Chapter 6

Variation in gene expression in phylogenetically-selected *V. cholerae*

Contribution statement

Nick Thomson supervised the work described in this chapter. George Salmond, Gordon Dougan, and Julian Parkhill made substantive intellectual contributions to experimental design. rRNA depletion, library construction, and cDNA sequencing was performed by the WSI Bespoke RNA team. The live isolates used in this project were supplied by IVI, NCTC, and PHE, and have been described in Chapter 5.

I performed RNA-seq experiments, analysed the data and produced all figures.

COVID-19 statement

The work described in this chapter was severely affected by the shutdown imposed on the University of Cambridge and the Sanger Institute by the COVID-19 pandemic. This affected both the generation and analysis of RNA-seq data, as well as *in vitro* validation of the results from transcriptomic analysis. Aspects of the analysis herein may therefore be incomplete, but are the subject of current and future research.

6.1 – Overview

Throughout the preceding chapters, it has been demonstrated and discussed that the O1 serogroup and El Tor biotype do not exclusively describe the lineage of *V. cholerae* causing current pandemic cholera. It is also clear that pathogenicity islands and virulence determinants once thought to be found solely in pandemic *V. cholerae* are not exclusively found in pandemic lineages of *V. cholerae* (Chapter 5). In spite of this, the pandemic 7PET lineage continues to present a considerably elevated public health threat relative to that of local or endemic *V. cholerae*, as has been shown previously [189] and is exemplified by the Argentinian case study presented in Chapter 3. The biological basis of this phenomenon remains unclear.

Although *V. cholerae* can be compared and contrasted in terms of their gene complements (as has been done throughout this thesis), it is also necessary to consider functional differences between these bacteria. One approach that can be applied to understanding the biological differences between related bacteria is that of comparative transcriptomics, whereby bacteria are cultured under similar conditions and the expression profile of the genes they harbour is determined. Such experiments provide a snapshot of gene expression under appropriately-controlled *in vitro* (or *in vivo*) growth conditions.

Comparative transcriptomic experiments have been carried out in *V. cholerae* since the development of microarrays using the sequence of the N16961 reference genome [59]. Microarray-based experiments rely on the hybridisation of fluorescently-labelled cDNA, generated from total RNA extracted from *V. cholerae* grown under specific conditions, to spots of complementary ssDNA corresponding to genes on glass microarray slides. Experiments using microarrays have been used to characterise *V. cholerae* gene expression *in vitro* [468] and during infection [128], in the hyperinfectious state post-shedding [469], and to define regulons by comparing isogenic knock-outs in master regulators to a wild-type background [129]. This approach has also been used to characterise differences in gene expression between classical and El Tor biotype isolates (O395 and A1552, respectively) [470].

Over time, transcriptomic experiments shifted from microarray-based experiments to those employing RNA-seq, the use of next-generation sequencing to sequence quantitatively cDNA produced from *V. cholerae*. These experiments have included exploring bacterial gene expression during infection [471] and defining gene regulation networks [472, 473]. Recent

work has sought to collate previously-published *V. cholerae* gene expression studies for the purposes of meta-analyses [474]. By sequencing transcripts from the bacterium, RNA-seq also has the advantage of being able to monitor changes in expression of small RNAs (sRNAs) (*e.g.*, [475]), and other elements that microarray experiments cannot capture. It is important to note that the majority of transcriptomic experiments in *V. cholerae* have made use of a small number of laboratory strains. These include the 7PET strains N16961 and A1552, and the Classical strain O395.

In this chapter, I describe pilot experiments designed to optimise RNA-seq analysis and experiments for use under CL3 conditions. As mentioned in section 4.1, *V. cholerae* is classified as a bioterrorism agent under ATCSA Schedule 5. The Schedule 5-compliant laboratories at WSI operate at CL3, which imposes considerable constraint on the scale and scope of experimental work on *V. cholerae*. Therefore, preliminary experiments were important to optimise RNA isolation protocols under these containment conditions. As part of establishing these methodologies, I performed a transcriptomic experiment designed to explore the variation in gene expression of eight *V. cholerae* across multiple growth conditions, chosen to represent multiple lineages and genotypes of *V. cholerae*. I present the rationale for choosing these isolates, which was principally to determine whether pandemic and non-pandemic *V. cholerae* O1 exhibit different patterns of gene expression. I highlight a number of these results which are amenable to future study, and discuss how these follow-up experiments might be performed and what they might tell us about *V. cholerae* biology.

6.2 – Specific aims

In this final chapter, I aimed to:

1. Optimise experimental methods for CL3 operations by comparing the transcriptomes of Classical and 7PET *V. cholerae* grown under identical conditions, and compare the lists of differentially-expressed genes to those identified in previous studies,
2. Select rationally a total of eight live *V. cholerae* isolates suitable for transcriptomic comparisons, and
3. Execute a large pilot experiment to compare gene expression across eight *V. cholerae*, and to carry out an initial analysis of the resulting data.

6.3 – Results

6.3.1 – Methods optimisation and initial transcriptomic studies

Initial experiments were carried out in order to optimise transcriptomic methodologies for use in our CL3 laboratory. These experiments were also carried out in order to assess (i) the quality of RNA produced by our extraction protocol, (ii) the reproducibility within an experiment carried out under CL3 conditions, and (iii) the degree to which transcriptomic data generated using the live isolates at our disposal were concordant with previously-published datasets. In preliminary experiments, biological triplicate 25 ml cultures of *V. cholerae* were grown to an OD_{600} of 0.55 in LB liquid media at a defined temperature (either 30 or 37 ˚C) in baffled flasks with aeration (180 rpm). The OD_{600} of each culture was measured hourly. Once the target OD_{600} was reached, bacteria were collected from the culture by centrifugation, and the cell pellet frozen immediately at -80 ˚C. Total RNA was then extracted from cell pellets (details set out in Methods, section 2.2.16). This methodology is also outlined in Figure 6.1.

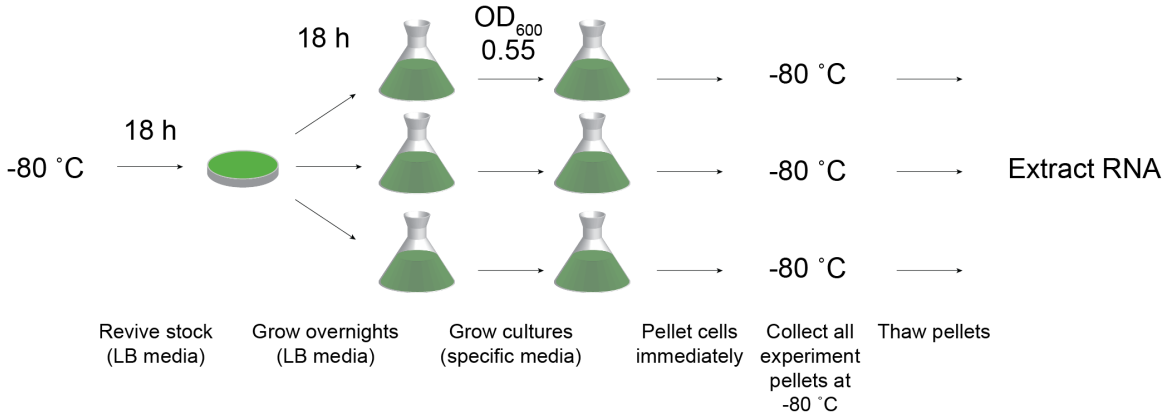

Figure 6.1 – Overview of RNA-seq experimental methodology for pilot experiment. All manipulations with the exception of culture incubation were performed inside a Class II MSC.

As well as assessing whether intact and sequence-able RNA could be isolated in the CL3 laboratory, this experiment aimed to determine whether differences in gene expression could be detected between a 7PET and a Classical culture grown to the same density. Cultures of bacteria were grown at both 37 ˚C and 30 ˚C, to assess the influence of temperature on gene expression in a strain, and whether these differences were similar or different across strains.

6.3.2 - RNA integrity and sequencing

Prior to submitting extracted RNA for rRNA depletion, library preparation and sequencing, the integrity of extracted RNA was assessed both by gel electrophoresis and by analysis on an Agilent Bioanalyser 2100 (Methods, section 2.2.17; data not shown). Only intact RNA samples were used for library construction; any which appeared to be degraded were discarded. Following the depletion of rRNA, strand-specific library preparation was performed and libraries were sequenced on one Illumina HiSeq 4000 lane. All of the samples for this experiment were batched together and sequenced on the same HiSeq lane, aiming to produce at least 10 million reads per sample. Details of total numbers of reads per sample, and the number of reads mapped to the N16961 genome sequence, are listed in Table 6.1 (see section 2.1.18 for computational details). Briefly, sequencing reads were mapped to the sequences of chromosomes 1 and 2 separately, gene expression was quantified, and differential gene expression analysis was performed using the Rockhopper package.

Sample ID	Strain	Condition	Number of reads	Reads mapped Chr1	Reads mapped Chr2
25900_7#1	7PET	37 ˚C, rep 1	14,092,482	12,200,919 (87%)	353,889 (3%)
25900_7#2	7PET	37 ˚C, rep 2	16,353,593	14,384,549 (88%)	283,329 (2%)
25900_7#3	7PET	37 ˚C, rep 3	22,897,140	19,977,810 (87%)	239,410 (1%)
25900_7#4	Classical	37 ˚C, rep 1	13,233,279	11,288,322 (85%)	334,318 (3%)
25900_7#5	Classical	37 ˚C, rep 2	18,441,714	15,733,617 (85%)	516,216 (3%)
25900_7#6	Classical	37 ˚C, rep 3	17,065,558	14,213,502 (83%)	730,105 (4%)
25900_7#19	7PET	30 ˚C, rep 1	16,724,553	12,862,968 (77%)	2,058,588 (12%)
25900_7#20	7PET	30 ˚C, rep 2	25,965,794	20,038,037 (77%)	3,092,668 (12%)
25900_7#21	7PET	30 ˚C, rep 3	26,594,182	21,059,504 (79%)	2,645,698 (10%)
25900_7#22	Classical	30 ˚C, rep 1	27,133,194	21,954,557 (81%)	1,552,086 (6%)
25900_7#23	Classical	30 ˚C, rep 2	20,970,321	16,216,919 (77%)	1,659,157 (8%)
25900_7#24	Classical	30 ˚C, rep 3	22,291,274	18,175,615 (82%)	1,026,833 (5%)

Table 6.1 – Lane IDs and summary statistics for RNA sequenced in pilot experiment. All reads mapped to the N16961 reference genome, accession # GCA_000006745.1. Strains chosen were MJD1402 (7PET) and MJD1404 (Classical); details presented in Table 6.6.

