdh	EQL83622.1 (vp2007-095) AB831_22110 (MAVP-Q)

Figure 8