

## Supplementary Material

### **Complete genome sequence of *Vibrio gazogenes* PB1: An estuarine bacterium capable of producing prodigiosin from starch or cellulose**

***Bincy Baby<sup>1,2</sup>, Dhanya Vijay<sup>1</sup>, Pretty S. Philip<sup>3</sup>, Afra A. Alnuaimi<sup>1</sup>, Hessa M. Almansoori<sup>1</sup>, Sarah O. Areidat<sup>1</sup>, Gulfaraz Khan<sup>3,5</sup>, Ranjit Vijayan<sup>2,4,5\*</sup> and M. Kalim Akhtar<sup>1\*</sup>***

*<sup>1</sup>Department of Chemistry, College of Science, United Arab Emirates University, Al Ain, United Arab Emirates*

*<sup>2</sup>Department of Biology, College of Science, United Arab Emirates University, Al Ain, United Arab Emirates*

*<sup>3</sup>Department of Medical Microbiology and Immunology, College of Medicine and Health Sciences, United Arab Emirates University, Al Ain, United Arab Emirates*

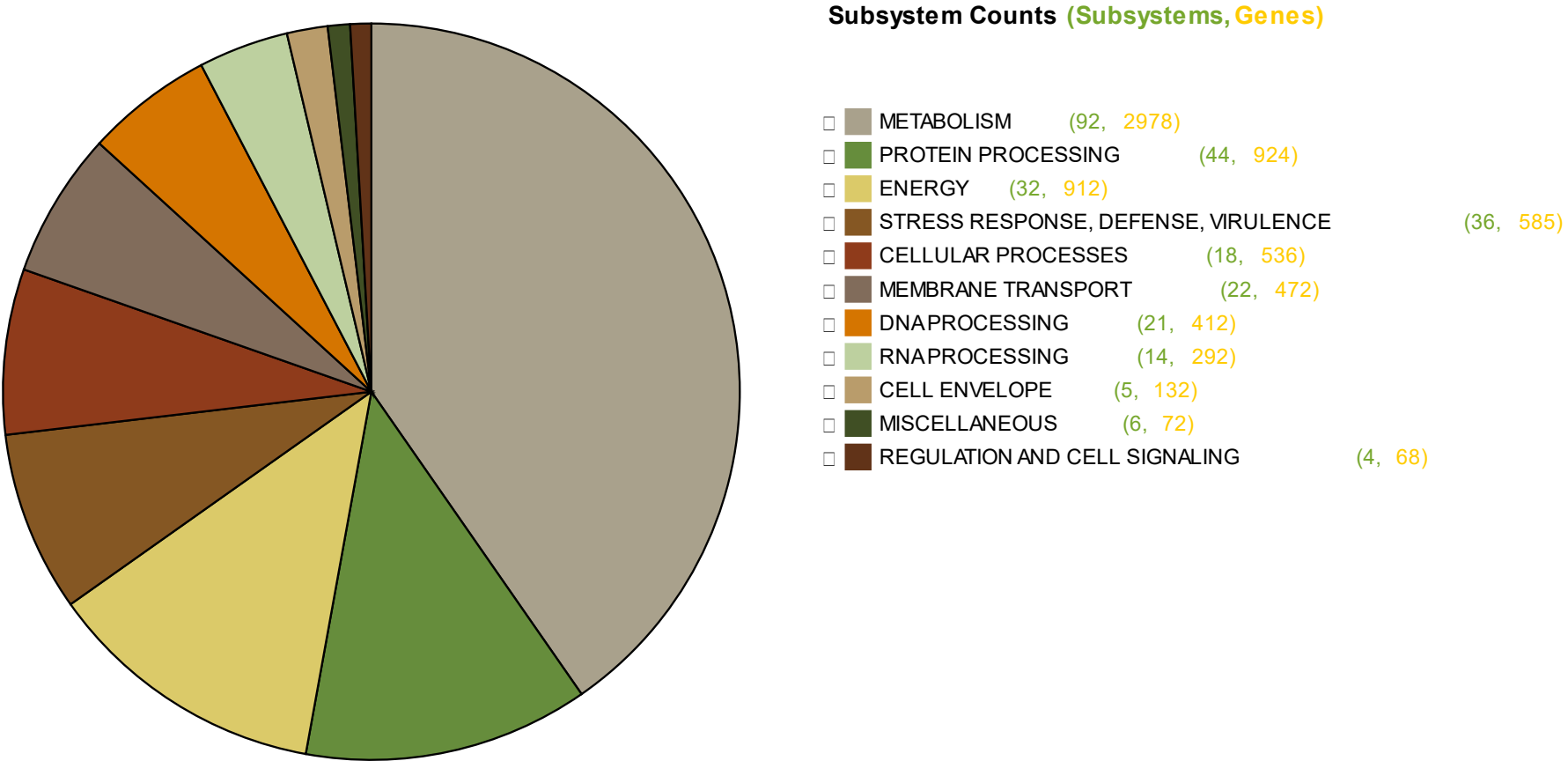
*<sup>4</sup>The Big Data Analytics Center, United Arab Emirates University, Al Ain, United Arab Emirates.*

*<sup>5</sup>Zayed Center for Health Sciences, United Arab Emirates University, Al Ain, United Arab Emirates.*

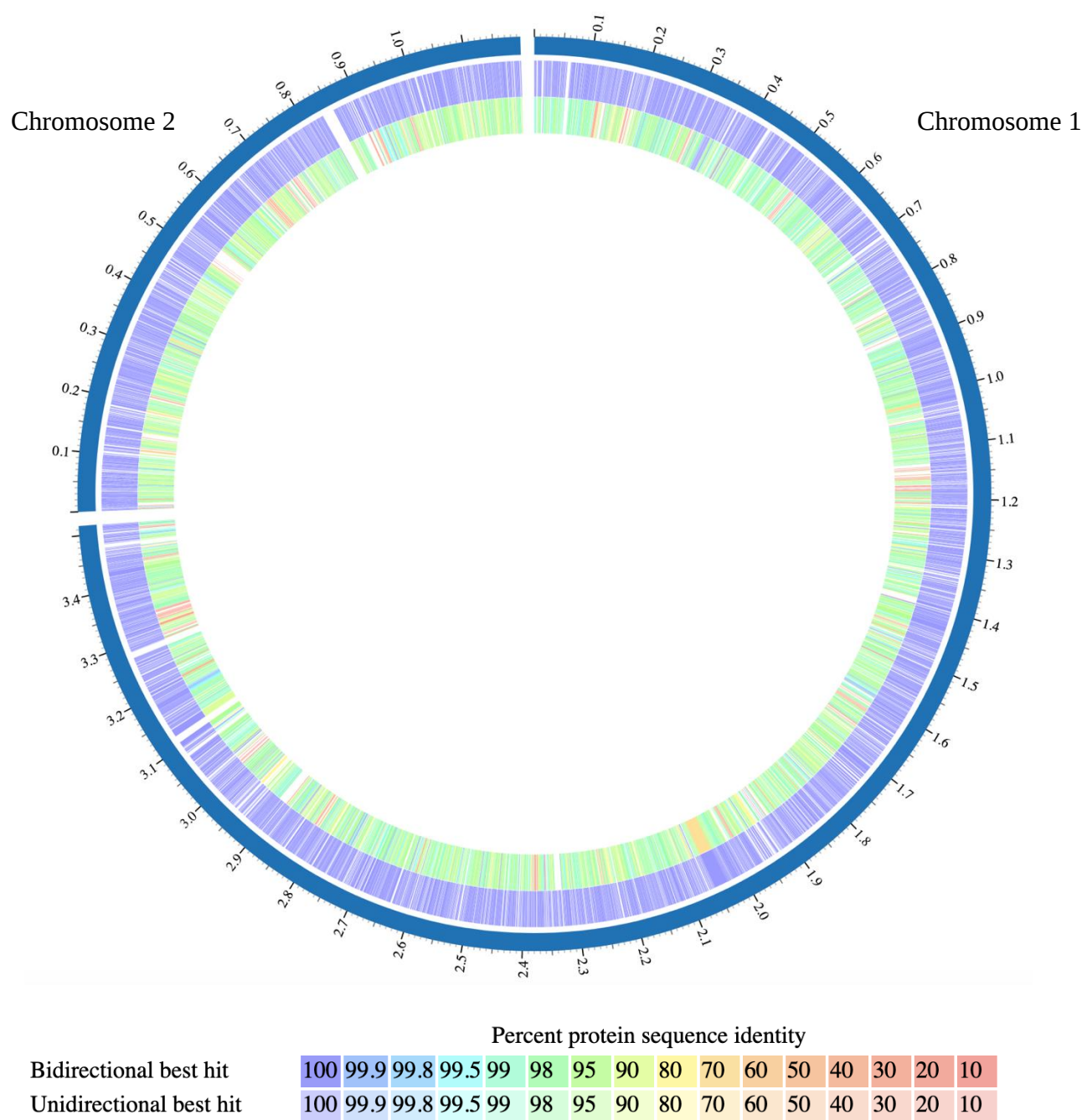
***\*Corresponding authors:***

*E-mails: [ranjit.v@uaeu.ac.ae](mailto:ranjit.v@uaeu.ac.ae), [mk.akhtar@uaeu.ac.ae](mailto:mk.akhtar@uaeu.ac.ae)*