6.3.3 – Identification of differentially-expressed genes in pilot data

Previous studies of inter-biotype differential gene expression had been performed using microarray technologies, and had quantified differential gene expression for *V. cholerae* cultured in M9 + NRES minimal media at 30 ˚C to an OD_{600} of 0.65 [470]. Supplementation

of minimal media with asparagine, arginine, glutamate and serine (NRES) is required to induce expression of TCP and CT in minimal media [476]. This previous work had compared gene expression in the O395 strain of Classical lineage *V. cholerae* to that of A1552, a 7PET LAT-1 isolate from 1992 (detailed in section 3.4.5). Although the target OD_{600} and growth temperature in our pilot experiment was similar to those used by Beyhan *et al.*, the growth media were not identical.

Therefore, the first analysis carried out using these data aimed to identify genes which were differentially-expressed at 30 ˚C between 7PET and Classical *V. cholerae*. The fact that Beyhan and colleagues had performed a similar experiment using non-identical reference strains provided an important benchmark for our experiment, allowing for any expression differences that might be expected from the literature to be identified. Genes were determined to be differentially expressed if the $\log_2$ fold-change of expression between conditions was greater than or equal to 2, or less than or equal to -2, and that the q-value for that calculation was less than 0.01 (to account for multiple testing). Genes which met these criteria are listed in Tables 6.2 and 6.3 - triplicate data have been normalised to produce average expression values.

Gene ID	Product	Expression.7PET	Expression.Cl	pValue	qValue	log2Ratio
		Chromosome 1				
VC_0107	hypothetical protein	1323	209	9.07E-33	1.79E-31	2.6622 3821
VC_0163	conserved hypothetical protein	237	39	5.23E-15	7.21E-14	2.6033 4103
VC_0181	conserved hypothetical protein	339	1	0	0	8.4051 4146
VC_0205	hypothetical protein	141	0	0	0	Inf
VC_0285	4-hydroxy-2-oxoglutarate aldolase/2-deydro-3-deoxyphosphogluconate aldolase	146	8	2.09E-122	6.04E-121	4.1898 2456
VC_0471	SprT protein, putative	91	9	6.80E-27	1.26E-25	3.3378 6964
VC_0502	type IV pilin, putative	707	1	0	0	9.4655 664
VC_0512	methyl-accepting chemotaxis protein	331	0	0	0	Inf
VC_0707	hypothetical protein	1298	45	0	0	4.8502 2157
VC_0732	transcriptional regulator, LysR family	90	14	7.19E-19	1.12E-17	2.6844 9817
VC_0771	vibriobactin-specific isochorismatase	224	31	1.09E-43	2.45E-42	2.8531 5861
VC_0775	vibriobactin synthesis protein, putative	4864	460	1.18E-239	4.13E-238	3.4024 3746
VC_0867	hypothetical protein	132	23	4.96E-08	4.92E-07	2.5208 3216
VC_0990	transcriptional activator RfaH, putative	814	69	1.29E-154	4.01E-153	3.5603 6053
VC_1008	sodium-type flagellar protein MotY	980	94	9.33E-201	3.10E-199	3.3820 4909
VC_1292	hypothetical protein	4296	714	2.68E-53	6.41E-52	2.5889 9801
VC_1455	transcriptional repressor RstR	96	0	0	0	Inf
VC_1464	transcriptional repressor RstR	96	0	0	0	Inf
VC_1616	glutaredoxin, putative	290	15	2.93E-114	8.38E-113	4.2730 1849
VC_1654	hypothetical protein	160	13	2.48E-38	5.32E-37	3.6214 8838
VC_1788	hypothetical protein	135	16	2.57E-34	5.16E-33	3.0768 156
VC_1822	PTS system, fructose-specific IIABC component	268	32	7.56E-104	2.13E-102	3.0660 8919
VC_1824	PTS system, nitrogen regulatory IIA component, putative	900	140	5.47E-72	1.39E-70	2.6844 9817
VC_1828	conserved hypothetical protein	1246	77	0	0	4.0163 0181
VC_1829	hypothetical protein	455	62	7.41E-20	1.19E-18	2.8755 2642
VC_1970	benzoate transport protein	1354	92	0	0	3.8794 5007
VC_2147	hypothetical protein	173	28	1.23E-15	1.74E-14	2.6272 7331
VC_2324	transcriptional regulator, LysR family	96	4	5.67E-172	1.83E-170	4.5849 625
VC_2327	hypothetical protein	4128	205	0	0	4.3317 4716
VC_2328	hypothetical protein	260	25	7.67E-27	1.42E-25	3.3785 1162
VC_2329	2,3,4,5-tetrahydropyridine-2-carboxylate N-succinyltransferase	95	15	1.63E-21	2.75E-20	2.6629 6501
VC_2365	hypothetical protein	1352	190	1.52E-85	4.11E-84	2.8310 2383
VC_2509	hypothetical protein	93	8	2.01E-20	3.30E-19	3.5391 5881
VC_2534	magnesium transporter	216	27	4.18E-67	1.04E-65	3
VC_2544	fructose-1,6-bisphosphatase	106	7	4.93E-110	1.40E-108	3.9205 6553
VC_2545	inorganic pyrophosphatase	566	27	0	0	4.3897 7074
VC_2563	hypothetical protein	73	12	2.25E-12	2.83E-11	2.6048 6206
VC_2608	ABC transporter, ATP-binding protein	176	20	1.56E-85	4.19E-84	3.1375 0352

VC_2613	phosphoribulokinase	73	3	8.29E-207	2.78E-205	4.6048 6206
VC_2707	hypothetical protein	126	21	7.79E-09	8.09E-08	2.5849 625
VC_2759	fatty oxidation complex, beta subunit	143	20	1.30E-39	2.81E-38	2.8379 4324
VC_t038	tRNA-Met	186	15	5.27E-35	1.07E-33	3.6322 6822
VC_t068	tRNA-Leu	95	14	4.27E-09	4.57E-08	2.7625 0069
VC_0107	hypothetical protein	1323	209	9.07E-33	1.79E-31	2.6622 3821
VC_0163	conserved hypothetical protein	237	39	5.23E-15	7.21E-14	2.6033 4103
VC_0181	conserved hypothetical protein	339	1	0	0	8.4051 4146
VC_0205	hypothetical protein	141	0	0	0	Inf
VC_0285	4-hydroxy-2-oxoglutarate aldolase/2-deydro-3-deoxyphosphogluconate aldolase	146	8	2.09E-122	6.04E-121	4.1898 2456
VC_0471	SprT protein, putative	91	9	6.80E-27	1.26E-25	3.3378 6964
VC_0502	type IV pilin, putative	707	1	0	0	9.4655 664
VC_0512	methyl-accepting chemotaxis protein	331	0	0	0	Inf
VC_0707	hypothetical protein	1298	45	0	0	4.8502 2157
VC_0732	transcriptional regulator, LysR family	90	14	7.19E-19	1.12E-17	2.6844 9817
VC_0771	vibriobactin-specific isochorismatase	224	31	1.09E-43	2.45E-42	2.8531 5861
VC_0775	vibriobactin synthesis protein, putative	4864	460	1.18E-239	4.13E-238	3.4024 3746
VC_0867	hypothetical protein	132	23	4.96E-08	4.92E-07	2.5208 3216
VC_0990	transcriptional activator RfaH, putative	814	69	1.29E-154	4.01E-153	3.5603 6053
VC_1008	sodium-type flagellar protein MotY	980	94	9.33E-201	3.10E-199	3.3820 4909
VC_1292	hypothetical protein	4296	714	2.68E-53	6.41E-52	2.5889 9801
VC_1455	transcriptional repressor RstR	96	0	0	0	Inf
VC_1464	transcriptional repressor RstR	96	0	0	0	Inf
VC_1616	glutaredoxin, putative	290	15	2.93E-114	8.38E-113	4.2730 1849
VC_1654	hypothetical protein	160	13	2.48E-38	5.32E-37	3.6214 8838
VC_1788	hypothetical protein	135	16	2.57E-34	5.16E-33	3.0768 156
VC_1822	PTS system, fructose-specific IIABC component	268	32	7.56E-104	2.13E-102	3.0660 8919
VC_1824	PTS system, nitrogen regulatory IIA component, putative	900	140	5.47E-72	1.39E-70	2.6844 9817
VC_1828	conserved hypothetical protein	1246	77	0	0	4.0163 0181
	Chromosome 2					
VC_A0029	transcriptional regulator, putative	96	21	8.07E-08	4.60E-07	2.1926 4508
VC_A0077	sulfate permease family protein	188	40	2.35E-22	2.27E-21	2.2326 6076
VC_A0080	GGDEF family protein	360	38	8.89E-121	1.72E-119	3.2439 2558
VC_A0106	hypothetical protein	92	20	1.44E-10	9.50E-10	2.2016 3386
VC_A0224	hypothetical protein	92	11	2.21E-16	1.78E-15	3.0641 3034
VC_A0233	hypothetical protein	199	37	4.92E-10	3.10E-09	2.4271 7125
VC_A0296	hypothetical protein	4114	502	2.33E-70	3.91E-69	3.0347 8252
VC_A0298	hypothetical protein	76	8	3.62E-16	2.89E-15	3.2479 2751
VC_A0310	hypothetical protein	174	2	0	0	6.4429 435
VC_A0321	hypothetical protein	544	27	5.53E-113	1.00E-111	4.3325 7534

