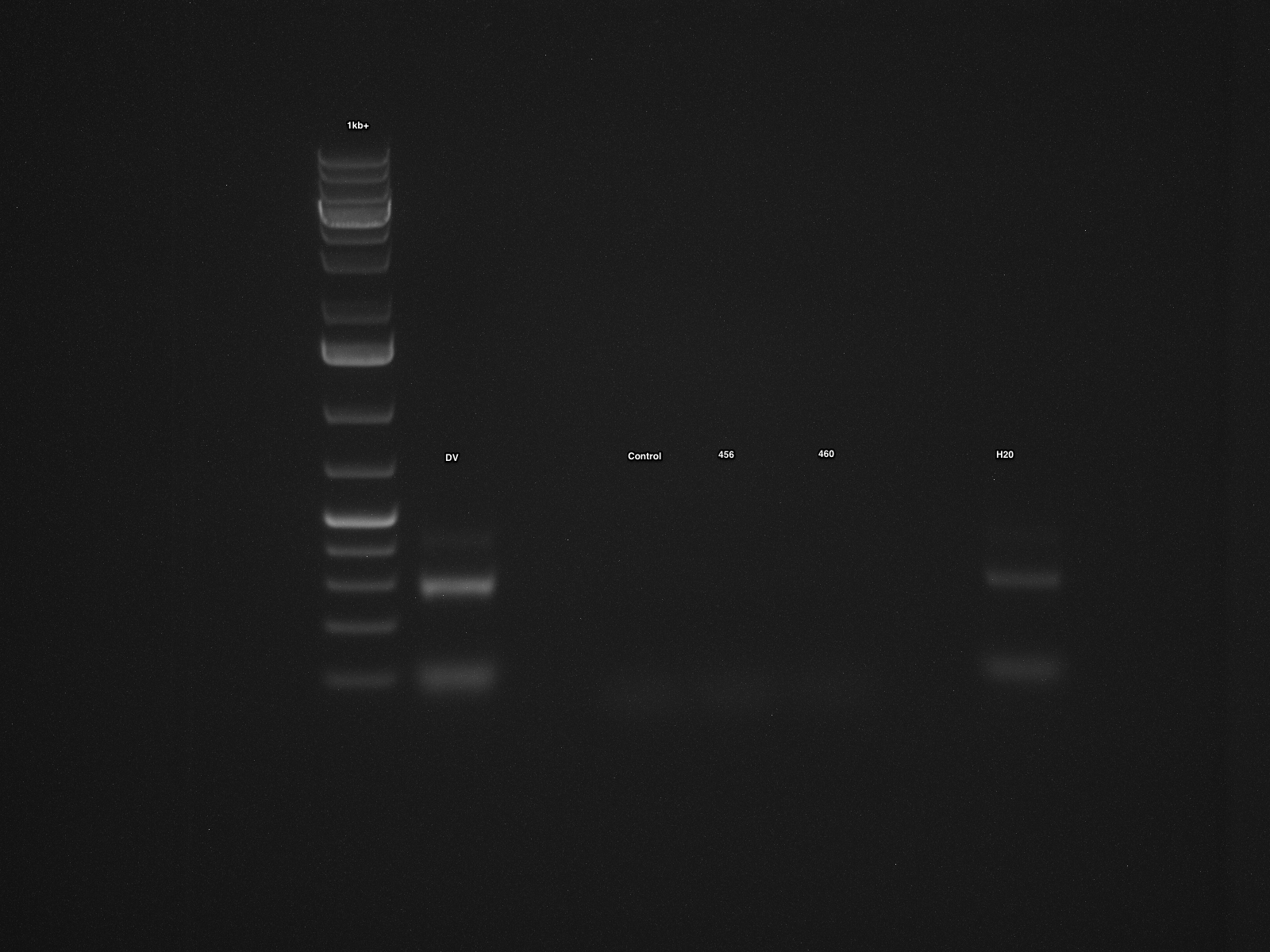
Supplementary information 1: Physical parameters measured while sampling the well. The following concentrations were below the detection limit- arsenic, cadmium, cobalt, cesium, lead, aluminum, uranium, silver, chromium, nickel, copper, selenium, phosphate, nitrate, bromide, and mercury. The centering material for the PVC ground water pipe in the wells contain mercury only in rare instances and in this case, mercury was not detected in the ground water samples from both wells. The value highlighted in grey was below the detection limit.