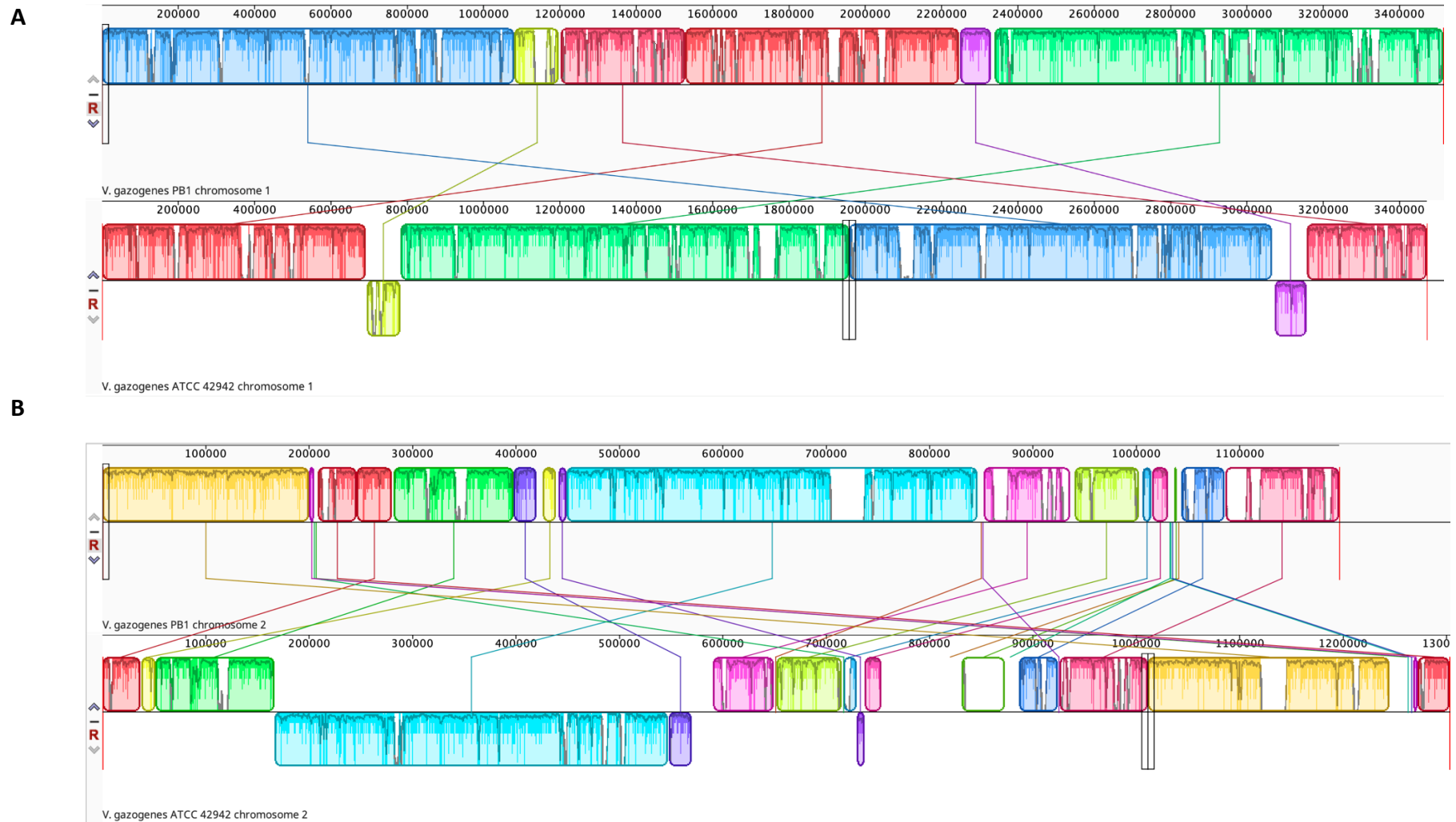
**Figure S1:** Subsystems identified in the *V. gazogenes* PB1 genome. The subsystem classification was produced by RAST annotation in PATRIC.



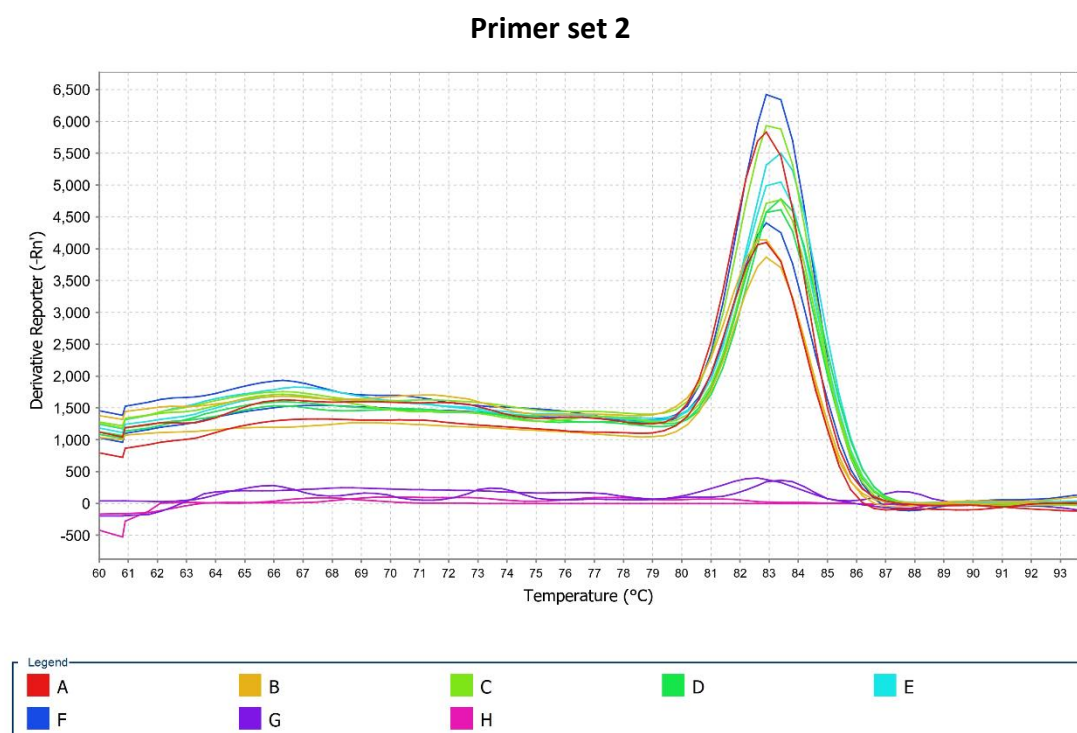
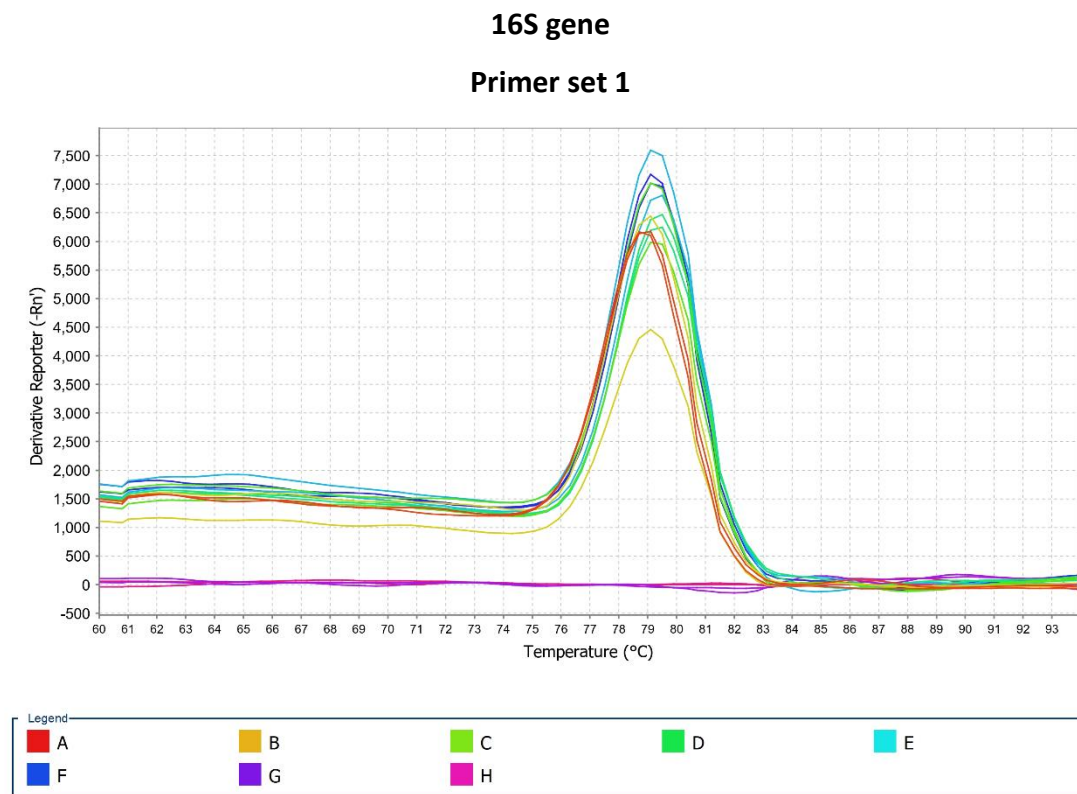
**Figure S2:** Comparison of the proteome of *V. gazogenes* PB1 and *V. gazogenes* ATCC 43942 using PATRIC. From outside to inside the tracks represent the genome size, including the two chromosomes, *V. gazogenes* PB1 proteins and the orthologous *V. gazogenes* ATCC 43942 protein.



**Figure S3:** Whole chromosome alignments of *V. gazogenes* PB1 and *V. gazogenes* ATCC 43942 **(A)** chromosome 1 and **(B)** chromosome 2 generated using progressive Mauve. Homologous blocks are shown in the same colour.