VC_A0329	hypothetical protein	93	8	7.91E-25	7.88E-24	3.5391 5881
VC_A0335	hypothetical protein	4685	516	2.46E-91	4.32E-90	3.1826 0608
VC_A0348	conserved hypothetical protein	694	31	1.05E-272	2.79E-271	4.4845 9554
VC_A0352	hypothetical protein	2113	316	3.36E-31	3.87E-30	2.7412 963
VC_A0359	plasmid stabilization element ParE, putative	1014	245	3.39E-18	2.89E-17	2.0492 04
VC_A0362	hypothetical protein	7734	1691	2.26E-69	3.76E-68	2.1933 381
VC_A0369	hypothetical protein	361	28	1.33E-76	2.27E-75	3.6885 001
VC_A0370	hypothetical protein	224	22	7.34E-52	1.06E-50	3.3479 233
VC_A0424	hypothetical protein	250	44	1.02E-15	8.12E-15	2.5063 5267
VC_A0429	hypothetical protein	2316	562	2.15E-15	1.68E-14	2.0429 9322
VC_A0432	hypothetical protein	75	10	2.61E-14	2.02E-13	2.9068 906
VC_A0434	hypothetical protein	3343	756	1.28E-25	1.32E-24	2.1446 8522
VC_A0448	hypothetical protein	104	14	1.04E-16	8.55E-16	2.8930 848
VC_A0466	hypothetical protein	2733	11	0	0	7.9568 3813
VC_A0503	conserved hypothetical protein	682	130	1.95E-18	1.69E-17	2.3912 6012
VC_A0515	hypothetical protein	3054	292	3.24E-92	5.75E-91	3.3866 5979
VC_A0531	sensor histidine kinase	336	47	1.36E-62	2.12E-61	2.8377 2857
VC_A0620	thiosulfate sulfurtransferase SseA, putative	125	30	3.14E-09	1.93E-08	2.0588 9369
VC_A0685	iron(III) ABC transporter, periplasmic iron-compound-binding protein	203	27	6.24E-37	7.90E-36	2.9104 4842
VC_A0686	iron(III) ABC transporter, permease protein	95	14	9.12E-29	9.99E-28	2.7625 0069
VC_A0687	iron(III) ABC transporter, ATP-binding protein	113	4	2.36E-239	5.97E-238	4.8201 7896
VC_A0713	hypothetical protein	73	6	2.46E-19	2.23E-18	3.6048 6206
VC_A0732	conserved hypothetical protein	152	3	0	0	5.6629 6501
VC_A0777	conserved hypothetical protein	263	50	3.00E-18	2.58E-17	2.3950 628
VC_A0874	hypothetical protein	1126	103	8.28E-79	1.44E-77	3.4504 9058
VC_A0879	hypothetical protein	2435	30	0	0	6.3428 1546
VC_A0892	hypothetical protein	101	24	1.25E-05	5.95E-05	2.0732 4898
VC_A0928	hypothetical protein	219	36	1.33E-17	1.12E-16	2.6048 6206
VC_A0945	maltose ABC transporter, periplasmic maltose-binding protein	119	24	3.76E-14	2.89E-13	2.3098 5526
VC_A0978	amino acid ABC transporter, periplasmic amino acid-binding protein, putative	100	21	1.25E-08	7.49E-08	2.2515 3877
VC_A0997	hypothetical protein	65	11	1.84E-12	1.32E-11	2.5629 3619
VC_A1007	hypothetical protein	66	5	2.86E-31	3.32E-30	3.7224 6602
VC_A1028	maltoporin	114	28	3.05E-10	1.95E-09	2.0255 3509

Table 6.2 - Genes upregulated in 7PET relative to Classical (*i.e.*, $\log_2$FC ≥ 2) at 30 ˚C. Genes that were previously identified as differentially-regulated between 7PET and Classical isolates at 30 ˚C [470] are indicated in blue. Two copies of *rstR* are present in the N16961 reference genome (Figure 4.2).

Gene ID	Product	Expression.7PET	Expression.Cl	pValue	qValue	log2Ratio
	Chromosome 1					
VC_0160	hypothetical protein	91	659	2.70E-37	5.68E-36	-2.85634
VC_0198	hypothetical protein	44	712	0	0	-4.0163018
VC_0204	conserved hypothetical protein	30	236	3.28E-85	8.65E-84	-2.9757525
VC_0292	hypothetical protein	40	168	7.63E-07	6.72E-06	-2.0703893
VC_0294	hypothetical protein	21	147	1.05E-16	1.54E-15	-2.8073549
VC_0314	hypothetical protein	843	7664	1.49E-150	4.59E-149	-3.184493
VC_0388	hypothetical protein	50	964	1.23E-251	4.34E-250	-4.2690331
VC_0458	conserved hypothetical protein	31	171	2.14E-17	3.21E-16	-2.4636562
VC_0485	pyruvate kinase I	101	464	1.10E-25	2.01E-24	-2.1997695
VC_0621	hypothetical protein	85	416	3.70E-31	7.10E-30	-2.2910488
VC_0713	hypothetical protein	768	6871	1.12E-135	3.34E-134	-3.1613419
VC_0765	conserved hypothetical protein	69	351	2.38E-22	4.12E-21	-2.3468028
VC_0824	TagD protein	99	765	5.50E-127	1.60E-125	-2.9499593
VC_0829	toxin co-regulated pilus biosynthesis protein B	16	78	3.73E-24	6.62E-23	-2.2854022
VC_0830	toxin co-regulated pilus biosynthesis protein Q	2	132	0	0	-6.0443941
VC_0831	toxin co-regulated pilus biosynthesis outer membrane protein C	20	109	4.16E-37	8.69E-36	-2.4462562
VC_0832	toxin co-regulated pilus biosynthesis protein R	9	195	3.98E-282	1.47E-280	-4.4374053
VC_0833	toxin co-regulated pilus biosynthesis protein D	4	182	0	0	-5.5077946
VC_0834	toxin co-regulated pilus biosynthesis protein S	9	340	0	0	-5.2394659
VC_0835	toxin co-regulated pilus biosynthesis protein T	13	156	2.64E-177	8.59E-176	-3.5849625
VC_0837	toxin co-regulated pilus biosynthesis protein F	43	1484	0	0	-5.1090106
VC_0838	TCP pilus virulence regulatory protein	5	66	5.95E-80	1.53E-78	-3.722466
VC_0839	leader peptidase TcpJ	16	83	7.15E-18	1.08E-16	-2.3750394
VC_0841	accessory colonization factor AcfC	26	330	2.31E-225	7.89E-224	-3.6658825
VC_0843	TagE protein	39	2266	0	0	-5.8605299
VC_0844	accessory colonization factor AcfA	14	968	0	0	-6.1115083
VC_0939	hypothetical protein	16	97	1.19E-18	1.83E-17	-2.5999128
VC_1233	conserved hypothetical protein	15	68	7.53E-09	7.86E-08	-2.1805722
VC_1290	DNA polymerase III, epsilon subunit, putative	49	211	5.01E-24	8.86E-23	-2.1063893
VC_1324	hypothetical protein	595	60038	0	0	-6.6568424
VC_1325	galactoside ABC transporter, periplasmic D-galactose/D-glucose-binding protein	12	365	0	0	-4.9267902
VC_1326	hypothetical protein	491	37233	0	0	-6.244715
VC_1327	galactoside ABC transporter, ATP-binding protein	8	80	3.10E-85	8.23E-84	-3.3219281
VC_1456	cholera enterotoxin, B subunit	34	465	4.56E-184	1.50E-182	-3.7736241
VC_1457	cholera enterotoxin, A subunit	27	462	0	0	-4.0968615
VC_1459	accessory cholera enterotoxin	78	455	5.86E-40	1.27E-38	-2.5443205
VC_1462	RstB2 protein	70	435	1.68E-49	3.94E-48	-2.6355886
VC_1463	RstA2 protein	24	101	2.61E-19	4.12E-18	-2.073249
VC_1554	glycerophosphoryl diester phosphodiesterase, putative	73	296	1.08E-28	2.04E-27	-2.0196288
VC_1569	hypothetical protein	18	86	4.68E-09	4.98E-08	-2.2563398
VC_1593	GGDEF family protein	91	665	8.83E-71	2.22E-69	-2.8694159
VC_1650	collagenase	56	320	4.66E-41	1.02E-39	-2.5145732
VC_1772	hypothetical protein	1016	13882	0	0	-3.7722431
VC_1774	conserved hypothetical protein	14	82	9.52E-32	1.84E-30	-2.5501971
VC_1776	N-acetylneuraminate lyase, putative	23	111	2.25E-23	3.93E-22	-2.2708539
VC_1780	hypothetical protein	271	7815	0	0	-4.8498811
VC_1825	transcriptional regulator	31	135	4.04E-21	6.77E-20	-2.1226193
VC_2083	zinc ABC transporter, permease protein	88	482	3.96E-65	9.82E-64	-2.4534577
VC_2306	hypothetical protein	1285	7697	7.70E-44	1.74E-42	-2.5825279
VC_2495	hypothetical protein	59	634	3.80E-73	9.70E-72	-3.425696

VC_2752	hypothetical protein	54	794	1.24E-143	3.78E-142	-3.8781077
		Chromosome 2				
VC_A0086	hypothetical protein	53	559	2.81E-117	5.29E-116	-3.398784
VC_A0087	hypothetical protein	118	824	3.51E-43	4.67E-42	-2.8038575
VC_A0128	ribose ABC transporter, ATP-binding protein	6	39	7.45E-21	6.97E-20	-2.7004397
VC_A0130	ribose ABC transporter, periplasmic D-ribose-binding protein	21	147	8.92E-42	1.16E-40	-2.8073549
VC_A0163	conserved hypothetical protein	26	142	1.99E-12	1.41E-11	-2.4493074
VC_A0480	hypothetical protein	287	1287	2.26E-29	2.49E-28	-2.1648894
VC_A0544	conserved hypothetical protein	114	665	1.09E-65	1.75E-64	-2.5443205
VC_A0582	conserved hypothetical protein	36	156	2.10E-13	1.57E-12	-2.1154772
VC_A0587	conserved hypothetical protein	36	206	4.04E-20	3.76E-19	-2.5165755
VC_A0665	C4-dicarboxylate transporter, anaerobic	108	536	3.06E-43	4.11E-42	-2.3112017
VC_A0670	hypothetical protein	230	186143	0	0	-9.6605618
VC_A0671	hypothetical protein	38	223	4.11E-51	5.87E-50	-2.5529724
VC_A0707	regulatory protein UhpC, putative	15	60	3.02E-12	2.13E-11	-2
VC_A0743	conserved hypothetical protein	7	451	0	0	-6.0096287
VC_A0744	glycerol kinase	45	186	4.20E-28	4.54E-27	-2.0473057
VC_A0750	hypothetical protein	237	5266	0	0	-4.4737486
VC_A0857	hypothetical protein	13	137	1.51E-30	1.71E-29	-3.3975924
VC_A0914	hemin ABC transporter, permease protein, putative	30	157	1.96E-30	2.21E-29	-2.3877302
VC_A0915	hemin ABC transporter, ATP-binding protein HutD	340	1652	9.95E-36	1.22E-34	-2.280607
VC_A0986	conserved hypothetical protein	88	474	1.99E-53	2.93E-52	-2.4293116
VC_A0989	conserved hypothetical protein	139	1238	2.78E-185	6.31E-184	-3.1548545
VC_A1003	hypothetical protein	344	1856	5.02E-51	7.12E-50	-2.4317162
VC_A1041	phosphomannomutase, putative	373	1896	1.10E-27	1.17E-26	-2.3457114
VC_A1062	putrescine-ornithine antiporter	9	36	7.85E-08	4.49E-07	-2

Table 6.3 - Genes upregulated in Classical relative to 7PET (*i.e.*, $\log_2$FC ≥ 2) at 30 ˚C. Genes that were previously identified as differentially-regulated between 7PET and Classical isolates at 30 ˚C [470] are indicated in blue. rRNA loci have been deleted from these results.

In each sample, the vast majority of reads mapped to loci on chromosome 1 rather than chromosome 2 (Table 6.1), suggesting that chromosome 1 is more transcriptionally active than chromosome 2. A total of 110 genes were found to be significantly elevated in expression in 7PET relative to Classical (67 on chromosome 1 and 43 on chromosome 2) and 82 genes were significantly elevated in expression in Classical relative to 7PET (58 on chromosome 1, 24 on chromosome 2) (Tables 6.2, 6.3). However, the overlap between these results and those of Beyhan and colleagues was limited. Behyan *et al* identified 270 genes that were elevated in expression in their Classical strain relative to their El Tor strain, and 252 genes that were elevated in their El Tor strain relative to their Classical strain. However, of these genes, just 11 overlap with our results from 7PET, and 16 overlap with our Classical isolate (Tables 6.2, 6.3). This may reflect strain-specific differences, the differences in growth media between the experiments, or differences in sensitivity between microarray-based analyses and this RNA-seq experiment.