AODC: Acridine Orange Direct Count, DIC: Dissolved Inorganic Carbon, DOC: Dissolved Organic Carbon.

|  |  |  |
| --- | --- | --- |
| Sample ID | GW456 (11-11-14) | GW460 (12-1-14) |
| Weather Conditions | Partly sunny | Cloudy |
| Temperature (°C) | 20.03 | 14.71 |
| Daily Rainfall (mm) | 0 | 5.60 |
| AODC cells/mL | **49858.22** | **17013.23** |
| Dissolved Oxygen (mg/L) | 0.51 | 0.11 |
| Conductivity (μS/cm) | 232.90 | 258.50 |
| Redox (mV) | -43.00 | -40 |
| pH | 7.98 | 7.46 |
| Sulfide (ppm) | 0.013 | 0.006 |
| Ferrous Ion (mg/L) | 0.04 | 0.02 |
| DIC mg/L | 29.93 | 38.63 |
| DOC mg/L | 0.28 | 0.72 |
| Chloride mg/l | 0.80 | 0.87 |
| Sulfate mg/l | 6.11 | 7.41 |
| Barium mg/l | 0.58 | 0.58 |
| Calcium mg/l | 26.68 | 31.58 |
| Gallium mg/l | 0.02 | 0.02 |
| Potassium mg/l | 3.53 | 2.99 |
| Magnesium mg/l | 10.99 | 11.33 |
| Manganese mg/l | 0.02 | 0.06 |
| Sodium mg/l | 34.77 | 14.14 |
| Strontium mg/l | 1.10 | 1.08 |
| Mercury | 0 | 0 |
| Zinc mg/l | 0.02 | 0.01 |