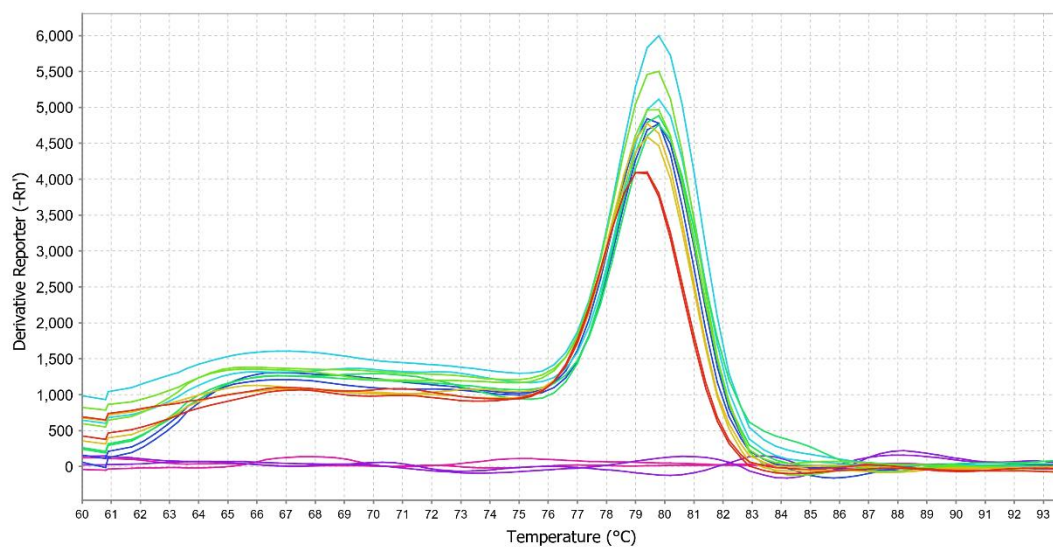


**Figure S4:** RT-qPCR melt curves for target genes related to the metabolism of starch and cellulose.

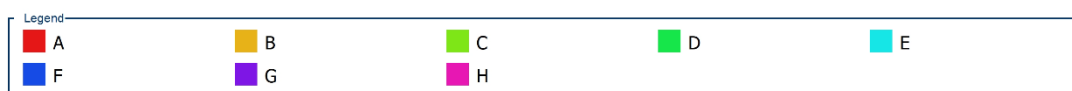
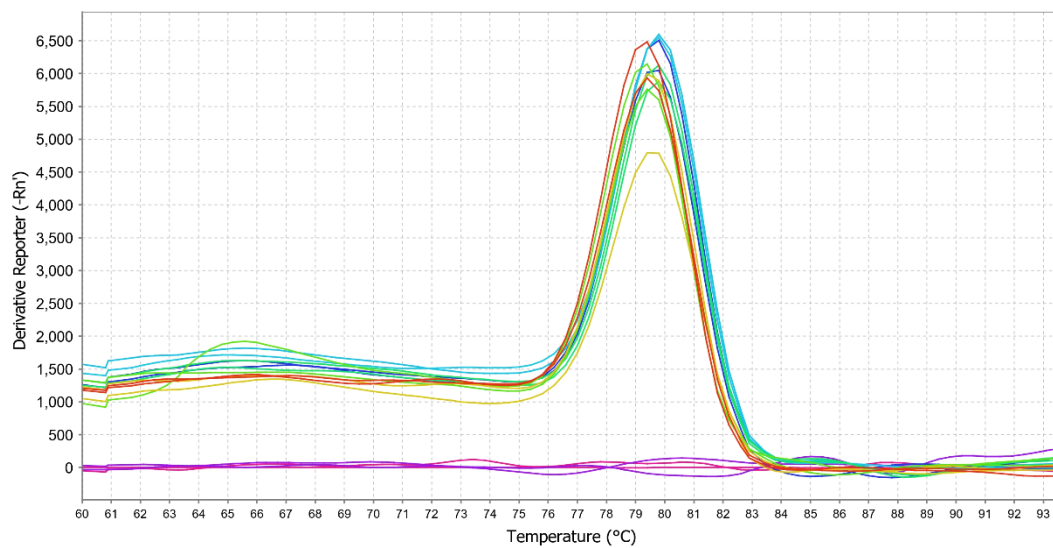