In spite of this, it is notable that the differential expression of genes encoded by CTXφ and VPI-1 between 7PET and Classical was recapitulated in our experiment. This included the expression of the *rstR* gene in 7PET (encoding the RstR repressor of CTXφ genes, and under "normal growth conditions", this is the only CTXφ protein produced by lysogens [99]) and the expression of *ctxAB*, *tcp*, and *acf* virulence genes in Classical cultures under these *in vitro* conditions. This is consistent with previous studies discussed in Chapter 5 (section 5.3.8.2) which showed that genetic determinants such as P_{aphA} and P_{tcpPH} mutations drive overexpression of the ToxT regulon in Classical bacteria *in vitro*. Further experimentation and validation is required to determine whether the genes that were differentially regulated in this study but not in that of Beyhan *et al* are due to differences in culture conditions between these studies, or whether they reflect strain-specific variation in gene expression.

We then determined the sets of genes that were differentially expressed by these strains when cultured at 37 ˚C, rather than at 30 ˚C. All other experimental conditions were kept constant with the 30 ˚C experiment. The results are illustrated in volcano plots (Figure 6.2) and significantly differentially regulated genes are listed in Tables 6.4 and 6.5.

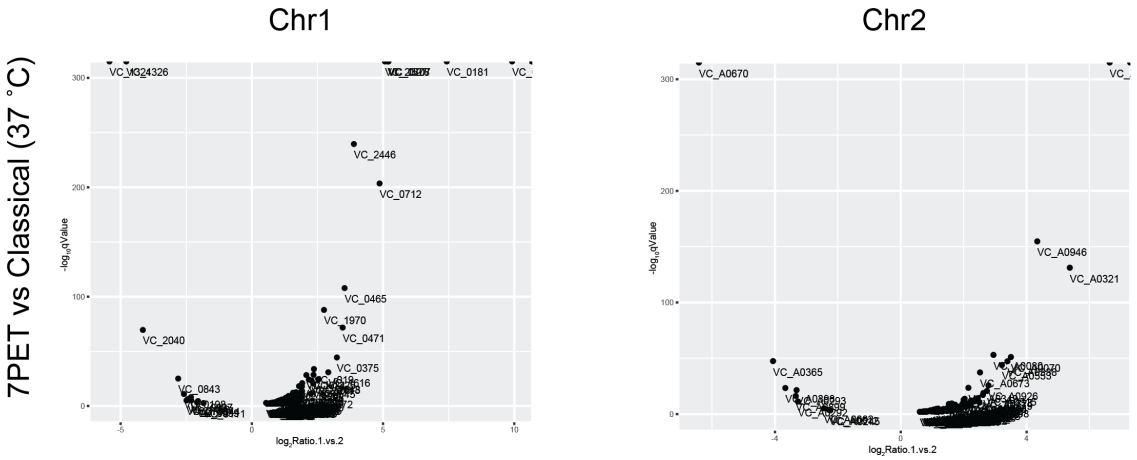

Figure 6.2 – Volcano plots comparing gene expression in 7PET and Classical strains at 37 ˚C. Genes that were not detected in one strain or the other are not represented in these plots (log_2FC = infinite). Genes which passed statistical thresholds described in section 2.1.18 are listed in Tables 6.4 and 6.5.

Gene ID	Product	Expression.7 PET	Expressio n.Cl	pValue	qValue	log2Rati o
		Chromosome 1				
VC_0175	deoxycytidylate deaminase-related protein	185	0	0	0	Inf
VC_0176	transcriptional regulator, putative	105	0	0	0	Inf
VC_0179	hypothetical protein	72	0	0	0	Inf
VC_0180	conserved hypothetical protein	87	0	0	0	Inf
VC_0181	conserved hypothetical protein	172	1	0	0	7.42626 4755
VC_0182	hypothetical protein	212	0	0	0	Inf
VC_0183	hypothetical protein	131	0	0	0	Inf
VC_0184	hypothetical protein	223	0	0	0	Inf
VC_0185	transposase, putative	204	0	0	0	Inf
VC_0375	hypothetical protein	236	25	1.22E-46	4.07E-45	3.23878 686
VC_0465	tyrosyl-tRNA synthetase	81	7	3.61E-110	1.50E-108	3.53249 5081
VC_0471	SprT protein, putative	88	8	5.14E-74	1.90E-72	3.45943 1619
VC_0490	conserved hypothetical protein	133	0	0	0	Inf
VC_0491	hypothetical protein	99	0	0	0	Inf
VC_0492	hypothetical protein	120	0	0	0	Inf
VC_0493	hypothetical protein	130	0	0	0	Inf
VC_0495	conserved hypothetical protein	164	0	0	0	Inf
VC_0496	hypothetical protein	77	0	0	0	Inf
VC_0502	type IV pilin, putative	345	0	0	0	Inf
VC_0503	conserved hypothetical protein	175	0	0	0	Inf
VC_0504	hypothetical protein	1705	0	0	0	Inf
VC_0505	hypothetical protein	966	1	0	0	9.91587 9379
VC_0506	hypothetical protein	529	0	0	0	Inf
VC_0507	hypothetical protein	359	0	0	0	Inf
VC_0508	hypothetical protein	143	0	0	0	Inf
VC_0509	hypothetical protein	112	0	0	0	Inf
VC_0510	DNA repair protein RadC-related protein	104	0	0	0	Inf
VC_0512	methyl-accepting chemotaxis protein	119	0	0	0	Inf
VC_0516	phage integrase	110	0	0	0	Inf
VC_0707	hypothetical protein	369	10	0	0	5.20554 8911
VC_0712	hypothetical protein	175	6	7.54E-206	3.43E-204	4.86624 8611
VC_1225	hypothetical protein	719	141	8.88E-31	2.61E-29	2.35029 6608
VC_1452	RstC protein	256	0	0	0	Inf
VC_1572	hypothetical protein	76	16	2.03E-13	4.89E-12	2.24792 7513
VC_1578	hypothetical protein	108	3	0	0	5.16992 5001
VC_1616	glutaredoxin, putative	143	19	4.80E-33	1.42E-31	2.91194 3823
VC_1648	hypothetical protein	633	109	8.75E-27	2.50E-25	2.53787 7365
VC_1970	benzoate transport protein	524	78	3.61E-90	1.41E-88	2.74802 0783
VC_2327	hypothetical protein	1983	59	0	0	5.07082 5913
VC_2446	tRNA nucleotidyltransferase	222	15	6.09E-242	2.93E-240	3.88752 5271
VC_2545	inorganic pyrophosphatase	172	34	1.31E-22	3.55E-21	2.33880 1913
VC_2566	conserved hypothetical protein	1526	339	2.36E-26	6.63E-25	2.17039 7784
VC_2611	hypothetical protein	1410	291	1.92E-25	5.32E-24	2.27660 4104
VC_2613	phosphoribulokinase	87	15	2.27E-26	6.43E-25	2.53605 29
		Chromosome 2				
VC_A0028	hypothetical protein	179	39	2.31E-11	2.93E-10	2.19841 356

VC_A0070	phosphate ABC transporter, periplasmic phosphate-binding protein	68	6	2.91E-53	8.46E-52	3.50250034
VC_A0080	GGDEF family protein	162	21	3.38E-55	1.00E-53	2.94753258
VC_A0168	hypothetical protein	164	36	7.49E-12	9.64E-11	2.187627
VC_A0315	hypothetical protein	799	8	0	0	6.64205169
VC_A0321	hypothetical protein	83	2	1.87E-133	7.60E-132	5.37503943
VC_A0323	conserved hypothetical protein	148	0	0	0	Inf
VC_A0328	biphenyl-2,3-diol 1,2-dioxygenase III-related protein	75	16	1.84E-09	2.04E-08	2.22881869
VC_A0335	hypothetical protein	957	233	1.12E-10	1.35E-09	2.03818897
VC_A0338	conserved hypothetical protein	63	6	1.75E-49	4.74E-48	3.39231742
VC_A0348	conserved hypothetical protein	263	49	5.93E-16	9.35E-15	2.42420915
VC_A0349	RelB protein	2128	479	1.46E-25	3.09E-24	2.15140059
VC_A0515	hypothetical protein	448	67	1.02E-22	2.07E-21	2.74126573
VC_A0527	conserved hypothetical protein	95	19	7.37E-14	1.06E-12	2.32192809
VC_A0531	sensor histidine kinase	131	32	1.42E-15	2.17E-14	2.033423
VC_A0553	hypothetical protein	223	24	4.55E-46	1.19E-44	3.2159374
VC_A0660	hypothetical protein	85	20	2.55E-08	2.66E-07	2.08746284
VC_A0673	transcriptional regulator, LacI family	373	65	2.15E-39	5.35E-38	2.52066401
VC_A0686	iron(III) ABC transporter, permease protein	84	21	1.24E-14	1.86E-13	2
VC_A0698	hypothetical protein	102	18	4.29E-12	5.62E-11	2.50250034
VC_A0719	sensor histide kinase	124	30	5.04E-12	6.54E-11	2.04730571
VC_A0774	UDP-glucose 4-epimerase	105	17	7.89E-21	1.54E-19	2.62678268
VC_A0879	hypothetical protein	159	26	4.26E-19	7.83E-18	2.61244324
VC_A0924	conserved hypothetical protein	66	14	4.92E-14	7.16E-13	2.2370392
VC_A0926	transcriptional regulator, AraC/XylS family	76	11	9.84E-28	2.17E-26	2.78849589
VC_A0946	maltose/maltodextrin ABC transporter, ATP-binding protein	81	4	4.12E-157	1.87E-155	4.33985
VC_A0994	hypothetical protein	93	17	6.80E-16	1.06E-14	2.45169597
VC_A1061	hypothetical protein	127	25	3.09E-13	4.42E-12	2.3448285

Table 6.4 - Genes upregulated in 7PET relative to Classical (*i.e.*, $\log_2FC \geq 2$) at 37 ˚C. Note that pathogenicity islands VSP-1 (*VC_0174-0186*) and VSP-2 (*VC_0489-0517*) are present and transcribed in 7PET isolates, but not in Classical isolates. rRNA and tRNA loci have been deleted from these results.