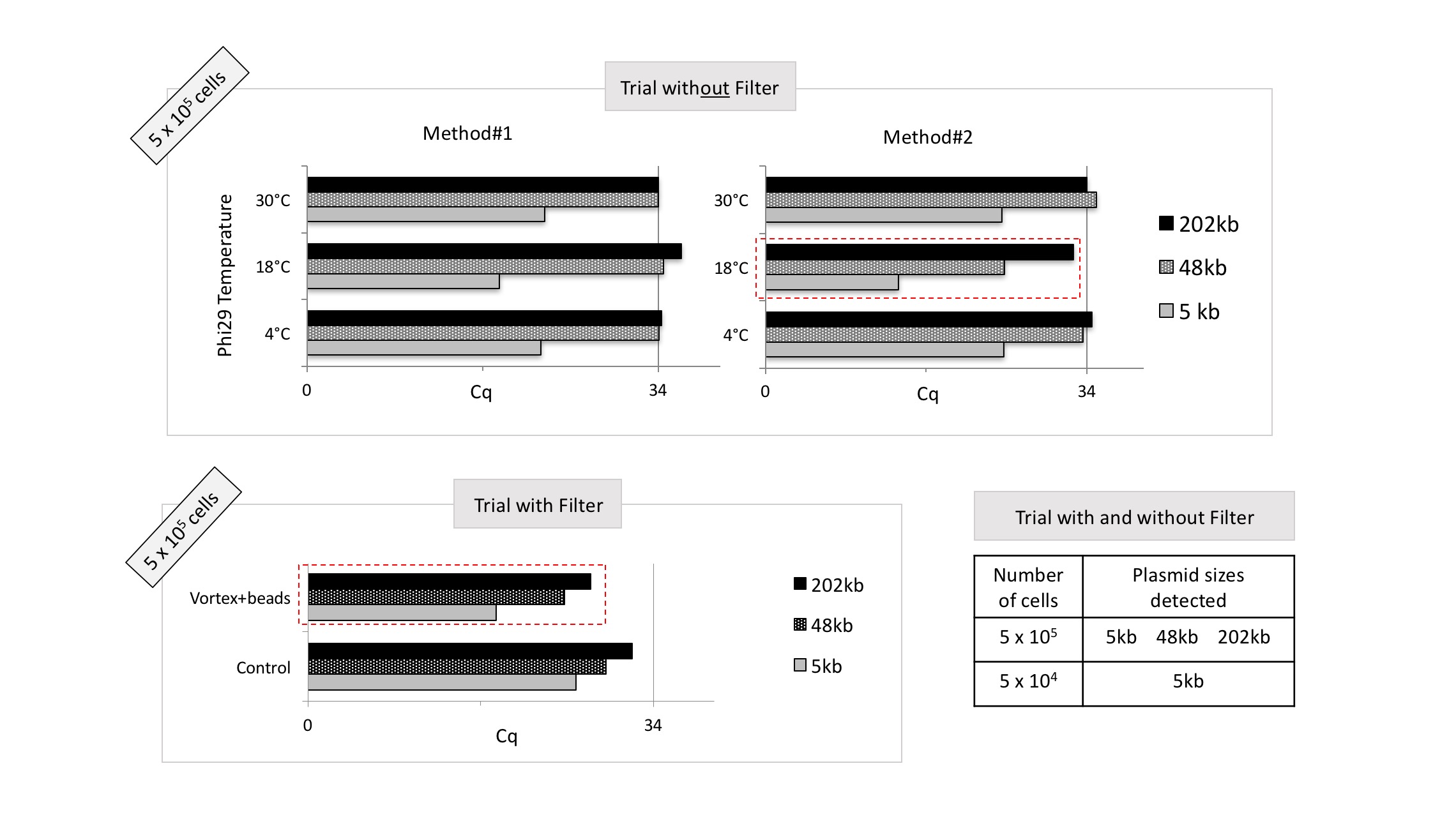
Supplementary Information 2: The lack of genomic DNA contamination in the plasmid DNA samples from wells GW456 and GW460 depicted by the lack of PCR products on agarose gel electrophoresis after PCR with degenerate 16S rRNA primers. The *Desulfovibrio* plasmid DNA was used as a positive control. The image depicts 1kb+ ladder (lane 1), *Desulfovibrio* genomic DNA (lane 2), blank (lane 3), *Desulfovibrio* plasmid DNA (lane 4), GW456 plasmid DNA (lane 5), GW460 plasmid DNA (lane 6), and water (lane 7). The expected product size is 300 bp.