## USP13629.1 $\alpha$ -glucosidase

### Primer set 1

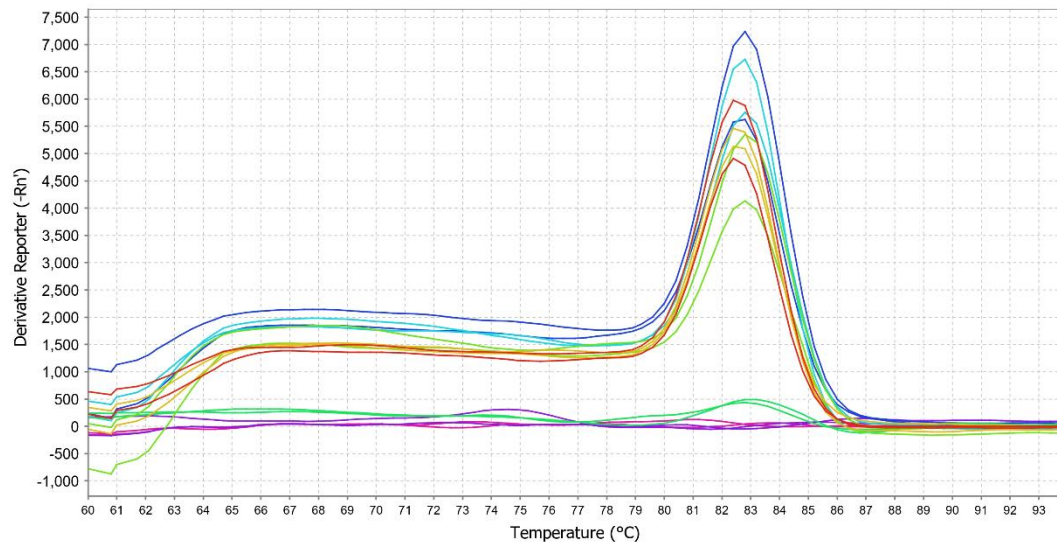


### Primer set 2

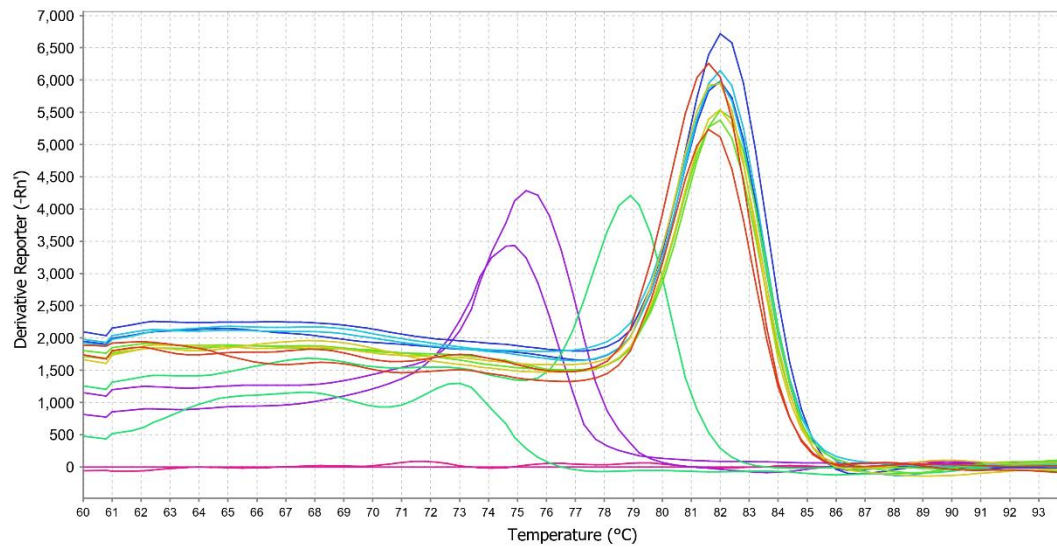


## USP14760.1 exoglucanase

### Primer set 1

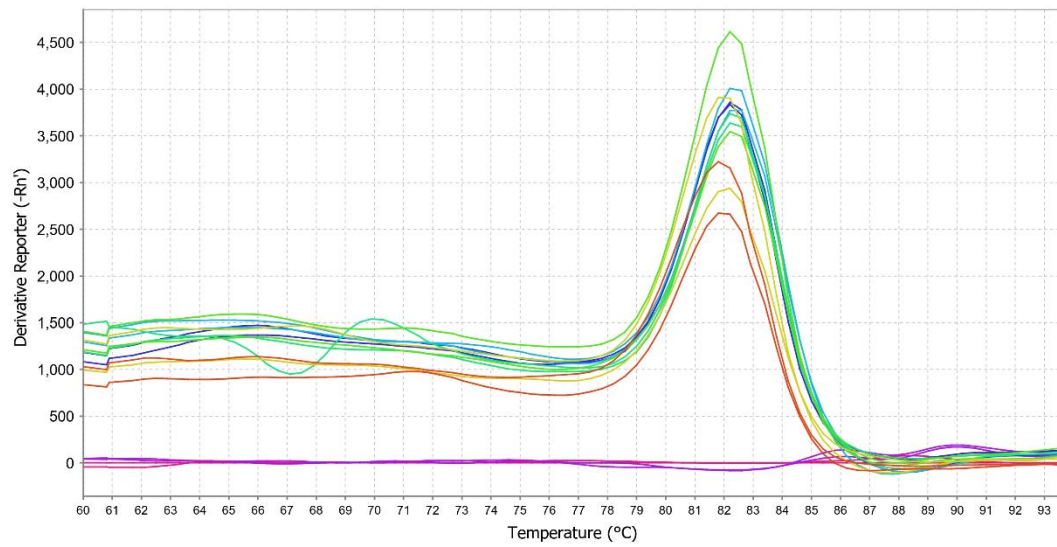


### Primer set 2

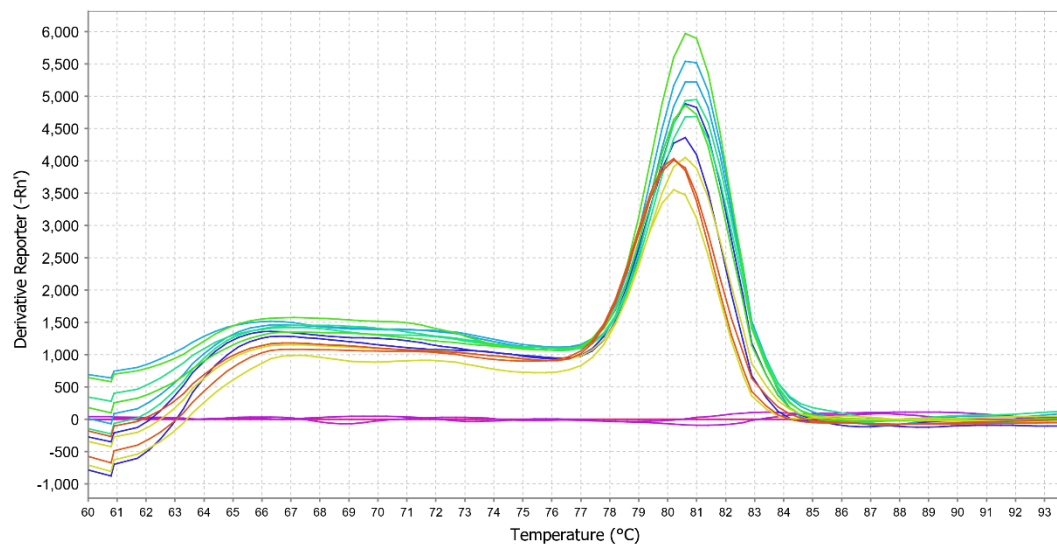


## USP16020.1 isoamylase

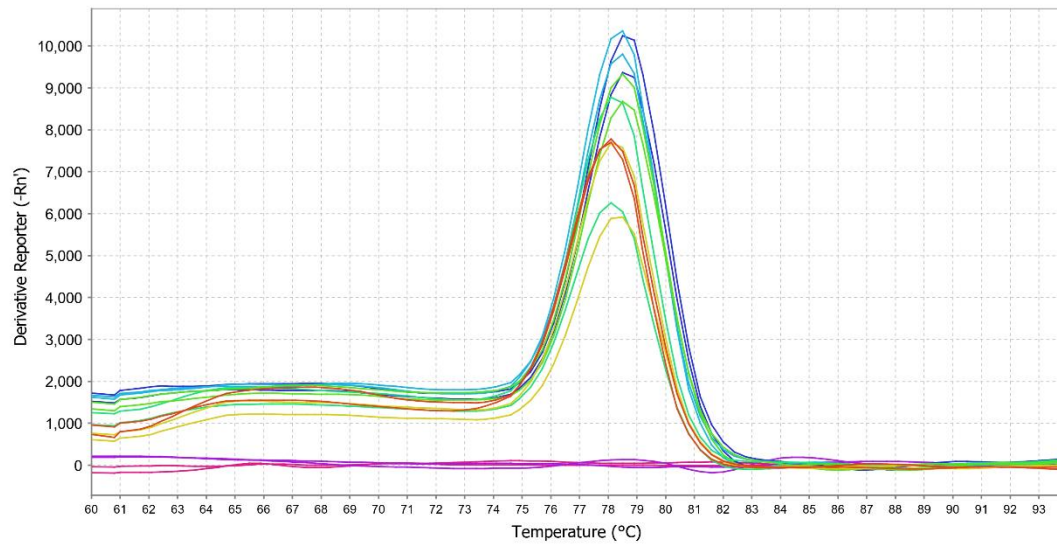
### Primer set 1



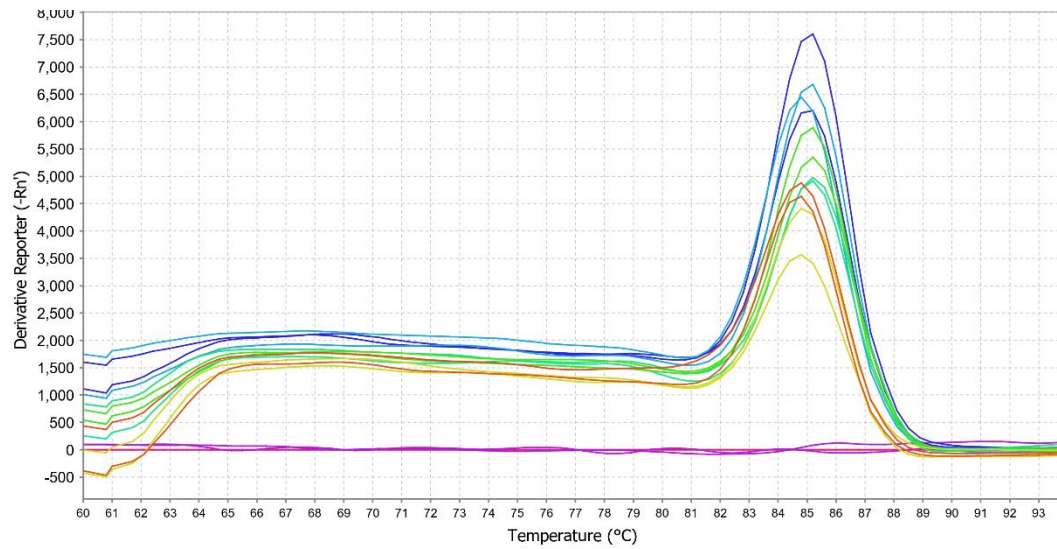
### Primer set 2



USP15696.1  $\alpha$ -amylase  
Primer set 1

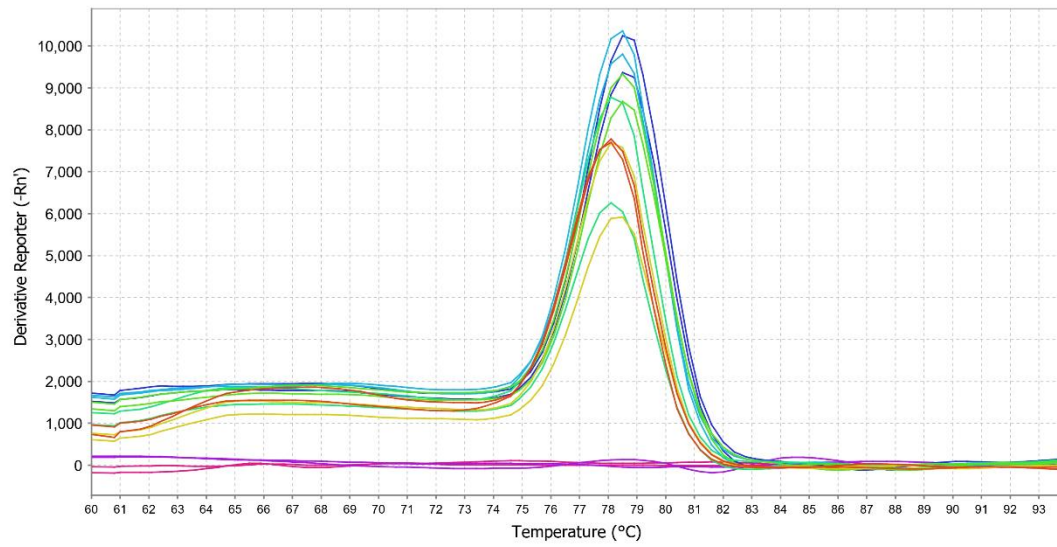


Primer set 2

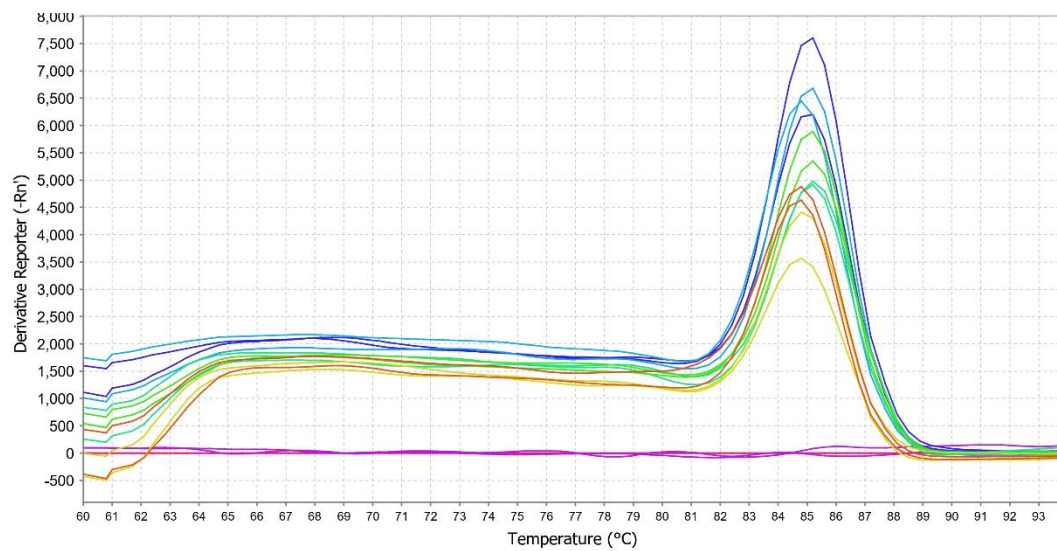


## USP15545.1 $\beta$ -glucosidase

### Primer set 1



### Primer set 2



**Table S1:** Genome attributes of *V. gazogenes* PB1

| Attribute             | Value                                                                     |
|-----------------------|---------------------------------------------------------------------------|
| Genome size           | 4,712,479 bp                                                              |
| GC %                  | 45.3 %                                                                    |
| Coverage              | 293X                                                                      |
| Number of chromosomes | 2 (3,516,273 bp and 1,196,206 bp)                                         |
| NCBI Accessions       | Chromosome 1: CP092587/ NZ_CP092587<br>Chromosome 2: CP092588/NZ_CP092588 |
| Genes (total)         | 4,178                                                                     |
| Genes (coding)        | 3,988                                                                     |
| Genes (RNA)           | 114                                                                       |
| rRNAs                 | 5S: 8; 16S: 7; 23S: 7                                                     |
| tRNAs                 | 87                                                                        |
| ncRNAs                | 5                                                                         |
| Pseudogenes           | 76                                                                        |

**Table S2:** *V. gazogenes* PB1 genes predicted by PATRIC to code for virulence factors

| Gene        | Gene Product                                        | Source    | Source ID   | Source Organism                                                            | Classification               | References           |
|-------------|-----------------------------------------------------|-----------|-------------|----------------------------------------------------------------------------|------------------------------|----------------------|
| <i>arcA</i> | Aerobic respiration control response regulator ArcA | Victors   | 218927655   | <i>Yersinia pestis</i> CO92                                                | Regulation                   | Merrell et al., 2002 |
| <i>atpA</i> | ATP synthase alpha chain                            | Victors   | 16767151    | <i>Salmonella enterica</i> subsp. enterica serovar Typhimurium str. LT2    | Metabolism                   | Merrell et al., 2002 |
| <i>bioB</i> | Biotin synthase                                     | Victors   | 147673516   | <i>Vibrio cholerae</i> O395                                                | Nutritional/Metabolic factor | Chen et al., 2005    |
| <i>cap</i>  | Cyclic AMP receptor protein                         | PATRIC_VF | SL1344_3433 | <i>Salmonella enterica</i> subsp. enterica serovar Typhimurium str. SL1344 |                              |                      |
| <i>carA</i> | Carbamoyl-phosphate synthase small chain            | PATRIC_VF | Z0037       | <i>Escherichia coli</i> O157:H7 str. EDL933                                | Nutritional/Metabolic factor | Chen et al., 2005    |
| <i>carB</i> | Carbamoyl-phosphate synthase large chain            | Victors   | 147673654   | <i>Vibrio cholerae</i> O395                                                | Nutritional/Metabolic factor | Chen et al., 2005    |
| <i>deoC</i> | Deoxyribose-phosphate aldolase                      | Victors   | 147674428   | <i>Vibrio cholerae</i> O395                                                |                              |                      |
| <i>dksA</i> | RNA polymerase-binding transcription factor DksA    | PATRIC_VF | S0140       | <i>Shigella flexneri</i> 2a str. 2457T                                     | Regulation                   | Merrell et al., 2002 |
| <i>fabG</i> | 3-oxoacyl-[acyl-carrier protein] reductase , FabG   | Victors   | 147675453   | <i>Vibrio cholerae</i> O395                                                | Metabolism                   | Merrell et al., 2002 |
| <i>fbp</i>  | Fructose-1,6-bisphosphatase, type I                 | Victors   | 147674062   | <i>Vibrio cholerae</i> O395                                                |                              |                      |
| <i>frdC</i> | Fumarate reductase subunit C                        | Victors   | 147674812   | <i>Vibrio cholerae</i> O395                                                |                              |                      |