Gene ID	Product	Expression .7PET	Expression .Cl	pValue	qValue	log2Ratio
		Chromosome 1				
VC_0198	hypothetical protein	76	457	4.80E-13	1.15E-11	-2.588122842
VC_0733	hypothetical protein	16	90	4.35E-07	7.64E-06	-2.491853096
VC_0837	toxin co-regulated pilus biosynthesis protein F	47	235	3.41E-10	7.35E-09	-2.321928095

VC_0843	TagE protein	84	588	2.90E-27	8.39E-26	-2.807354922
VC_0844	accessory colonization factor AcfA	61	253	1.80E-06	2.91E-05	-2.052256237
VC_0939	hypothetical protein	37	157	4.23E-05	0.00057203	-2.085167383
VC_1324	hypothetical protein	118	5072	0	0	-5.425695981
VC_1326	hypothetical protein	161	4439	0	0	-4.785102115
VC_1456	cholera enterotoxin, B subunit	17	92	2.44E-08	4.69E-07	-2.436099115
VC_1457	cholera enterotoxin, A subunit	23	114	4.96E-07	8.62E-06	-2.309328058
VC_2040	conserved hypothetical protein	5	89	7.88E-72	2.84E-70	-4.153805336
	Chromosome 2					
VC_A0245	PTS system, IIA component	14	66	7.39E-05	0.00051408	-2.2370392
VC_A0292	hypothetical protein	8	77	4.50E-13	6.26E-12	-3.2667865
VC_A0293	hypothetical protein	13	129	1.85E-23	3.81E-22	-3.3107875
VC_A0365	hypothetical protein	6	100	1.36E-49	3.74E-48	-4.0588937
VC_A0398	hypothetical protein	7	89	2.35E-25	4.91E-24	-3.6683785
VC_A0512	anaerobic ribonucleoside-triphosphate reductase activating protein	21	102	2.68E-05	0.00020681	-2.2801079
VC_A0662	CbbY family protein	14	76	7.44E-07	7.14E-06	-2.4405726
VC_A0670	hypothetical protein	281	23993	0	0	-6.4158996
VC_A0899	hypothetical protein	18	182	3.46E-18	6.15E-17	-3.3378696

Table 6.5 - Genes upregulated in Classical relative to 7PET (*i.e.*, $\log_2$FC ≤ 2) at 37 ˚C. Genes encoded on VPI-1 and CTXφ were up-regulated (*VC_0809-VC_0848* and *VC_1457-1465*, respectively).

Once again, virulence genes that are members of the ToxT regulon were expressed in Classical at a higher level than in 7PET at 37 ˚C, consistent with the observations made at 30 ˚C. This included *ctxAB*, *tcp* and *acf* genes, and the *tagE* gene. It was also evident that at 37 ˚C, genes encoded by VSP-1 and VSP-2 were transcribed in 7PET (Table 6.4). These genes are absent from Classical *V. cholerae* and from MJD1402 (Figures 5.7, 5.22), and therefore their expression was detected as having an infinite $\log_2$(expression ratio). These genes were not determined to be significantly transcribed at 30 ˚C (Tables 6.2, 6.3). We also noted that in spite of rRNA depletion using RiboZero (Methods, section 2.2.18), rRNA loci were identified as being transcribed at significantly different levels in 7PET and Classical isolates (data not shown). This indicates that RiboZero depletion in these samples, or across the experiment more generally, was not totally efficient, and emphasises that this is a depletion rather than an elimination step.

6.3.4 – Effect of temperature on transcriptome of Classical and 7PET isolates

Using these data, a comparison was then performed to identify genes that were differentially expressed at different temperatures (*i.e.*, comparing the 37 ˚C and 30 ˚C results for one strain to one another). At 37 ˚C, 122 genes were upregulated in 7PET relative to Classical (81 on chromosome 1, 41 on chromosome 2) and 245 genes were upregulated in Classical relative to 7PET (154 on chromosome 1, 91 on chromosome 2). At 30 ˚C, 93 genes were upregulated in 7PET relative to Classical (59 on chromosome 1, 34 on chromosome 2), and 105 genes were upregulated in Classical relative to 7PET (67 on chromosome 1, 38 on chromosome 2). Sixty-two genes were upregulated in 7PET relative to Classical at both 37 and 30 ˚C (42 on chromosome 1, 20 on chromosome 2), and 32 genes were upregulated in Classical relative to 7PET at both temperatures (21 on chromosome 1, 11 on chromosome 2). These results are summarised in Figure 6.3; the raw data are not shown.

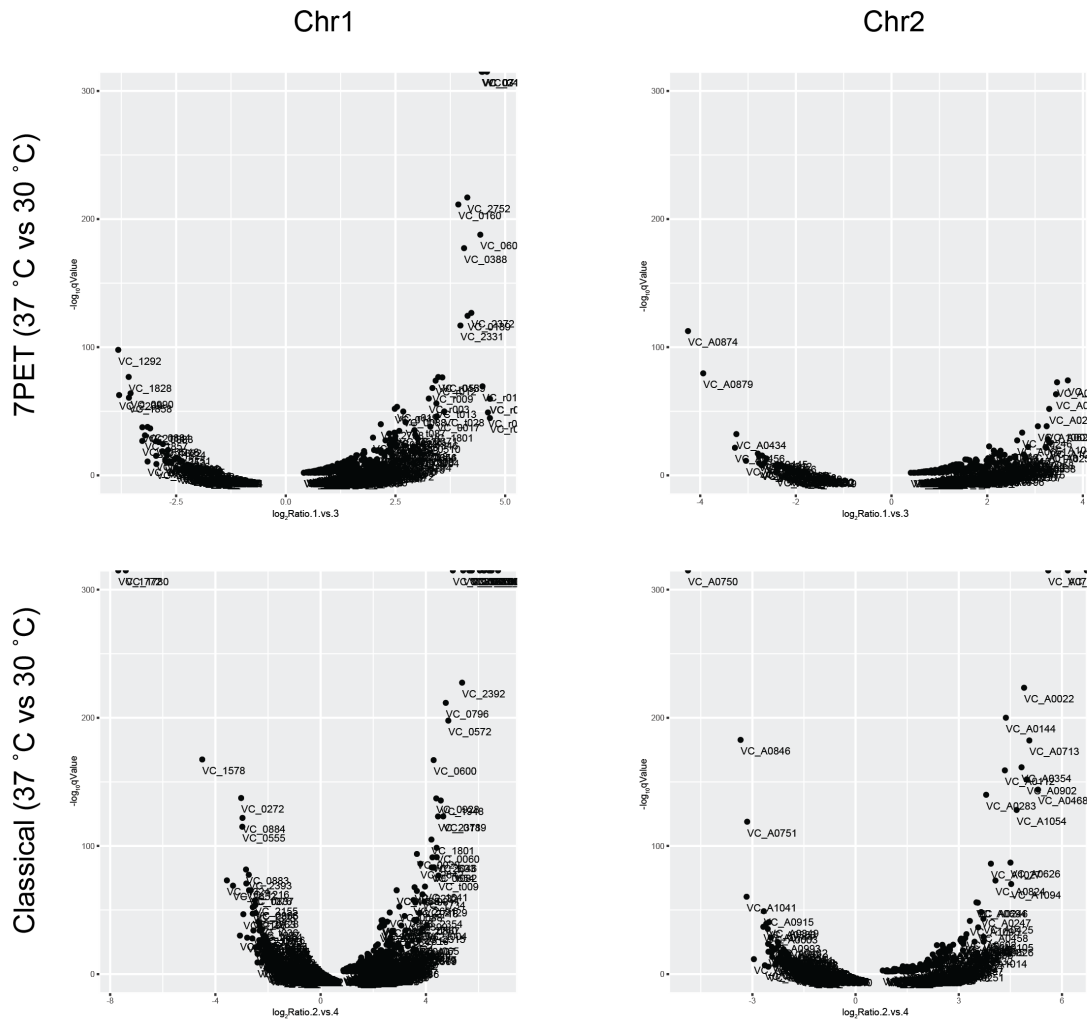

Figure 6.3 – Volcano plot comparing genes expressed at 37 and 30 ˚C in Classical and 7PET strains. Genes that were not detected in one strain or the other are not represented in these plots (log_2FC = infinite).

One of the most striking expression differences amongst these data was the effect of temperature on the expression of a subset of genes on VPI-2 in Classical *V. cholerae* (Figure 6.4). These genes encode sialidase and other enzymes required to catabolise and metabolise sialic acids (section 1.2.5). This difference in transcription is illustrated in Figure 6.4, and although at 30 ˚C there is a general increase in transcription of genes on VPI-2 in the Classical isolate, these sialic acid metabolism genes were specifically and dramatically up-regulated in comparison to their surrounding genes (Figure 6.4). It is important to note that neither VPI-2 nor the *nanH*/sialidase gene clusters are considered canonically to be part of the ToxR or ToxT regulons – therefore, further investigation is required to determine whether or not the up-regulation of virulence genes and sialidase genes at 30 ˚C in Classical *V. cholerae* are due to mechanisms or signals that are independent of one another.

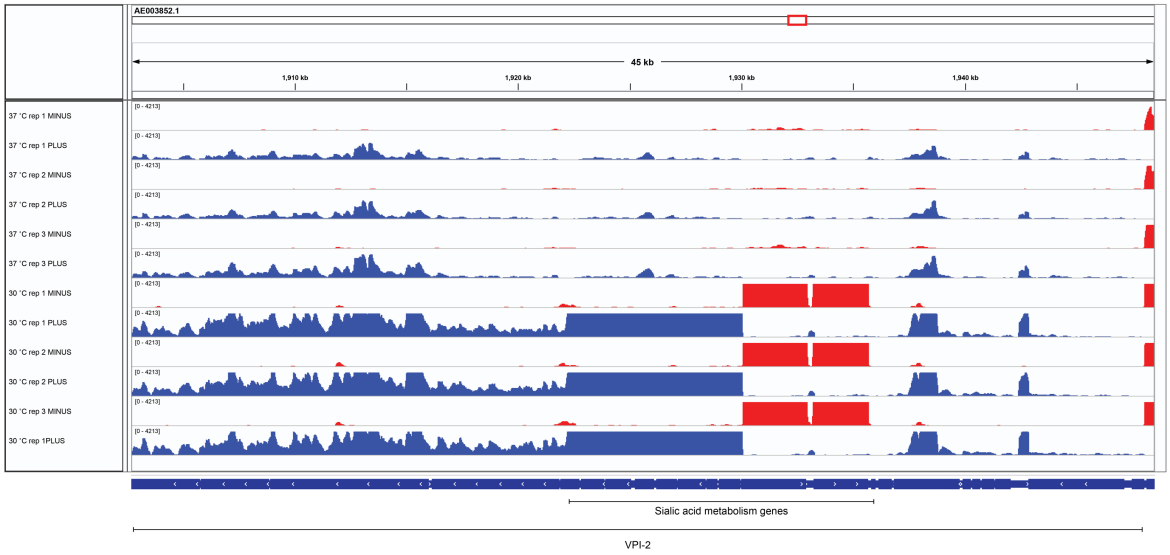

Figure 6.4 – Upregulation of sialic acid metabolisms genes in Classical *V. cholerae* at 30 ˚C relative to 37 ˚C. All lanes scaled identically.