1 2 3 4 5 6 7

300 bp

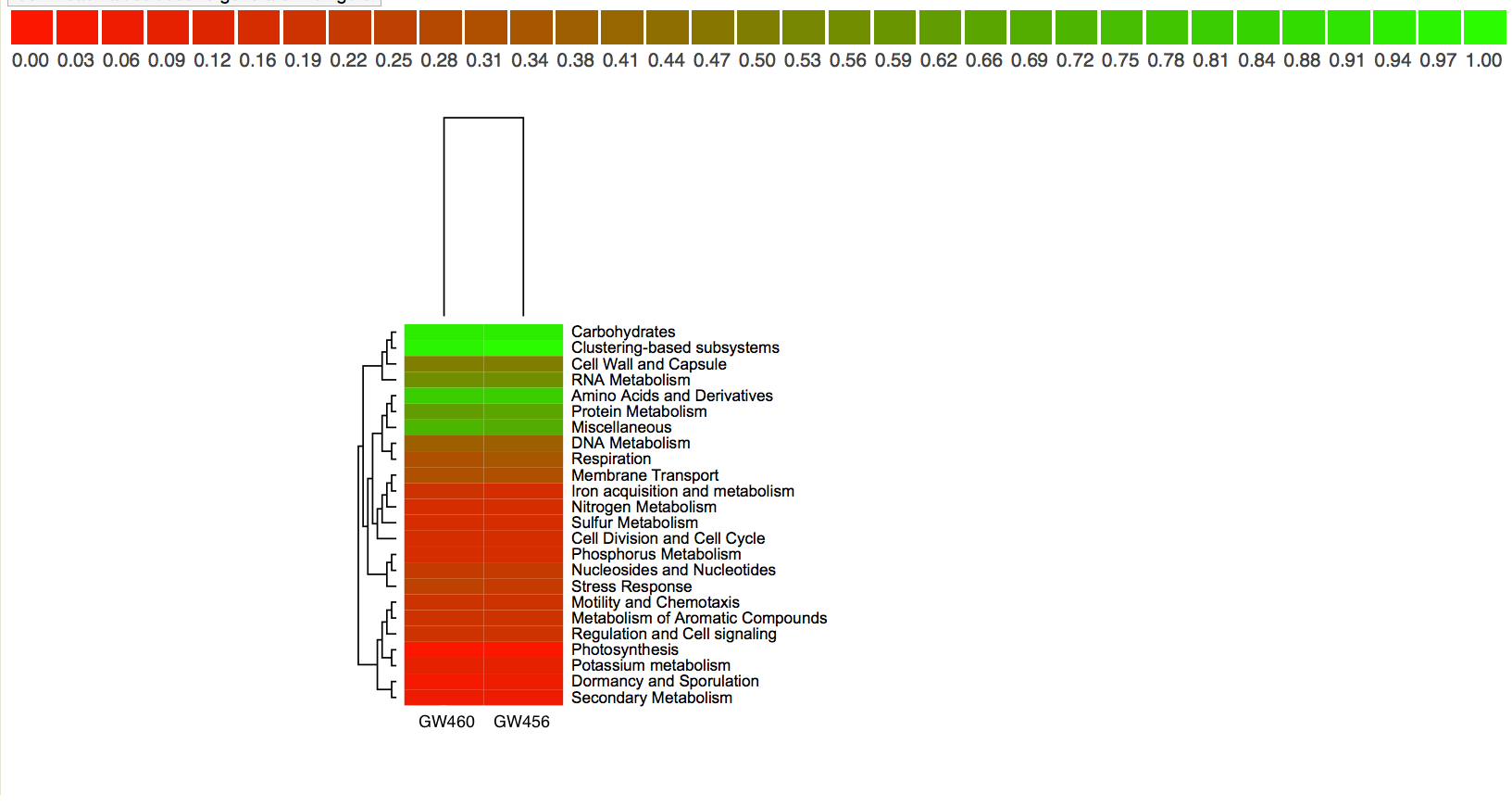
Supplementary Information 3: Optimization of plasmid isolation methods performed by plasmid isolation followed by qPCR based detection of the plasmid by targeting a unique plasmid encoded gene on each of the plasmids a) depicts the threshold values (Cq) obtained when isolating plasmids from 5 x 105 cells using the two methods (Birnboim and Doly, 1979; Anderson and Mckay, 1983), at three different Phi29 incubation temperatures b) depicts optimization of plasmid isolation procedure when cells are present on a filter. A lower Cq number indicates more copies of the plasmid were present in the DNA sample, and hence implies better plasmid detection. The Cq values refer to an average of two technical replicates.



a.