| <i>fur</i>  | Ferric uptake Regulation protein FUR                | Victors   | 15642106    | <i>Vibrio cholerae</i> O1 biovar El Tor str. N16961                        | Regulation                   | Chen et al., 2005    |
| <i>glmM</i> | Phosphoglucosamine mutase                           | Victors   | 15640659    | <i>Vibrio cholerae</i> O1 biovar El Tor str. N16961                        | Immune modulation            | Chen et al., 2005    |

|               |                                                                                                                                    |           |             |                                                                             |                          |                      |
|---------------|------------------------------------------------------------------------------------------------------------------------------------|-----------|-------------|-----------------------------------------------------------------------------|--------------------------|----------------------|
| <i>glyA-1</i> | Serine hydroxymethyltransferase                                                                                                    | Victors   | 227081115   | <i>Vibrio cholerae</i> M66-2                                                |                          |                      |
| <i>gmhA</i>   | D-sedoheptulose 7-phosphate isomerase                                                                                              | PATRIC_VF | SL1344_0306 | <i>Salmonella enterica</i> subsp. enterica serovar Typhimurium str. SL1344  | Immune modulation        | Chen et al., 2005    |
| <i>grxD</i>   | Monothiol glutaredoxin GrxD                                                                                                        | PATRIC_VF | E2348C_1740 | <i>Escherichia coli</i> O127:H6 str. E2348/69                               |                          |                      |
| <i>guaA</i>   | GMP synthase [glutamine-hydrolyzing], amidotransferase subunit / GMP synthase [glutamine-hydrolyzing], ATP pyrophosphatase subunit | PATRIC_VF | S2725       | <i>Shigella flexneri</i> 2a str. 2457T                                      |                          |                      |
| <i>hcp-2</i>  | T6SS component Hcp                                                                                                                 | VFDB      | VFG007138   | <i>Vibrio cholerae</i> O1 biovar El Tor str. N16961                         | Effector delivery system | Chen et al., 2005    |
| <i>hfq</i>    | RNA-binding protein Hfq                                                                                                            | PATRIC_VF | S4595       | <i>Shigella flexneri</i> 2a str. 2457T                                      |                          |                      |
| <i>kdsA</i>   | 2-Keto-3-deoxy-D-manno-octulosonate-8-phosphate synthase                                                                           | VFDB      | VFG013465   | <i>Haemophilus influenzae</i> Rd KW20                                       | Immune modulation        | Chen et al., 2005    |
| <i>luxS</i>   | S-ribosylhomocysteine lyase @ Autoinducer-2 production protein LuxS                                                                | VFDB      | VFG018241   | <i>Vibrio cholerae</i> O1 biovar El Tor str. N16961                         | Biofilm                  | Chen et al., 2005    |
| <i>lysA</i>   | Diaminopimelate decarboxylase                                                                                                      | Victors   | 147674363   | <i>Vibrio cholerae</i> O395                                                 | Lysine synthesis         | Merrell et al., 2002 |
| <i>nqrE</i>   | Na(+)-translocating NADH-quinone reductase subunit E                                                                               | Victors   | 147674521   | <i>Vibrio cholerae</i> O395                                                 | Immune modulation        | Chen et al., 2005    |
| <i>nqrF</i>   | Na(+)-translocating NADH-quinone reductase subunit F                                                                               | Victors   | 147675137   | <i>Vibrio cholerae</i> O395                                                 | Immune modulation        | Chen et al., 2005    |
| <i>ompR</i>   | Two-component system response regulator OmpR                                                                                       | PATRIC_VF | SEN3328     | <i>Salmonella enterica</i> subsp. enterica serovar Enteritidis str. P125109 |                          |                      |
| <i>prfC</i>   | Peptide chain release factor 3                                                                                                     | Victors   | 165975488   | <i>Actinobacillus pleuropneumoniae</i> serovar 3 str. JL03                  |                          |                      |
| <i>pta</i>    | BioD-like N-terminal domain / Phosphate acetyltransferase                                                                          | Victors   | 147675254   | <i>Vibrio cholerae</i> O395                                                 |                          |                      |