There was also evidence that genes involved in *V. cholerae* virulence were up-regulated in Classical cultures at 30 ˚C relative to 37 ˚C under these *in vitro* conditions, exemplified by the *ctxAB* operon (illustrated in Figure 6.5). This is consistent with previous reports that have shown that culturing Classical *V. cholerae* at 30 ˚C induces virulence gene expression and the production both of CT and TCP, relative to culturing at 37 ˚C [91, 476–478]. The biological reason for this difference remains unclear. However, this has previously been shown not to be the case for 7PET bacteria, and this observation was recapitulated in our experiment – temperature variation did not cause statistically significant differential expression of virulence genes in cultures of 7PET *V. cholerae*.

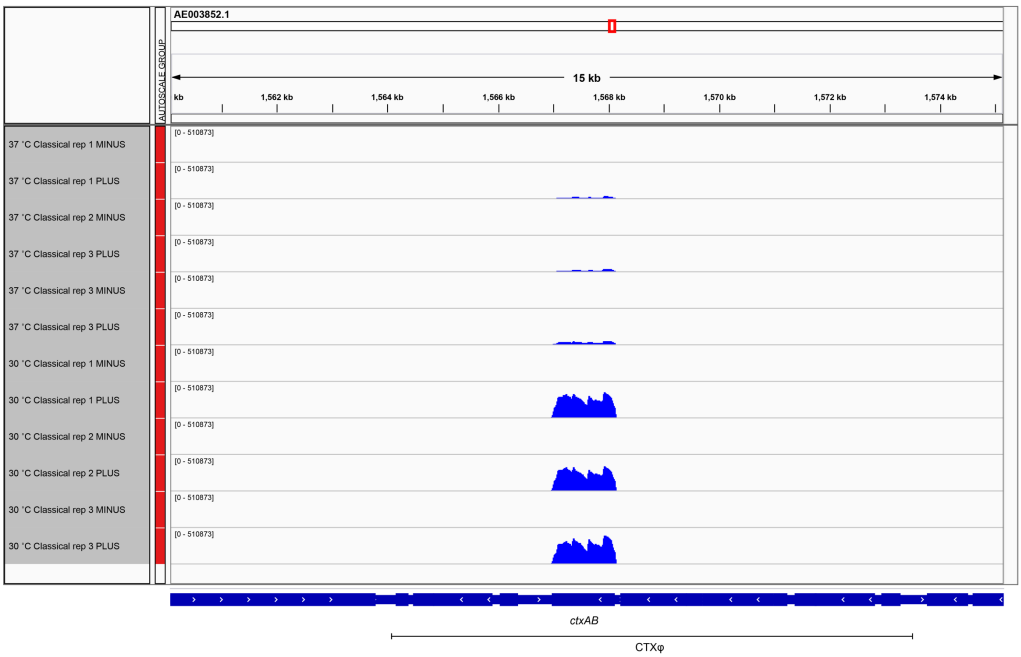

Figure 6.5 – Illustration of *ctxAB* transcript levels in Classical *V. cholerae* at 30 ˚C. The *ctxAB* operon is transcribed at very high levels at 30 ˚C in comparison to the levels at 37 ˚C. All lanes scaled identically.

6.3.5 – Selection of isolates for transcriptomic experiments

These initial experiments established that RNA-seq approaches could be used successfully with *V. cholerae* at WSI, and provided initial information about the practicalities of designing such experiments under CL3 conditions. Informed by these results, attention was then turned to extending these comparative transcriptomics methodologies to broader biological questions – specifically, investigating whether pandemic and non-pandemic lineages of *V. cholerae* O1 displayed different transcriptomic signatures when cultured under identical growth conditions.

Eight whole-genome-sequenced *V. cholerae* isolates were selected for comparative transcriptomic analysis. A number of factors were considered in choosing these isolates, including the complement of virulence genes and genomic islands which they harboured, their serogroup, the presence and absence of important promoter mutations, and their membership of key phylogenetic lineages. These factors were established in the course of the research described in Chapter 5. The strains are detailed in Table 6.6, and their phylogenetic position is indicated in Figure 6.6.

Internal ID	Other name	Serogroup	Details
MJD1402	NCTC 10256	O1	7PET. El Tor biotype. Toxigenic. Hong Kong, 1961
MJD1403	MJD474	O1	7PET. El Tor biotype. Toxigenic. UK traveller returning from Somalia, 2017
MJD1404	NCTC 10732	O1	Classical. Classical biotype. 1952
MJD1405	NCTC 5596	O1	Classical. Classical biotype. Pre-1939
MJD1406	A213	O1	Gulf Coast. El Tor biotype. Non-toxigenic. Georgia, 1984
MJD1407	A219	O1	Gulf Coast. El Tor biotype. Toxigenic. Georgia, 1986
MJD1408	MJD462	Non-O1	Related to MS-6. UK traveller returning from Thailand, 2017
MJD1409	NCTC 9422	O1	Non-7PET. El Tor biotype. Non-toxigenic. Pre-1955

Table 6.6 – Strains used in transcriptomic experiments. Each of these glycerol stocks were prepared from a single well-isolated colony taken from the original culture of each strain that was frozen at WSI. Isolation dates listed where known. Live isolates also listed in Table 2.1.

Two 7PET isolates were included, one from wave 1 (MJD1402), and a recent travel-associated isolate that phylogenetic analysis has shown to be a member of wave 3 (MJD1043). Wave 1 corresponds to the initial outbreak of the seventh cholera pandemic, and wave 3 to more recent pandemic cholera [234]. Both of these isolates are toxigenic, and harbour *ctxB3* and *ctxB7* alleles of *ctxB*, respectively. Phylogenetic analysis demonstrated that MJD1403 is a member of the same sub-lineage of 7PET as the isolates which caused the Yemeni cholera epidemic in 2017 – these isolates also harbour the *ctxB7* variant [309].

Two Classical isolates, MJD1404 and MJD1405, were included in this study. Both of these are toxigenic *V. cholerae* accessioned by NCTC, and harbour the *ctxB1* variant of *ctxB*, as expected for Classical *V. cholerae* [479]. These two isolates were included in the analysis because although MJD1404 is closely-related to the O395 reference sequence (Figure 5.10), MJD1405 harbours a wild-type P_{tcpPH} allele (Figure 5.20). Given the importance of this promoter mutation in virulence gene regulation (section 5.3.8.2), this second Classical isolate was included in the experiment.

Two *V. cholerae* O1 (A213, A219) that were members of the Gulf Coast lineage were included in these experiments. These were selected because they were non-toxigenic and toxigenic, respectively. Both originated from Georgia, USA, and were isolated in the 1980s. Their genome sequences were first reported in 2011 [234]. Like other members of the Gulf Coast lineage, these isolates harbour VPI-1 and VPI-2, but lack VSP-1 and VSP-2 (section 5.3.5). They are related to, but not identical to, strain 2740-80, the commonly-described representative of the Gulf Coast lineage [54, 189, 234]

Although this experiment was designed to explore whether there were differences in gene expression between 7PET and other *V. cholerae* O1, a single non-O1 isolate was included in this analysis. This non-toxigenic isolate was included because it is very closely related to MS-6, a toxigenic serogroup O1 isolate that harbours VPI-1, VPI-2, and VSP-1 [440]. MS-6 and MJD1408 are more closely related to Gulf Coast and MX-2 lineages, and to recent non-7PET Chinese *V. cholerae* O1 [396] than they are to 7PET. By including this isolate, similarities in gene expression to that of Gulf Coast isolates might be detectable which are independent of serogroup, though such candidate results would need to be validated.

The eighth strain included in the experiments was NCTC 9422, one of the non-pandemic El Tor *V. cholerae* O1 isolates characterised in Chapter 5 (section 5.3.7.2). This isolate is non-toxigenic and lacks all of the canonical *V. cholerae* pathogenicity islands, and was included to determine the transcriptional profile exhibited by a serogroup O1 biotype El Tor isolate that is more distantly related to 7PET than the Gulf Coast isolates.

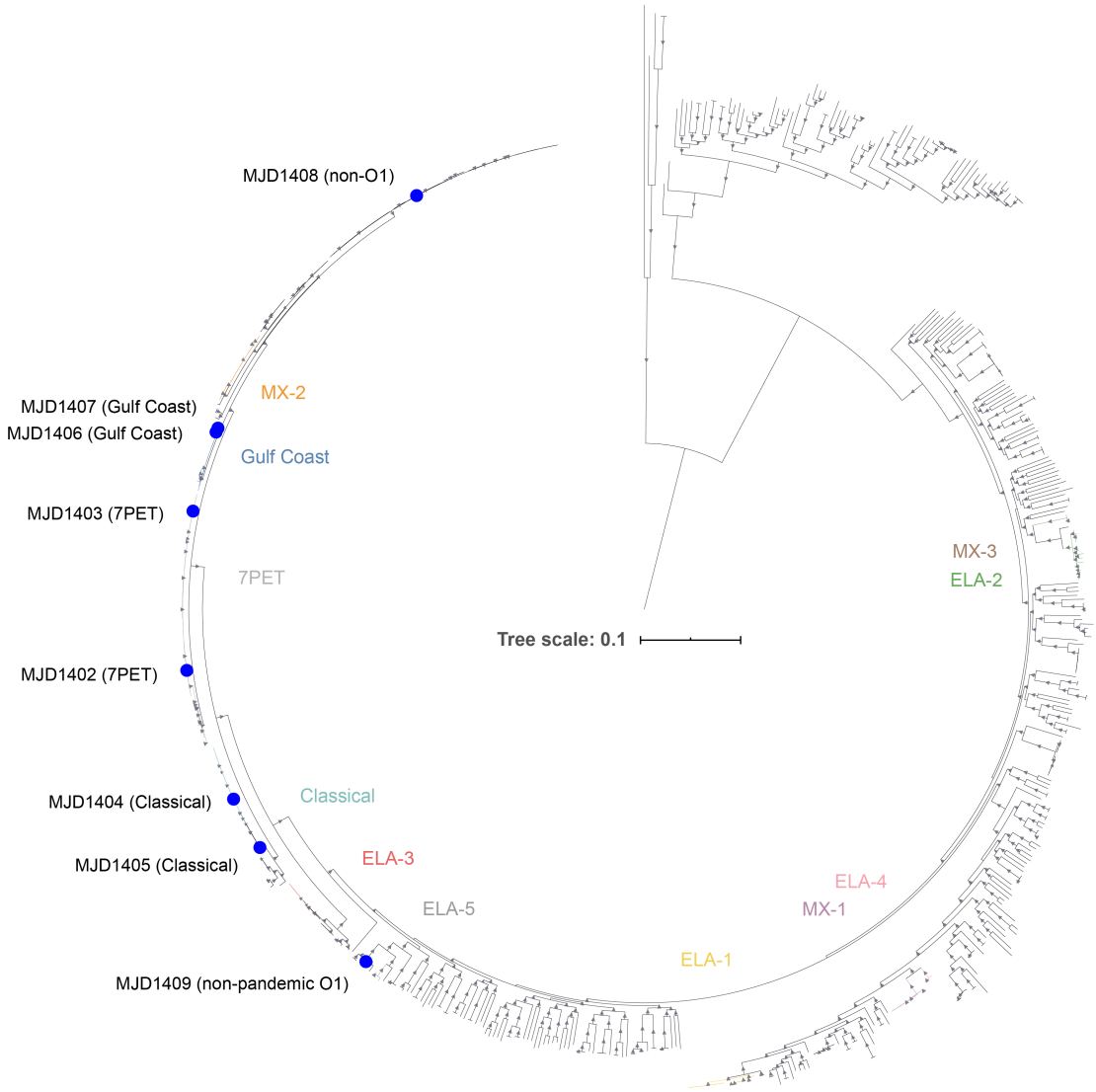

Figure 6.6 – Phylogenetic position of the live isolates chosen for this chapter research. Lineages are denoted as previously described [189]. The phylogeny presented here is the same as that used to generate Figure 5.1.