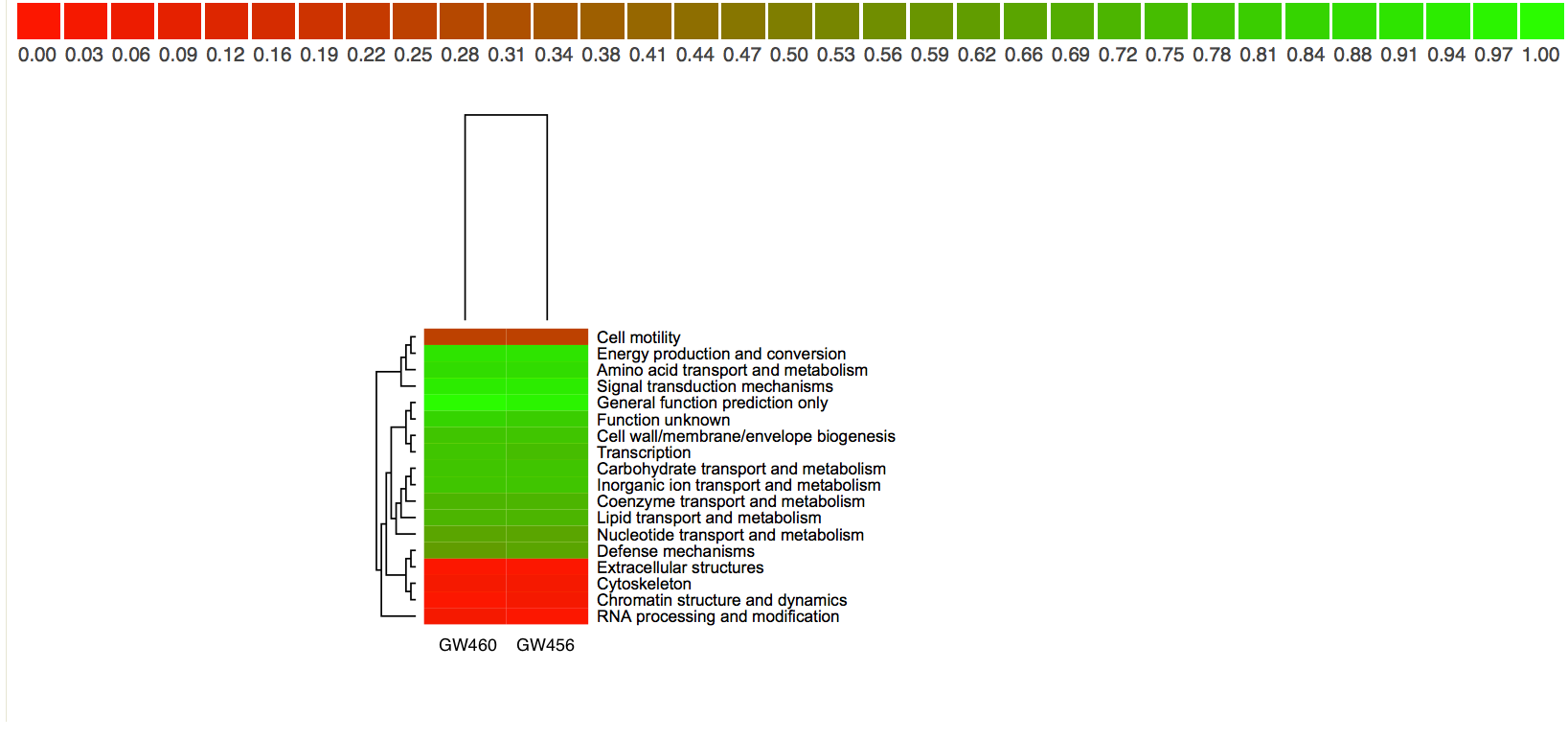
b.

Supplementary Information 4: Heat map of the normalized functional abundance of “all\_scaffolds” in MG-RAST based on the hierarchical classification a) Subsystems annotation level-1, and b) COG annotation level-2 (Parameters: 1e-5 maximum e-value cut off, 60 % maximum identity cut off, 15 bp minimum alignment length cut off)