|               |                                                                              |           |           |                                                                      |                                                            |                      |
|---------------|------------------------------------------------------------------------------|-----------|-----------|----------------------------------------------------------------------|------------------------------------------------------------|----------------------|
| <i>purD</i>   | Phosphoribosylamine--glycine ligase                                          | Victors   | 147673352 | Vibrio cholerae O395                                                 | Nutritional/Metabolic factor                               | Chen et al., 2005    |
| <i>purF</i>   | Amidophosphoribosyltransferase                                               | Victors   | 147673131 | Vibrio cholerae O395                                                 | Nutritional/Metabolic factor                               | Merrell et al., 2002 |
| <i>pykF</i>   | Pyruvate kinase                                                              | Victors   | 147674848 | Vibrio cholerae O395                                                 | Energy metabolism                                          | Merrell et al., 2002 |
| <i>rpmF</i>   | LSU ribosomal protein L32p @ LSU ribosomal protein L32p, zinc-independent    | Victors   | 126208850 | Actinobacillus pleuropneumoniae L20                                  |                                                            |                      |
| <i>rpmJ</i>   | LSU ribosomal protein L36p @ LSU ribosomal protein L36p, zinc-dependent      | Victors   | 56707499  | Francisella tularensis subsp. tularensis SCHU S4                     |                                                            |                      |
| <i>rpoE</i>   | RNA polymerase sigma factor RpoE                                             | Victors   | 147675349 | Vibrio cholerae O395                                                 | Regulation                                                 | Chen et al., 2005    |
| <i>rpoS</i>   | RNA polymerase sigma factor RpoS                                             | PATRIC_VF | SEN2763   | Salmonella enterica subsp. enterica serovar Enteritidis str. P125109 | Regulation                                                 | Chen et al., 2005    |
| <i>sspA</i>   | Stringent starvation protein A                                               | Victors   | 147675668 | Vibrio cholerae O395                                                 | Cellular process                                           | Merrell et al., 2002 |
| <i>trpB</i>   | Tryptophan synthase beta chain                                               | PATRIC_VF | Z2550     | Escherichia coli O157:H7 str. EDL933                                 |                                                            |                      |
| <i>tufA</i>   | Translation elongation factor Tu                                             | Victors   | 15676067  | Neisseria meningitidis MC58                                          | Adherence                                                  | Chen et al., 2005    |
| <i>VC0032</i> | AAA+ ATPase superfamily protein YifB/ComM, associated with DNA recombination | Victors   | 15640064  | Vibrio cholerae O1 biovar El Tor str. N16961                         | Unknown                                                    |                      |
| <i>VC0260</i> | Nucleoside-diphosphate sugar epimerase/dehydratase                           | Victors   | 15640289  | Vibrio cholerae O1 biovar El Tor str. N16961                         | Immune modulation                                          | Chen et al., 2005    |
| <i>VC0468</i> | Glutathione synthetase                                                       | Victors   | 15640495  | Vibrio cholerae O1 biovar El Tor str. N16961                         | Biosynthesis of co-factors, prosthetic groups and carriers | Merrell et al., 2002 |
| <i>VC0727</i> | Phosphate transport system regulatory protein PhoU                           | Victors   | 15640746  | Vibrio cholerae O1 biovar El Tor str. N16961                         | Regulation                                                 | Merrell et al., 2002 |

|                  |                                                              |         |           |                                                     |                                                     |                      |
|------------------|--------------------------------------------------------------|---------|-----------|-----------------------------------------------------|-----------------------------------------------------|----------------------|
| <i>VC0965</i>    | Phosphoenolpyruvate-protein phosphotransferase of PTS system | Victors | 15640981  | <i>Vibrio cholerae</i> O1 biovar El Tor str. N16961 | Transport and binding proteins                      | Merrell et al., 2002 |
| <i>VC1012</i>    | Electron transport complex protein RnfE                      | Victors | 15641027  | <i>Vibrio cholerae</i> O1 biovar El Tor str. N16961 | Energy metabolism                                   | Merrell et al., 2002 |
| <i>VC1098</i>    | Acetate kinase                                               | Victors | 15641111  | <i>Vibrio cholerae</i> O1 biovar El Tor str. N16961 | Energy metabolism                                   | Merrell et al., 2002 |
| <i>VC1715</i>    | Chromosome partition protein MukE                            | Victors | 15641719  | <i>Vibrio cholerae</i> O1 biovar El Tor str. N16961 | Cellular process                                    | Merrell et al., 2002 |
| <i>VC1716</i>    | Chromosome partition protein MukF                            | Victors | 15641720  | <i>Vibrio cholerae</i> O1 biovar El Tor str. N16961 | Cellular process                                    | Merrell et al., 2002 |
| <i>VC2067</i>    | Flagellar synthesis regulator FlcN                           | Victors | 15642067  | <i>Vibrio cholerae</i> O1 biovar El Tor str. N16961 | Motility                                            | Chen et al., 2005    |
| <i>VC2206</i>    | Flagellar protein FlgP                                       | Victors | 15642205  | <i>Vibrio cholerae</i> O1 biovar El Tor str. N16961 | Motility                                            | Chen et al., 2005    |
| <i>VC2295</i>    | Na(+)-translocating NADH-quinone reductase subunit A         | Victors | 15642293  | <i>Vibrio cholerae</i> O1 biovar El Tor str. N16961 | Immune modulation                                   | Merrell et al., 2002 |
| <i>VC2647</i>    | Transcriptional regulator, PadR family                       | Victors | 15642642  | <i>Vibrio cholerae</i> O1 biovar El Tor str. N16961 | Hypothetical conserved                              | Merrell et al., 2002 |
| <i>VC2660</i>    | Translation elongation factor P                              | Victors | 15642655  | <i>Vibrio cholerae</i> O1 biovar El Tor str. N16961 | Elongation factor P                                 | Merrell et al., 2002 |
| <i>VCA0006</i>   | FIG007491: hypothetical protein YeeN                         | Victors | 15600777  | <i>Vibrio cholerae</i> O1 biovar El Tor str. N16961 | Hypothetical conserved                              | Merrell et al., 2002 |
| <i>vipA/mglA</i> | T6SS component TssB (ImpB/VipA)                              | VFDB    | VFG002092 | <i>Vibrio cholerae</i> O1 biovar El Tor str. N16961 | Effector delivery system (Type VI secretion system) | Chen et al., 2005    |
| <i>vipB/mglB</i> | T6SS component TssC (ImpC/VipB)                              | VFDB    | VFG002093 | <i>Vibrio cholerae</i> O1 biovar El Tor str. N16961 | Effector delivery system (Type VI secretion system) | Chen et al., 2005    |
| <i>zwf</i>       | Glucose-6-phosphate 1-dehydrogenase                          | Victors | 147671692 | <i>Vibrio cholerae</i> O395                         | Energy metabolism                                   | Merrell et al., 2002 |

**Table S3:** Antibiotic resistance genes identified in *V. gazogenes* PB1 using PATRIC

| AMR Mechanism                                                      | Family/Genes                                                                                               |
|--------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------|
| Antibiotic activation enzyme                                       | <i>katG</i>                                                                                                |
| Antibiotic inactivation enzyme                                     | CatB family                                                                                                |
| Antibiotic target in susceptible species                           | Alr, Ddl, dxr, EF-G, EF-Tu, folA, Dfr, folP, gyrA, gyrB, Iso-tRNA, kasA, MurA, rho, rpoB, rpoC, S10p, S12p |
| Antibiotic target protection protein                               | QnrB family                                                                                                |
| Antibiotic target replacement protein                              | fabV                                                                                                       |
| Efflux pump conferring antibiotic resistance                       | EmrAB-OMF, EmrAB-TolC, MacA, MacB, Tet(35), TolC/OpmH                                                      |
| Gene conferring resistance via absence                             | gidB                                                                                                       |
| Protein altering cell wall charge conferring antibiotic resistance | GdpD, PgsA                                                                                                 |
| Regulator modulating expression of antibiotic resistance genes     | H-NS, OxyR                                                                                                 |

**Table S4:** Prophage regions identified by PHASTER in the *V. gazogenes* PB1 genome.

| Region                           | Region Length | Completeness | Score | Total Proteins | Region Position | Most Common Phage (Number of similar proteins) | NCBI Accession | GC %   |
|----------------------------------|---------------|--------------|-------|----------------|-----------------|------------------------------------------------|----------------|--------|
| <b>Chromosome 1 (CP092587.1)</b> |               |              |       |                |                 |                                                |                |        |
| 1                                | 9Kb           | incomplete   | 30    | 9              | 887974-897006   | PHAGE_Escher_500465_1 (7)                      | NC_049342      | 43.63% |
| 2                                | 21.1Kb        | intact       | 140   | 27             | 2061980-2083082 | PHAGE_Vibrio_VP882 (14)                        | NC_009016      | 45.64% |
| <b>Chromosome 2 (CP092588.1)</b> |               |              |       |                |                 |                                                |                |        |
| 1                                | 25.8Kb        | questionable | 80    | 32             | 526331-552169   | PHAGE_Vibrio_8 (14)                            | NC_022747      | 46.34% |
| 2                                | 39.2Kb        | intact       | 123   | 55             | 701354-740613   | PHAGE_Vibrio_12B12 (42)                        | NC_021070      | 46.34% |

**Table S5:** Genomic islands predicted by IslandPath-DIMOB in the *V. gazogenes* PB1 genome

| Chromosome 1 (CP092587.1) |         |         |
|---------------------------|---------|---------|
| Genome Island #           | Start   | End     |
| GI_1                      | 116587  | 131071  |
| GI_2                      | 767987  | 781104  |
| GI_3                      | 880889  | 892019  |
| GI_4                      | 1128508 | 1173292 |
| GI_5                      | 2330710 | 2340786 |
| GI_6                      | 2889764 | 2907705 |
| Chromosome 2 (CP092588.1) |         |         |
| Genome Island #           | Start   | End     |
| GI_1                      | 306912  | 323990  |
| GI_2                      | 697776  | 735882  |
| GI_3                      | 918370  | 940627  |
| GI_4                      | 1162062 | 1189180 |