It is important to state that the chosen isolates do not represent the totality of *V. cholerae* diversity, and were chosen for the specific reasons outlined above. However, this multi-strain transcriptomic experiment was in itself a pilot experiment, designed to determine whether or not lineage-specific signatures of gene expression could be detected. This was also the maximum number of isolates that could be handled safely in one working day from a practical perspective. In order to assay gene expression across multiple strains within the constraints of working at CL3, cultures were harvested for RNA extraction after a defined period of time post-inoculation (3 hr), rather than by monitoring optical density for each culture and stopping the experiment for each strain once a defined OD_{600} value had been reached.

6.3.6 – Assaying differential gene expression in eight strains

Eight *V. cholerae* isolates were cultured at 37 ˚C in LB media, under identical conditions to those described in section 6.3.1. After 3 hr, cultures were harvested and RNA was subsequently isolated from snap-frozen cell pellets. All other conditions were kept identical to that of the pilot experiment (section 6.3.1). Sequencing data were analysed in the same way as was carried out in the pilot experiment, and genes that were statistically significantly differentially expressed were determined, pairwise, between each of the eight strains. These data have been summarised in Table 6.7.

		X							
		1402	**1403**	**1404**	**1405**	**1406**	**1407**	**1408**	**1409**
Y	**1402**	-	40/0	67/30	110/1077	60/11	55/14	61/55	409/1110
	1403		-	64/96	102/1119	49/72	42/140	1/237	369/1129
	1404			-	5/1058	43/54	47/77	44/118	259/1129
	1405				-	1045/82	1038/73	999/96	204/127
	1406					-	0/3	22/42	335/1110
	1407						-	24/40	272/1102
	1408							-	348/1077
	1409								-

Table 6.7 – Summary numbers for differentially-expressed genes across 8-way experiment. Number of genes up-regulated in isolate X relative to isolate Y / number of genes up-regulated in isolate Y compared to isolate X. Note that genes which were present in only one of these isolates were excluded from these aggregate numbers.

It was immediately apparent that MJD1405 and MJD1409 exhibited drastically different gene expression profiles to all other isolates - over 1,000 genes were expressed in these isolates relative to all other isolates in the experiment (Table 6.7). Moreover, gene expression in these two isolates was much more similar between one another than between these and other isolates, (Table 6.7). To investigate this using a complementary approach, a principal component analysis (PCA) of normalised transcripts for all 24 samples was performed to summarise the variation across all of the samples (Figure 6.7). Principal component 1 (PC1), explaining 57.3% of the variation amongst these 24 isolates, clearly separated both MJD1405 and MJD1409 from the rest of the experiment, and also showed that the three replicates for each of these isolates were dissimilar to one another (Figure 6.7). The variability within these replicates is unlikely to be explained by differences in growth stage – all cultures were harvested at the same time

post-inoculation, and OD_{600} values for each replicate were similar at the time of harvesting. It was therefore decided not to compare the data from these two strains to the others collected in this experiment.

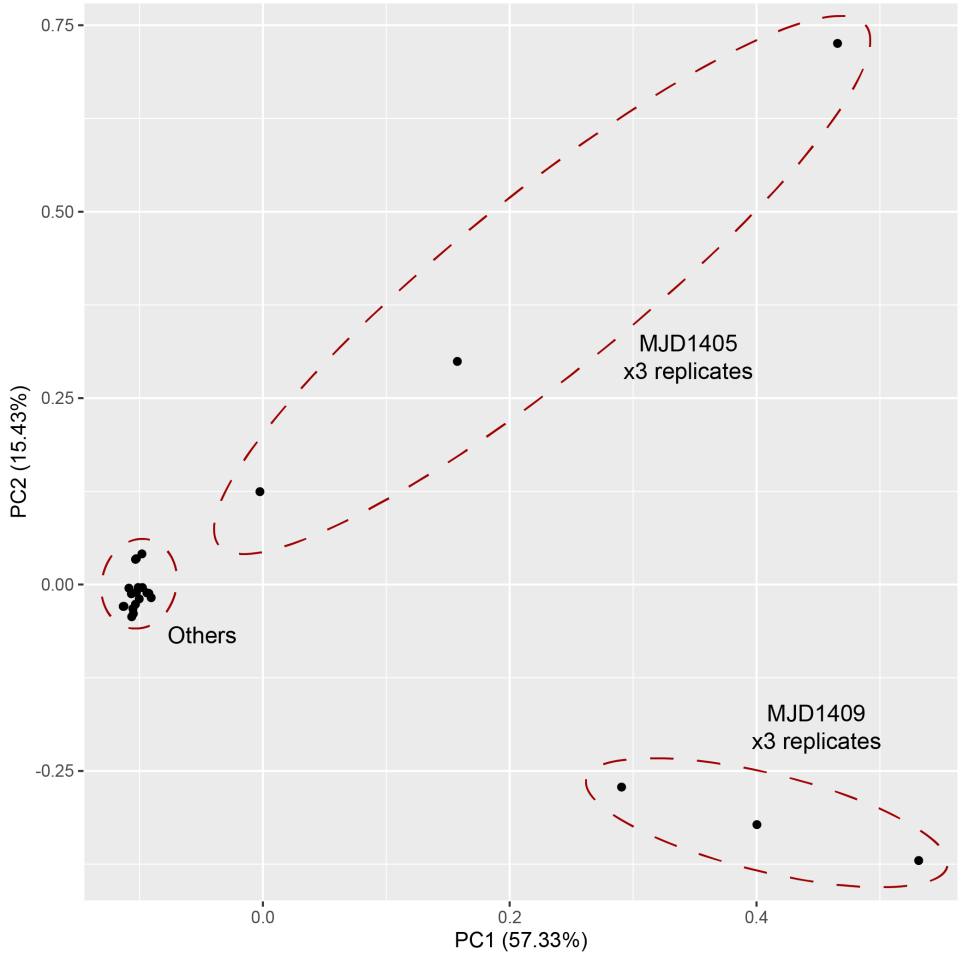

Figure 6.7 – PCA comparing all of the 24 sequenced samples in this experiment. The biological replicates of MJD1405 and MJD1409 were substantially different to the other strains in the experiment and to one another, separated by PC1. Ellipses added manually for illustrative purposes (Adobe Illustrator).

PCA was then repeated after excluding replicates of MJD1405 and MJD1409, on eighteen samples in total (Figure 6.8). In this analysis, PC1 (explaining 26.3% of the variance amongst these replicates) clearly separated Classical isolate MJD1404 from the rest of the samples. However, there appeared to be limited separation of samples along PC2; replicates of different strains were mixed amongst one another, and the three replicates of each strain did not cluster with one another in every instance. This was a surprising result, as it had been hypothesised that isolates would cluster on the basis of their genotype or membership of genetic lineage. This may reflect variation within the experiment, or that each culture had reached a slightly

different growth phase stage at the time of harvesting. This variability meant that we were conservative in selecting genes for subsequent consideration or discussion in the absence of confirmatory experiments.

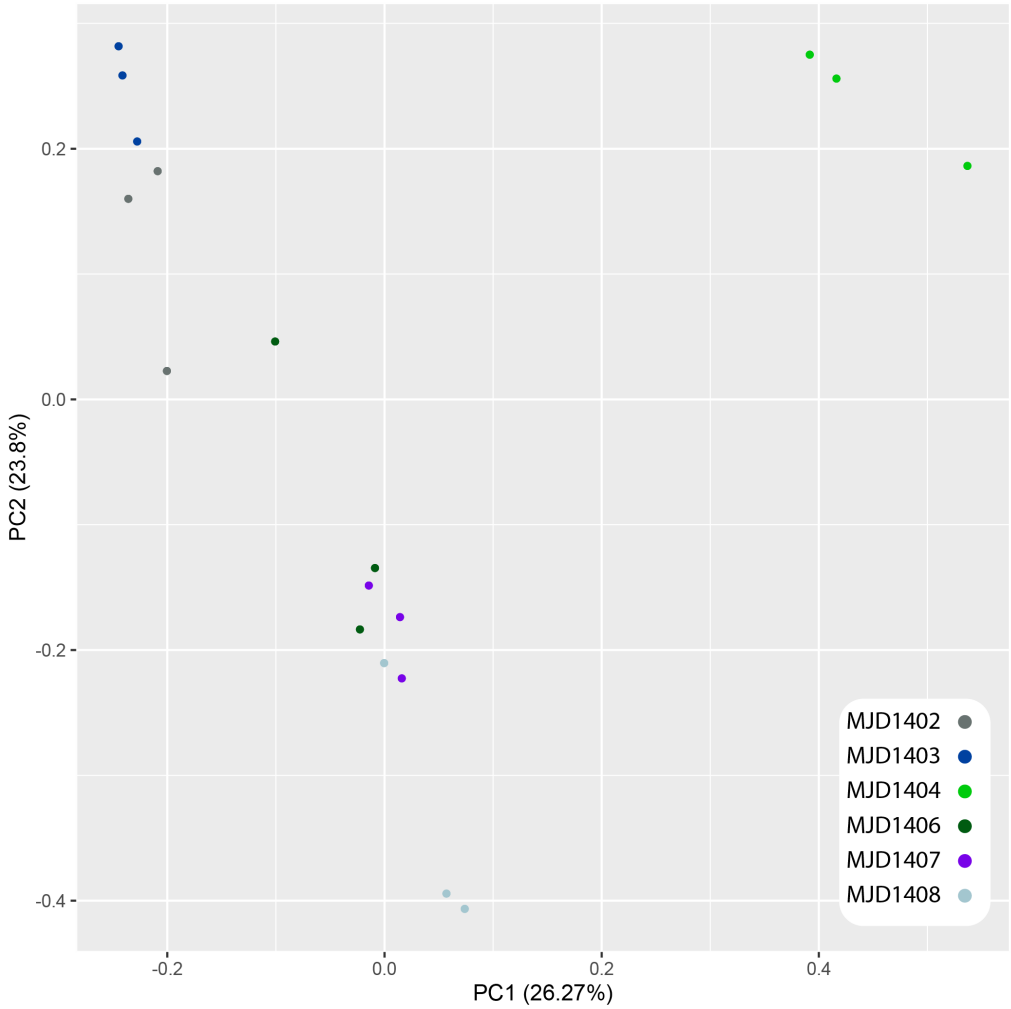

Figure 6.8 – PCA of the 18 samples remaining after the exclusion of MJD1405 and MJD1409. The biological replicates of MJD1404, the Classical isolate remaining in the experiment, were clearly separated from the other samples in the experiment by PC1 and PC2. However, the remaining samples failed to resolve into clearly-separated groups based on their genotype.