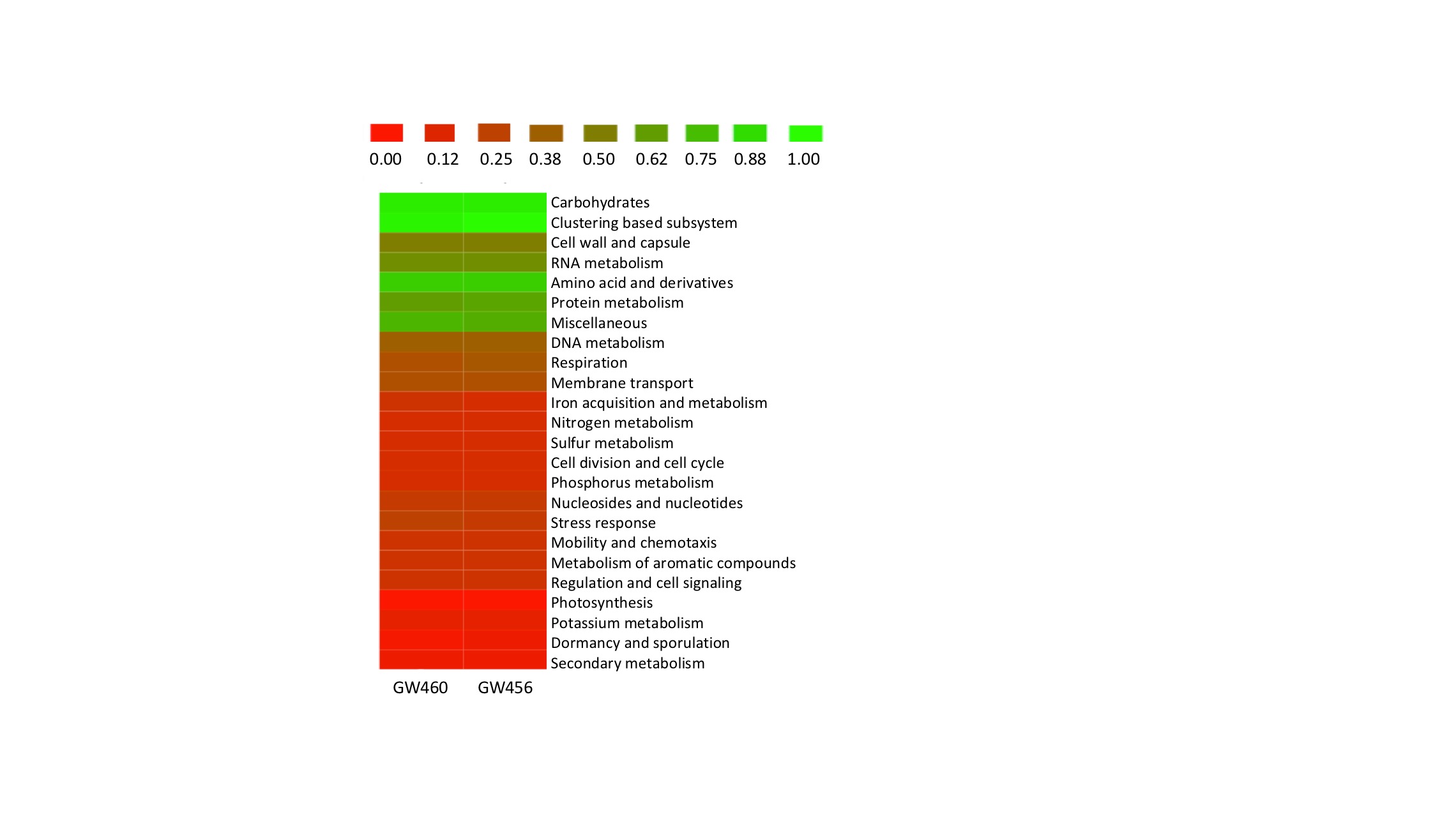


SEED Subsystem

Classification



COG Classification



Supplementary Information 5: Distribution of “all\_scaffolds” and “circular\_scaffolds” from both wells based on the scaffold size.

/Users/akothari/Desktop/ENIGMA Paper/Figures/Length_histogram.pdf

c.

d.

b.

a.