**Table S6:** Comparison of the genomes between *V. gazogenes* PB1 and other members of the *Vibrio* genus

| Species                                  | <i>Vibrio gazogenes</i> | <i>Vibrio gazogenes</i> | <i>Vibrio spartinae</i> | <i>Vibrio cholerae</i> | <i>Vibrio parahaemolyticus</i> | <i>Vibrio natriegens</i> | <i>Vibrio alginolyticus</i> | <i>Vibrio harveyi</i> | <i>Vibrio campbellii</i> |
|------------------------------------------|-------------------------|-------------------------|-------------------------|------------------------|--------------------------------|--------------------------|-----------------------------|-----------------------|--------------------------|
| Strain                                   | PB1                     | ATCC 43942              | 3.6                     | RFB16                  | O3:K6 substr.<br>RIMD 2210633  | ATCC 14048               | E110                        | ATCC 33843            | BoB-53                   |
| NCBI Genome ID                           | 14723                   | 14723                   | 56603                   | 505                    | 691                            | 14718                    | 2331                        | 685                   | 11077                    |
| Genome size (Mb)                         | 4.71                    | 4.79                    | 5.01                    | 4.14                   | 5.17                           | 5.18                     | 5.15                        | 5.88                  | 5.43                     |
| Chromosome 1<br>(NCBI RefSeq Accession)  | NZ_CP092587.1           | NZ_CP018835.1           | NZ_CP046268.1           | NZ_CP043554.1          | NC_004603.1                    | NZ_CP009977.1            | NZ_CP098033.1               | NZ_CP009467.1         | NZ_CP026321.1            |
| Chromosome 1 (Mb)                        | 3.52                    | 3.47                    | 3.82                    | 2.95                   | 3.29                           | 3.25                     | 3.29                        | 3.62                  | 3.47                     |
| Chromosome 2<br>(NCBI RefSeq Accession)  | NZ_CP092588.1           | NZ_CP018836.1           | NZ_CP046269.1           | NZ_CP043556.1          | NC_004605.1                    | NZ_CP009978.1            | NZ_CP098034.1               | NZ_CP009468.1         | NZ_CP026322.1            |
| Chromosome 2 (Mb)                        | 1.2                     | 1.3                     | 1.19                    | 1.14                   | 1.88                           | 1.93                     | 1.82                        | 2.26                  | 1.96                     |
| Total number of CDS                      | 3988                    | 4084                    | 4167                    | 3639                   | 4519                           | 4488                     | 4589                        | 5205                  | 4768                     |
| Average Nucleotide<br>Identity (ANI) (%) | -                       | 94.22                   | 91.92                   | 71.61                  | 71.4                           | 71.4                     | 71.15                       | 71.15                 | 71.26                    |
| Average Amino Acid<br>Identity (AAI) (%) | -                       | 96.58                   | 94.77                   | 73.21                  | 72.4                           | 72.02                    | 72.12                       | 71.87                 | 71.84                    |
| Percentage of<br>Conserved Proteins (%)  | -                       | 85.46                   | 84.4                    | 59.36                  | 54.81                          | 55.92                    | 55.56                       | 52.7                  | 53.75                    |

**Table S7:** Annotated genes in the 100 kbp region of chromosome 2 of *V. gazogenes* ATCC 43942 that has been deleted in *V. gazogenes* PB1.

| Annotation                               | Start   | End     | Length | Strand  |
|------------------------------------------|---------|---------|--------|---------|
| <b>Region 1: 569,917-590,569</b>         |         |         |        |         |
| toxin HipA CDS                           | 570,310 | 571,549 | 1,240  | forward |
| transcriptional regulator CDS            | 571,534 | 572,016 | 483    | forward |
| hypothetical protein CDS                 | 572,805 | 573,683 | 879    | forward |
| allulose-6-phosphate 3-epimerase CDS     | 573,858 | 574,550 | 693    | reverse |
| PTS fructose transporter subunit IIC CDS | 574,738 | 575,838 | 1,101  | reverse |
| PTS fructose transporter subunit IIB CDS | 575,895 | 576,218 | 324    | reverse |
| PTS fructose transporter subunit IIA CDS | 576,287 | 576,754 | 468    | reverse |
| PTS fructose transporter subunit IIA CDS | 577,149 | 577,604 | 456    | reverse |
| hypothetical protein CDS                 | 577,628 | 579,268 | 1,641  | reverse |
| DNA helicase CDS                         | 579,722 | 583,243 | 3,522  | forward |
| zeta toxin family protein CDS            | 583,779 | 584,492 | 714    | reverse |
| metal-dependent hydrolase CDS            | 584,551 | 584,819 | 269    | reverse |
| hypothetical protein CDS                 | 584,941 | 587,802 | 2,862  | reverse |
| hypothetical protein CDS                 | 587,846 | 590,104 | 2,259  | reverse |
| <b>Region 2: 753,693-819,888</b>         |         |         |        |         |
| hypothetical protein CDS                 | 753,681 | 754,148 | 468    | forward |
| hypothetical protein CDS                 | 754,418 | 755,185 | 768    | reverse |
| threonine--tRNA ligase CDS               | 755,245 | 757,200 | 1,956  | reverse |
| hypothetical protein CDS                 | 757,671 | 758,885 | 1,215  | reverse |
| hypothetical protein CDS                 | 759,021 | 763,778 | 4,758  | reverse |
| hypothetical protein CDS                 | 763,775 | 767,893 | 4,119  | reverse |
| hypothetical protein CDS                 | 767,893 | 773,442 | 5,550  | reverse |
| hypothetical protein CDS                 | 773,439 | 781,196 | 7,758  | reverse |
| hypothetical protein CDS                 | 781,228 | 787,713 | 6,486  | reverse |
| hypothetical protein CDS                 | 787,809 | 793,187 | 5,379  | reverse |
| hypothetical protein CDS                 | 793,216 | 799,848 | 6,633  | reverse |
| hypothetical protein CDS                 | 799,845 | 807,335 | 7,491  | reverse |
| transposase CDS                          | 808,357 | 808,623 | 267    | forward |
| transposase CDS                          | 808,665 | 809,468 | 804    | forward |
| hypothetical protein CDS                 | 809,717 | 810,127 | 411    | reverse |
| hypothetical protein CDS                 | 810,142 | 811,065 | 924    | reverse |
| hypothetical protein CDS                 | 811,995 | 812,870 | 876    | forward |
| endonuclease V CDS                       | 812,963 | 813,604 | 642    | forward |
| hypothetical protein CDS                 | 813,596 | 814,486 | 891    | reverse |
| hypothetical protein CDS                 | 814,488 | 816,266 | 1,779  | reverse |
| hypothetical protein CDS                 | 816,786 | 817,661 | 876    | forward |

|                                      |           |           |       |         |
|--------------------------------------|-----------|-----------|-------|---------|
| hypothetical protein CDS             | 817,763   | 818,062   | 300   | reverse |
| hypothetical protein CDS             | 818,055   | 818,285   | 231   | reverse |
| hypothetical protein CDS             | 818,433   | 821,108   | 2,676 | reverse |
| <b>Region 3: 821,079-831,326</b>     |           |           |       |         |
| hypothetical protein CDS             | 821,535   | 822,413   | 879   | forward |
| hypothetical protein CDS             | 822,556   | 822,837   | 282   | reverse |
| hypothetical protein CDS             | 823,240   | 823,659   | 420   | reverse |
| hypothetical protein CDS             | 823,656   | 823,925   | 270   | reverse |
| hypothetical protein CDS             | 824,355   | 824,813   | 459   | reverse |
| hypothetical protein CDS             | 824,820   | 826,034   | 1,215 | reverse |
| hypothetical protein CDS             | 826,037   | 826,219   | 183   | forward |
| hypothetical protein CDS             | 826,286   | 826,624   | 339   | reverse |
| hypothetical protein CDS             | 826,627   | 827,766   | 1,140 | reverse |
| IS256 family transposase CDS         | 827,940   | 828,178   | 239   | forward |
| hypothetical protein CDS             | 828,252   | 828,701   | 450   | reverse |
| hypothetical protein CDS             | 828,815   | 829,054   | 240   | reverse |
| hypothetical protein CDS             | 829,390   | 829,815   | 426   | reverse |
| hypothetical protein CDS             | 829,974   | 830,339   | 366   | reverse |
| hypothetical protein CDS             | 830,649   | 831,032   | 384   | reverse |
| hypothetical protein CDS             | 831,223   | 831,489   | 267   | forward |
| <b>Region 4: 1,244,876-1,261,341</b> |           |           |       |         |
| hypothetical protein CDS             | 1,245,021 | 1,245,206 | 186   | forward |
| amino acid adenylation CDS           | 1,245,262 | 1,245,930 | 669   | forward |
| MBL fold hydrolase CDS               | 1,246,119 | 1,247,210 | 1,092 | forward |
| hypothetical protein CDS             | 1,247,212 | 1,247,853 | 642   | forward |
| 5'-nucleotidase CDS                  | 1,248,154 | 1,249,113 | 960   | forward |
| hypothetical protein CDS             | 1,249,214 | 1,250,101 | 888   | forward |
| endonuclease V CDS                   | 1,250,093 | 1,250,773 | 681   | reverse |
| hypothetical protein CDS             | 1,250,827 | 1,251,702 | 876   | reverse |
| hypothetical protein CDS             | 1,252,196 | 1,252,570 | 375   | forward |
| hypothetical protein CDS             | 1,252,765 | 1,253,187 | 423   | reverse |
| hypothetical protein CDS             | 1,253,254 | 1,253,982 | 729   | forward |
| hypothetical protein CDS             | 1,253,997 | 1,254,401 | 405   | forward |
| hypothetical protein CDS             | 1,254,678 | 1,254,929 | 252   | reverse |
| hypothetical protein CDS             | 1,255,120 | 1,255,503 | 384   | forward |
| hypothetical protein CDS             | 1,255,791 | 1,256,174 | 384   | forward |
| hypothetical protein CDS             | 1,256,512 | 1,256,910 | 399   | forward |
| hypothetical protein CDS             | 1,257,198 | 1,257,754 | 557   | forward |
| hypothetical protein CDS             | 1,258,329 | 1,258,715 | 387   | forward |
| hypothetical protein CDS             | 1,258,811 | 1,259,689 | 879   | reverse |