It was decided to interrogate these data to identify potentially novel biological insights. For example, to our knowledge, this experiment is the first to compare the transcriptome of Gulf Coast isolates to those of either Classical or 7PET bacteria. Amongst the two Gulf Coast strains, just three genes were significantly differentially expressed (Table 6.7). These were *VC_1865*, *VC_A0540*, and one putative CDS of unknown function. Of these, only the product

of *VC_A0540* has a known or predicted function (formate transporter). This lack of variability is consistent with one of the aims of this experiment – namely, to determine whether strains within a *V. cholerae* lineage have transcriptional profiles more similar to the profiles of other lineages.

We interrogated the genes differentially expressed amongst Gulf Coast and 7PET isolates (Table 6.7), and immediately found it to be striking that multiple components of T6SS were up-regulated in Gulf Coast isolates in comparison to 7PET isolates. The macromolecular T6SS is an apparatus that enables *V. cholerae* to prey upon bacteria and eukaryotes, by secreting and injecting effector proteins into target cells (*e.g.*, [480, 481]). Three such systems are usually encoded by three gene clusters in *V. cholerae*, and among the genes involved in producing T6SS is *hcp*, which encodes the Hcp protein (*VC_1415*, *VC_A0017*), and is considered to be a 'hallmark' of T6SS function [482]. Hcp is thought to form hexameric rings that constitute the tube through which T6SS secrete effectors [483]. These *hcp* genes, and other genes encoding structural components of *V. cholerae* T6SS, were substantially differentially expressed in these isolates, exemplified in Figure 6.9. Examining transcription profiles across all replicates and strains also showed that MJD1408 exhibited a similar *hcp* transcription profile to Gulf Coast isolates (to which it is most closely related phylogenetically, Figure 6.6).

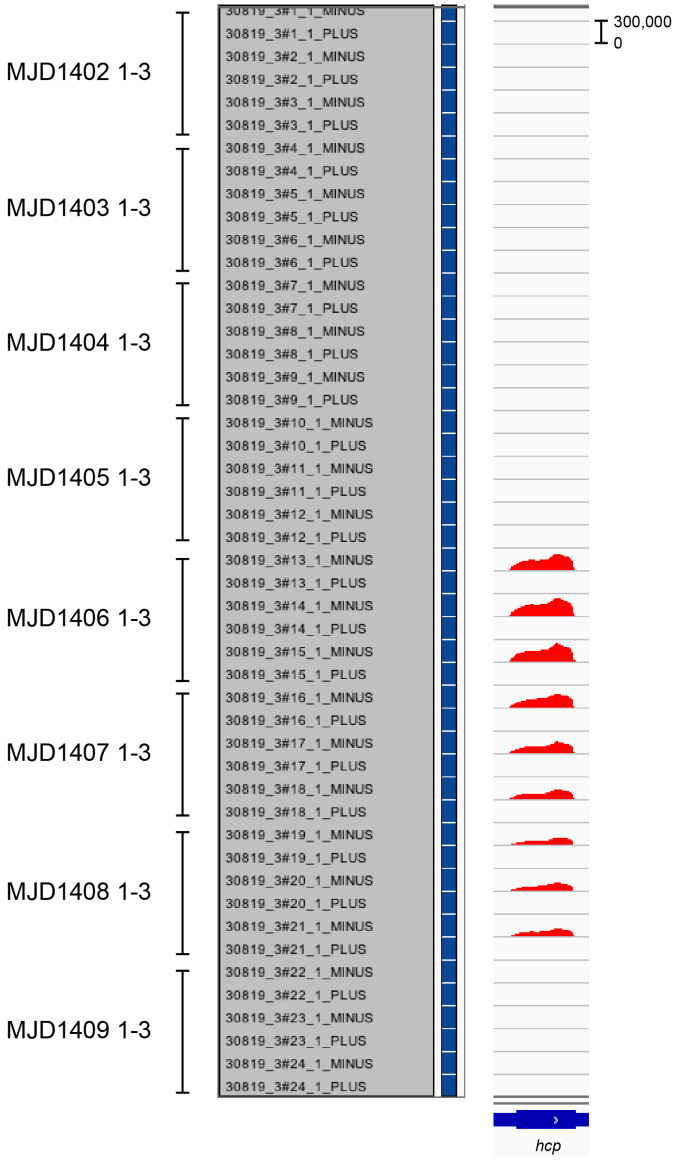

Figure 6.9 – The *hcp* gene (*VC_A0017*) is upregulated in Gulf Coast and a related strain of *V. cholerae* (MJD1408) relative to all other strains in this experiment. All tracks scaled identically.

6.4 – Discussion

The design of these comparative experiments, and these initial data, lay the groundwork for future analyses that should begin to describe lineage-specific as well as strain-specific differences in gene expression amongst *V. cholerae*. These data have shown that genes which are highly over- or under-expressed can be identified, and have also highlighted the homogeneity of expression profiles within lineages and other closely-related *V. cholerae*. Characterising aspects of gene regulation that are hallmarks of a *V. cholerae* lineage should add to our understanding of the biological differences between pandemic and non-pandemic bacterial lineages. It is also notable that these data recapitulated some, but not all, of the previously-published differences in gene expression between Classical and 7PET *V. cholerae* [470]. The genes that were detected in both studies were highly differential-expressed in both studies, and this suggests that in spite of the differences in growth conditions between these experiments, any fundamental and dramatic lineage-specific gene expression patterns can still be detected (such as the hyperactivation of the ToxR regulon in Classical isolates grown *in vitro*). This is reinforced by the fact that the strains used in both experiments were different, though members of the same phylogenetic lineages. This suggests that this approach of comparing gene expression across lineages has merit, and although variation between experiments might influence expression patterns, dramatically-different expression profiles may still be detectable.

However, these data also highlight the risks inherent in working with diverse bacteria under defined *in vitro* conditions. Variation in growth rates, auxotrophies, and the presence of mutations that lower fitness *in vitro* are all factors which could introduce confounding variation into experiments, including the large experiment described in this chapter. Such variations might also explain why replicate cultures of MJD1405 and MJD1409 did not resemble one another in PCA analyses (Figure 6.7). Nevertheless, these data show that observations such as differential expression of virulence genes and T6SS can be made in such an experiment, provided the magnitude of the signal in question is large enough to be detected and to be trustworthy. Additional experiments such as qRT-PCR and the construction of targeted mutants will still be required to validate results from transcriptomic experiments. It is also evident from these data that using this approach, dramatic differences in gene expression can be detected between strains and across growth conditions (Figures 6.5, 6.9). This lends itself to the development of genetic tools, such as fusions between differentially-regulated promoters

(*e.g.*, P_{hcp}, P_{ctxAB}, P_{nanH}) and reporter genes such as *gfp* and *lux*, with which to identify *cis* and *trans*-factors involved in these differences in gene expression. Using the large collection of genome sequences now available to us, the distribution of any such factors, or variants of factors, can be easily determined across the sequenced *V. cholerae* species, such as was carried out in sections 3.4.9, 4.3.11, 5.3.3, and 5.3.8.

The stark differences in T6SS gene expression in multiple *V. cholerae* were particularly intriguing because it has been shown that the 2740-80 Gulf Coast strain of *V. cholerae* O1 constitutively expresses T6SS *in vitro*, whereas in 7PET bacteria, suitable inductive signals must be present for T6SS to be expressed (*e.g.*, [50, 484]). Similarly, the O37 isolate V52 expresses T6SS constitutively *in vitro*, and this strain has been used to study the biology of T6SS in *V. cholerae* [485, 486]. Crucially, although 2740-80 has been shown to express T6SS constitutively, neither of the Gulf Coast strains used in this analysis are identical to 2740-80 [234]. This strongly suggests that the constitutive expression of T6SS is a feature of the Gulf Coast lineage. Moreover, MJD1408 is closely related to the Gulf Coast lineage (Figure 6.6), suggesting that this T6SS expression phenotype might be common to other related lineages within *V. cholerae*. Testing of additional isolates, including those that belong to MX-2 or resemble the Chinese *V. cholerae* O1 that are related to the Gulf Coast lineage (section 5.3.1), may be warranted.

The VasH protein [487], the sigma factor RpoN [485], exposure to chitin, and quorum sensing [50, 484] are all factors which regulate the expression of T6SS in *V. cholerae*. VasH is different between N16961 and V52, and ectopic expression of $VasH_{V52}$ can drive aspects of T6SS expression in N16961, but is not sufficient to activate full T6SS killing activity *in vitro* [487]. The VasH protein sequence has also been shown to be variable across non-O1/O139 *V. cholerae* [488, 489]. The constitutive expression of T6SS is an area of current research and although the mechanisms by which this expression is regulated constitutively in V52 remain unclear, the activity of the WigKR two-component system has recently been implicated in governing T6SS regulation in V52 [490]. In future work, it will be necessary to determine the variation of *vasH* amongst these isolates and across the *V. cholerae* phylogeny, and to determine whether introducing variation in this or other regulators can elicit a constitutive expression phenotype in 7PET or Classical isolates (or, indeed, if abolishing these regulators' activity in Gulf Coast isolates can confer a 7PET-like T6SS expression phenotype *in vitro*).

Future experiments such as these should consider the utilisation of alternative growth conditions, to characterise further any gene expression differences that are lineage-specific. For instance, LB liquid media was used in all of these experiments. It would be worthwhile using minimal media supplemented with relevant carbon sources (*e.g.*, glucose, chitin, etc), both to determine the ability of chosen strains to grow on these media, and to assess whether these nutritional stresses provoke differences in the transcriptome of these isolates and lineages.

It will also be important to repeat these analyses using a panel of strains that includes the reference *V. cholerae* strains that are used by molecular biologists, to ascertain with confidence whether the previous work published on those reference strains describes the transcriptional behaviour of those strains, or of lineages more generally. Similarly, it is important to recognise that the gene expression data calculated in these experiments were generated by mapping to the N16961 reference genome. Differences in the expression of genes present in other isolates that are absent from N16961 will not be detectable using this approach. Producing closed genome assemblies for each strain being assayed will be an important next step in the optimisation of these large-scale transcriptomic assays.