listed in the Database of prOkaryotic OpeRons (DOOR) [59].

RESULTS

Production and sequencing of the HDTM libraries

With the goal of identifying the essential genome of *B. cenocepacia* K56-2, we used HDTM followed by Illumina sequencing of the transposon insertion sites. We generated a HDTM library of one million mutants by introducing the suicide plasmid pRBrhaBoutgfp [60] into *B. cenocepacia* K56-2 by triparental mating. To avoid insertions in non-essential genes from being lethal due to polar effects on downstream essential genes, the transposon contains a rhamnose-inducible promoter. Selection of transposon mutants in the presence of rhamnose allows expression of genes in a transcriptional unit downstream of the transposon insertion. To enrich for the transposon–genome junctions and identify the location of the insertion sites, we used the Tn-seq circle method [21] (Fig. S1).

As species of the genus *Burkholderia* have large multireplicon genomes with approximately 67 % G+C content [29], we first considered using a Hi-fidelity KAPA polymerase (KAPA bioscience) to amplify the transposon insertions by PCR. The KAPA DNA polymerase has been successfully used to increase the proportion of reads from transposon junctions in GC-rich regions in *B. thailandensis* [61] and has minimal amplification bias [62]. However, PCR-amplification of the transposon junctions may not be favoured by the KAPA DNA polymerase as the G+C content of the transposon sequence inserted into the genome is much lower (Fig. S2). To test the ability of the KAPA DNA polymerase to PCR-amplify the transposon junctions, the sequences of the HDTM library produced after PCR amplification with the KAPA DNA polymerase were compared with those obtained with the iTaq DNA polymerases (Bio-Rad). Sequencing the HDTM library after PCR-amplification with the KAPA polymerase resulted in 89 983 unique insertion sites with an average of 1 insert every 87 bases and a read G+C content of 61.0 % (Table 1). However, PCR-amplification with the iTaq DNA polymerase revealed 293 568 unique insertion sites, with an average of 1 insert every 27 bases, a read G+C content of 59.7 % (Table 1) and a lower proportion of insertion sites in GC-rich regions (Fig. S3). The total reads from PCR-amplification of the HDTM library with the iTaq DNA polymerase were more evenly distributed over the insertion sites, whereas the use of the KAPA DNA polymerase resulted in many insertions with a low read count and a large proportion of reads mapping to a small number of insertion sites (Fig. S4). For these reasons, the identification of essential genes was performed with the data produced using the iTaq DNA polymerase.

Essential gene identification

The first step in Tn-seq data analysis is to map the reads against the genome of the micro-organism under investigation (Fig. S5). As a complete genome of *B. cenocepacia* K56-2 is not available, we used the contigs from the draft genome of *B. cenocepacia* K56-2 [41] to map the Tn-seq reads and removed the 100 highest-read sites. After removing these reads, it was evident that there were positional effects on the insertion density and read counts (Fig. S6). These effects were not due to transposon insertion or Tn-seq method biases as the distribution of reads from sequencing the whole genome of *B. cenocepacia* K56-2 on the Illumina MiSeq platform exhibited the same trend. Notably, 17 of the 20 insertion sites with high read counts were in regions predicted to be genomic islands by Islandviewer3 [63] (Fig. 1). Overall, the reads were not evenly scaled across the contigs and the insertion density was strongly correlated with the G+C content (Fig. 1). To account for these biases, we created a model of read depth as a function of the position along each contig and G+C content. We then corrected the read count based on the model prediction to normalize the reads prior to the essentiality analysis, which minimized the effects of position and G+C content on read density (Fig. S7).

After normalizing the reads by contig and G+C content, we generated pseudocounts in 2000 simulations of randomly rearranging the read counts for each insertion site around the genome (Fig. S3). We then compared the mean read count from the actual data to the variance of the pseudocounts using the R package, EdgeR [53]. Essential genes were identified based on whether the mean read counts for a gene were statistically different than expected if the gene were not essential. Using this analysis, we identified 508 essential genes in *B. cenocepacia* K56-2 (Table S3). Included in our predicted essential gene set are genes in which disruptive mutations in *B. cenocepacia* causes a lethal phenotype: *dxs* (BCAM0911), *hemE* (BCAL0040), *infB* (BCAL1507), *gyrB* (BCAL0421), *ubiB* (BCAL0876), *valS* (BCAL1448), BCAL3369, and *murJ* (BCAL2764) [64–67]. In addition, we confirmed the essentiality of genes previously characterized by our laboratory. These genes encode EtfAB, an essential electron transfer flavoprotein [68] and EsaR, an essential response regulator involved in cell envelope integrity [69].

Table 1. Summary of results from sequencing the HDTM library

HDTM library preparation	Total reads	No. of reads containing the transposon sequence and mapping to the genome	G+C content of reads	No. of unique insertion sites	Frequency of Tn insertion
KAPA	20 459 975	4 370 457 (21 %)	61.0 %	89 983	1/87 bp
iTaq	15 132 067	6 936 891 (46 %)	59.7 %	293 568	1/27 bp