**Table S8.** CRISPR array and CRISPR-Cas operons identified in *V. gazogenes* PB1 using CRISPRCasFinder

| Chromosome | Start   | End     | Array length | Direct repeat sequence              | Repeat Length | # of spacers | Evidence level | CRISPR-Cas type | Cas genes                                             |
|------------|---------|---------|--------------|-------------------------------------|---------------|--------------|----------------|-----------------|-------------------------------------------------------|
| 1          | 1916871 | 1918302 | 1432         | CGGTTTCATCCCCGTGTGTACGGGGAACAC      | 29            | 23           | 4              | -               | -                                                     |
| 1          | 2855893 | 2856160 | 268          | TTTCTAAGCTGCCTATGCGGCAGTGAAC        | 28            | 4            | 4              | -               | -                                                     |
| 1          | 3441994 | 3442426 | 433          | ATCGGAAGACACTGGCCGAGACAGGCCGTT      | 30            | 6            | 3              | -               | -                                                     |
| 2          | 495287  | 496047  | 761          | GTGTTCCCCGTATGCACGGGGATGAACCG       | 29            | 12           | 4              | I-E             | Cas3, Cas8, Cse2, Cas7, Cas5e, Cas6, Cas1, Cas2       |
| 2          | 874590  | 875029  | 440          | GTCTTAACGGCCTGTTTCGGCCAAGGTATTTCCAC | 35            | 6            | 4              | III-B           | Cmr6, Cmr5, Cmr4, Cmr3, Cmr2, Cas10, Csm3, Cas1, Cas2 |
| 2          | 1088527 | 1089043 | 517          | CGGTTTCATCCCCGTGCATACGGGGAACAC      | 29            | 8            | 4              | III-D           | Cas10, Csm3, Csm3, Cas1, Cas2                         |
| 2          | 1110778 | 1111161 | 384          | GTTTCAGTAGGTCTGGCCGTTTAGGCCGTTGAGAC | 35            | 5            | 4              | -               | -                                                     |

**Table S9:** Results of spacer BLASTN search

| Chromosome | CRISPR array start | CRISPR array end | Spacer start | Spacer end | Spacer sequence                  | % identity | Query coverage | NCBI accession | Organism                                                            |
|------------|--------------------|------------------|--------------|------------|----------------------------------|------------|----------------|----------------|---------------------------------------------------------------------|
| 1          | 1916871            | 1918302          | 1916900      | 1916931    | GATAATGACCTGCTGGTTGCGGTCGCCGTTA  | 100%       | 100%           | JX556417       | <i>Vibrio</i> phage vB_VpaM_MAR                                     |
|            |                    |                  |              |            |                                  | 100%       | 100%           | FN297812       | <i>Vibrio</i> phage VP58.5                                          |
|            |                    |                  |              |            |                                  | 100%       | 100%           | AY133112       | <i>Vibrio</i> phage VHML                                            |
| 2          | 495287             | 496047           | 495926       | 495957     | CGTTCATTACCGACCGGGATTCGTGGGAGTAC | 100%       | 100%           | CP020037.1     | <i>Vibrio parahaemolyticus</i> strain 20130629002S01 plasmid pVPGX2 |

**Table S10:** Primers used for the RT-qPCR analysis

| Gene                                                 | Approximate amplicon sizes (bp) |     | Primer Set 1                                                                                 | Primer set 2                                                                                 |
|------------------------------------------------------|---------------------------------|-----|----------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------|
|                                                      |                                 |     |                                                                                              |                                                                                              |
| USP16020.1<br>(glgX)                                 | 72                              | 123 | <i>glgX_set 1_F</i> : CACATCGATGGCTTCCGTTTCG<br><i>glgX_set 1_R</i> : CGCATTGCGACTGAAGGTTT   | <i>glgX_set 2_F</i> : CTGATCGCAGAACCTTGGG<br><i>glgX_set 2_R</i> : CTTGATCACCACGCCAGAAG      |
| USP15696.1<br>(starch-binding protein)               | 96                              | 171 | <i>SBP_set 1_F</i> : CGCACCAATGCCAGTAATGA<br><i>SBP_set 1_R</i> : CGGATAGCCGTAAGGGTAGG       | forward <i>SBP_set 2_F</i> : CCACAGCGATGATGCACCG<br><i>SBP_set 2_R</i> : CCGTTATCCCACAGTGAG  |
| USP13629.1<br>(alpha-glucosidase)                    | 91                              | 95  | <i>AG_set 1_F</i> : GGCGGTTATACGACCCTAC<br><i>AG_set 1_R</i> : CCTTCATGGCTACGCATAATC         | forward <i>AG_set 2_F</i> : CAAGGTTATCTGGCAACGCA<br><i>AG_set 2_R</i> : GCATCCGGATTGGTGAAGTC |
| USP15540.1<br>(glycoside hydrolase family 9 protein) | 102                             | 126 | <i>GHF9_set 1_F</i> : GAGCCGGTGCAAATTATCG<br><i>GHF9_set 1_R</i> : ATCAATCGCAAAGGGATGGC      | <i>GHF9_set 2_F</i> : AATGGGAAAGGCAGATCCGA<br><i>GHF9_set 2_R</i> : GGGCCATCAACGAAGTAACC     |
| USP14760.1<br>(glycoside hydrolase family 6 protein) | 161                             | 102 | <i>GHF6_set 1_F</i> : GTAGCAGCGATGACCGTAAC<br><i>GHF6_set 1_R</i> : GCATACGTGTGTGATGTGCTG    | <i>GHF6_set 2_F</i> : CGAAACTAACCGTCAGTGATGGC<br><i>GHF6_set 2_R</i> : GTTACGGTCATCGCTGCTAC  |
| USP15545.1<br>(GH1 family beta-glucosidase)          | 179                             | 127 | <i>GH1BG_set 1_F</i> : GTAACAAACCCGGCAAAGTGG<br><i>GH1BG_set 1_R</i> : CGCTAAACCTTGCGGATTGAC | <i>GH1BG_set 2_F</i> : CTATACCCGGAATGTGGTTTCG<br><i>GH1BG_set 2_R</i> : CGGACGAGTAAATCCGTC   |
| 16S RNA                                              | 95                              | 154 | <i>HK1_16S_set 1_F</i> : CGATGCAACGCGAAGAACC<br><i>HK1_16S_set 1_R</i> : GCACCTGTCTCTCAGTTCC | <i>HK1_16S_set 2_F</i> : GGAAGTGAAGACAGGTGC<br><i>HK1_16S_set 2_R</i> : CCCACCTTCTCCGGTTTAT  |

**Table S11:** A one-way ANOVA was performed to compare the effect of three different carbon sources on the transcript levels of selected genes related to cellulose and starch metabolism, as determined from the PGAP annotation pipeline. A p-value of less than 0.05 was considered significant.

|                                                                   | Cultivation time<br>(18 hours) |           |        | F-value | P-value    |
|-------------------------------------------------------------------|--------------------------------|-----------|--------|---------|------------|
|                                                                   | Glucose                        | Cellulose | Starch |         |            |
| <b>USP16020.1</b><br><b>isoamylase</b>                            | 1.0                            | 0.2       | 0.1    | 44.6    | 0.00000049 |
|                                                                   | 1.0                            | 0.2       | 0.7    |         |            |
|                                                                   | 1.0                            | 0.3       | 0.2    |         |            |
|                                                                   | 1.0                            | 0.5       | 0.3    |         |            |
|                                                                   | 1.0                            | 0.3       | 0.3    |         |            |
|                                                                   | 1.0                            | 0.4       | 0.2    |         |            |
| <b>USP15696.1</b><br><b><math>\alpha</math>-amylase</b>           | 1.0                            | 0.1       | 16.7   | 11.0    | 0.00116875 |
|                                                                   | 1.0                            | 0.1       | 8.0    |         |            |
|                                                                   | 1.0                            | 0.0       | 2.6    |         |            |
|                                                                   | 1.0                            | 0.2       | 18.3   |         |            |
|                                                                   | 1.0                            | 0.1       | 6.9    |         |            |
|                                                                   | 1.0                            | 0.1       | 4.0    |         |            |
| <b>USP15545.1</b><br><b><math>\beta</math>-glucosidase</b>        | 1.0                            | 0.9       | 0.3    | 64.6    | 0.00000004 |
|                                                                   | 1.0                            | 0.3       | 0.1    |         |            |
|                                                                   | 1.0                            | 0.4       | 0.1    |         |            |
|                                                                   | 1.0                            | 0.5       | 0.3    |         |            |
|                                                                   | 1.0                            | 0.5       | 0.2    |         |            |
|                                                                   | 1.0                            | 0.3       | 0.0    |         |            |
| <b>USP14760.1</b><br><b>exoglucanase</b>                          | 1.0                            | 30.0      | 7.1    | 15.7    | 0.00021070 |
|                                                                   | 1.0                            | 35.3      | 6.9    |         |            |
|                                                                   | 1.0                            | 18.1      | 1.2    |         |            |
|                                                                   | 1.0                            | 37.0      | 7.9    |         |            |
|                                                                   | 1.0                            | 9.2       | 1.5    |         |            |
|                                                                   | 1.0                            | 10.2      | 0.9    |         |            |
| <b>USP13629.1</b><br><b><math>\alpha</math>-glucosidase</b>       | 1.0                            | 43.5      | 13.2   | 18.8    | 0.00008263 |
|                                                                   | 1.0                            | 49.6      | 9.8    |         |            |
|                                                                   | 1.0                            | 41.5      | 12.7   |         |            |
|                                                                   | 1.0                            | 17.0      | 3.6    |         |            |
|                                                                   | 1.0                            | 19.4      | 3.4    |         |            |
|                                                                   | 1.0                            | 17.8      | 3.1    |         |            |
| <b>USP15540.1</b><br><b>endo-1,4-<math>\beta</math>-glucanase</b> | ND                             | ND        | ND     |         |            |