Scaffold size (kb)

Scaffold size (kb)

Scaffold size (kb)

Scaffold size (kb)

GW456 “circular\_scaffolds”

GW460 “circular\_scaffolds”

GW456 “all\_scaffolds”

GW460 “all\_scaffolds”

Number of scaffolds

Number of scaffolds

Number of scaffolds

Number of scaffolds

Supplementary Information 6: Distribution of scaffolds from both wells based on the GC content.

a.

/Users/akothari/Desktop/ENIGMA Paper/Figures/GC_histogram (1).pdf

c.

d.

b.

GC content (%)

GC content (%)

GC content (%)

GC content (%)

GW460 “circular\_scaffolds”

GW456 “circular\_scaffolds”

GW456 “all\_scaffolds”

GW460 “all\_scaffolds”

Number of scaffolds

Number of scaffolds

Number of scaffolds

Number of scaffolds

Supplementary Information 7: Top hits obtained by the comparison of “all\_scaffolds” and “circular\_scaffolds” from the wells GW456 and GW460 with A CLAssification of Mobile genetic Elements (ACLAME) database. The number of hits are indicated in the bracket.

|  |  |  |
| --- | --- | --- |
| **Category** | **Well GW456** | **Well GW460** |
| Gene annotation  – “all\_scaffolds” | Total Hits: 417828  hypothetical protein (79052)  ABC transporter related (3061) | Total Hits: 362921  hypothetical protein (70055)  ABC transporter related (2298) |
| Gene source  – “all\_scaffolds” | Total Hits: 417828  *Ralstonia solanacearum* GMI1000(31019)  Sinorhizobium meliloti 1021 (27117) | Total Hits: 362921  *Ralstonia solanacearum* GMI1000 (29419)  *Sinorhizobium meliloti* (23173) |
| Gene annotation  – “circular\_scaffolds” | Total Hits: 406  hypothetical protein (121)  outer membrane efflux protein (7) | Total Hits: 1486  hypothetical protein (449)  Mobilization protein (24) |
| Gene source  – “circular\_scaffolds” | Total Hits: 406  *Pelobacter propionicus* DSM 2379 (26)  *Verminephrobacter eiseniae EF01-2* (22) | Total Hits: 1486  *Sphingomonas wittichii* RW1 (66)  *Ralstonia solanacearum GMI1000 (47)* |

Supplementary Information 8: Antibiotic resistance encoded by “all\_scaffolds” from the wells GW456 and GW460, by comparison with the Antibiotic Resistance Genes Database (ARDB) and the Comprehensive Antibiotic Resistance Database (CARD).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | GW456 | | | GW460 | |
| **ARDB-** Number of genes with hits | |  | 236 |  | 312 |
| Top categories | Bacitracin resistance  *bacA* | | 70.8 % | Bacitracin resistance  *bacA* | 42.9 % |
|  | Chlaramphenicol resistance *catA1* | | 2.5 % | Chlaramphenicol resistance *catA1* | 22.8 % |
|  | Nodulation cell division resistance *mexF* | | 2.5 % | Nodulation cell division resistance *mexF* | 4.8 % |
| **CARD-** Number of genes with hits | |  | 1002 |  | 862 |
| Top categories | Aminocoumarin resistance *alaS* | | 14.9 % | Aminocoumarin resistance *alaS* | 11.5 % |
|  | Elfamycin resistance | | 10.2 % | Elfamycin resistance | 7.8 % |
|  | mupirocin resistance *ileS* | | 7.7 % | DNA repair *mfd* | 7.1 % |

Supplementary Information 9: The percentage of scaffolds encoding homologs of toxin, antitoxin and regulator gene sequences.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Scaffolds | Toxin (%) | Antitoxin (%) | Regulator (%) | Toxin and Antitoxin (%) | Toxin, Antitoxin, and Regulator (%) |
| GW456 “all\_scaffolds” | 2.1 | 2.0 | 0.3 | 0.8 | 0.1 |
| GW460 “all\_scaffolds” | 1.6 | 1.6 | 0.3 | 0.6 | 0.1 |
| GW456 “circular\_scaffolds” | 23.5 | 29.4 | 7.4 | 19.1 | 4.4 |
| GW460 “circular\_scaffolds” | 10.4 | 13.1 | 3.3 | 7.7 | 1.8 |

Supplementary Information 10: Top hits obtained by the comparison of “all\_scaffolds” and “circular\_scaffolds” from the wells GW456 and GW460 with the Antibacterial Biocide and Metal Resistance Genes (BacMet) database. The number of hits are indicated in the bracket.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Category** | **Well GW456** | | | **Well GW460** |
| *BacMet- Predicted genes database* | |  | |  |
| Gene annotation  – “all\_scaffolds” | Total Hits: 487  heavy metal translocating P-type ATPase (12)  Arsenic resistance protein (12) | | | Total Hits: 646  Copper-translocating P-type ATPase (16)  LysR family transcriptional regulator (13) |
| Gene source  – “all\_scaffolds” | Total Hits: 487  *Pseudomonas fluorescens* SBW25 (39)  *Pseudomonas fluorescens* WH6 (28) | | | Total Hits: 646  *Pseudomonas fluorescens* SBW25 (66)  *Pseudomonas fluorescens* Pf0-1 (62) |
| Resistance genes to compounds  – “all\_scaffolds” | Total Hits: 487  Copper (66)  Triclosan (36)  Arsenic (33) | | | Total Hits: 646  Copper (99)  Triclosan (70)  Arsenic (34) |
| No results when the “circular\_scaffolds” were compared. | | | | |
| *BacMet - Experimentally confirmed genes database* | | |  |  |
| Gene annotation  – “all\_scaffolds” | Total Hits: 162  Mercury reductase enzyme, merA (9)  Multidrug efflux RND transporter, MexF (4) | | | Total Hits: 192  Multidrug efflux RND transporter, permease protein MexF; part of mexE-mexF-oprN efflux operon (14)  Mercury reductase enzyme, merA (14) |
| Gene source  – “all\_scaffolds” | Total Hits: 162  *Escherichia coli (strain K12) (80)*  *Pseudomonas stutzeri* (10) | | | Total Hits: 192  *Escherichia coli (strain K12) (84)*  *Pseudomonas fluorescens Pf-5/ATCC BAA-477* (20) |
| Resistance genes to compounds  – “all\_scaffolds” | Total Hits: 162  Copper (23)  Mercury (17)  Triclosan (16) | | | Total Hits: 192  Copper (32)  Triclosan (28)  Mercury (20) |
| No results when the “circular\_scaffolds” were compared. | | | | |

Supplementary Information 10: Antibiotic resistance encoded by “all\_scaffolds” from the wells GW456 and GW460, by comparison with the Antibiotic Resistance Genes Database (ARDB) and the Comprehensive Antibiotic Resistance Database (CARD).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | GW456 | | | GW460 | |
| **ARDB-** Number of genes with hits | |  | 236 |  | 312 |
| Top categories | Bacitracin resistance  *bacA* | | 70.8 % | Bacitracin resistance  *bacA* | 42.9 % |
|  | Chlaramphenicol resistance *catA1* | | 2.5 % | Chlaramphenicol resistance *catA1* | 22.8 % |
|  | Nodulation cell division resistance *mexF* | | 2.5 % | Nodulation cell division resistance *mexF* | 4.8 % |
| **CARD-** Number of genes with hits | |  | 1002 |  | 862 |
| Top categories | Aminocoumarin resistance *alaS* | | 14.9 % | Aminocoumarin resistance *alaS* | 11.5 % |
|  | Elfamycin resistance | | 10.2 % | Elfamycin resistance | 7.8 % |
|  | mupirocin resistance *ileS* | | 7.7 % | DNA repair *mfd* | 7.1 % |

**References**

Anderson DG, Mckay LL. (1983). Simple and rapid method for isolating large plasmid DNA from lactic streptococci. *Appl Environ Microbiol* **46**: 549–552.

Birnboim HC, Doly J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* **7**: 1513–1523.