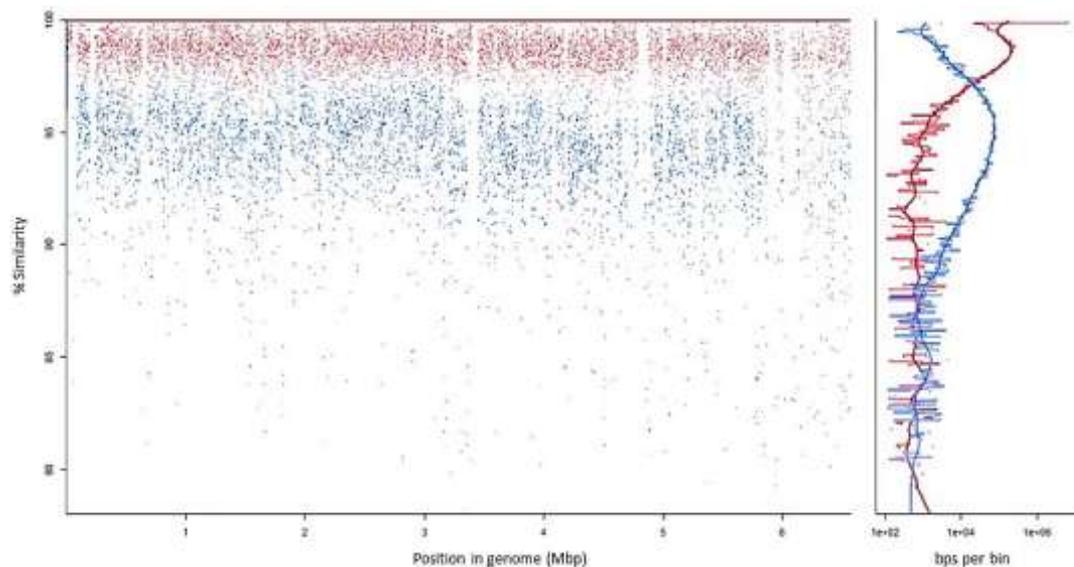


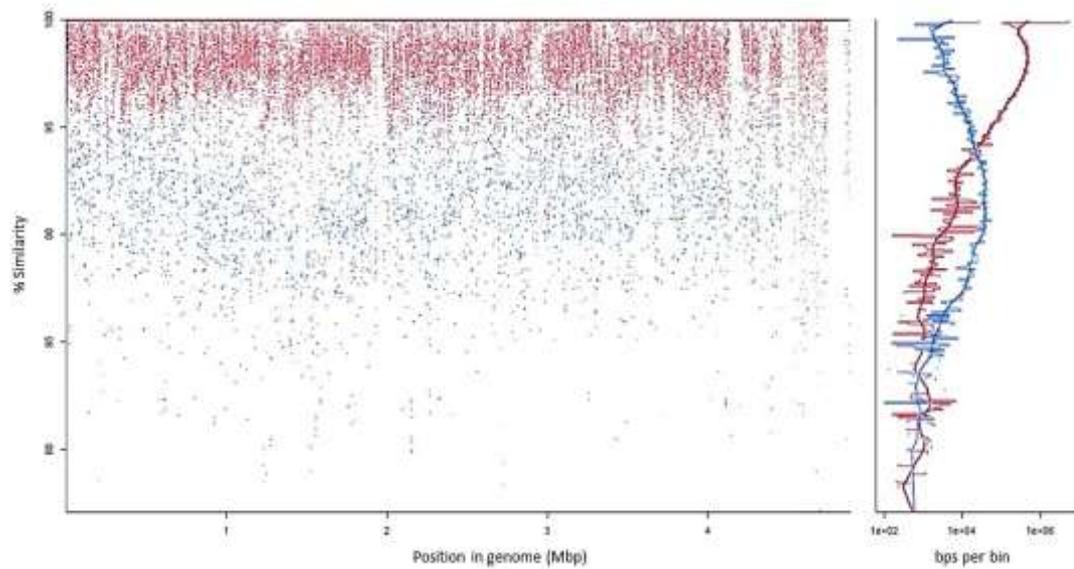
Supplementary File 4

Recruitment plots

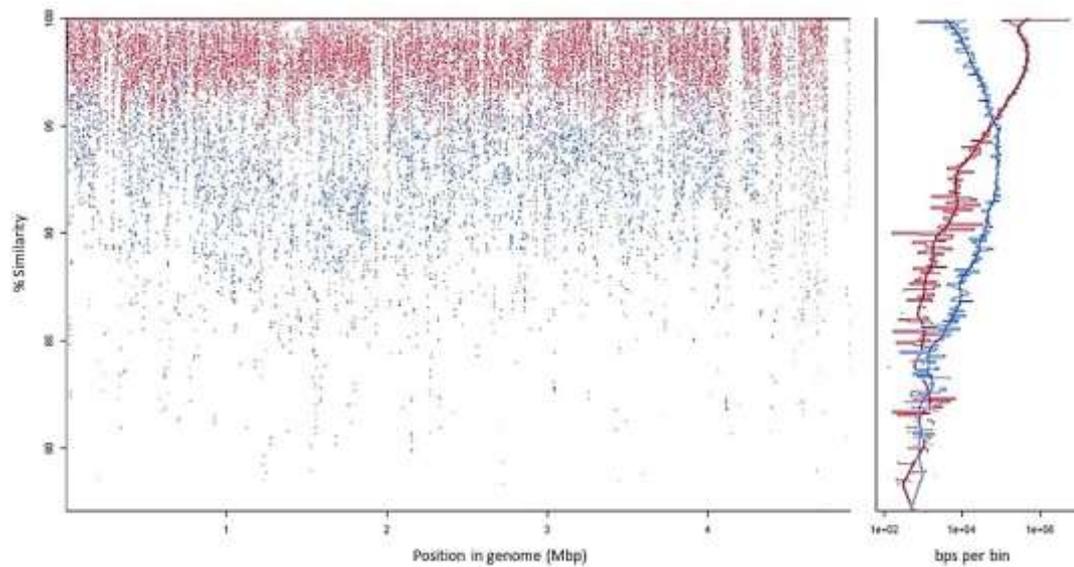
The similarity of genes present on different whole genome sequences was assessed using the recruitment plot approach initially described by Ghai et al. (2010) for metagenome data. Gene similarities were determined using the BLASTN, where the nucleotide sequences of the gene were compared to a specified reference genome. Only genes sharing at least 70% coverage with a gene on the reference genome were included. The similarity value determined for the identified gene and the corresponding gene on the reference genome was then plotted using the `replot2` function of the `enveomics.R` package in R. Genes were ordered on the x-axis according to their position on the reference genome. The recruitment plots for *Paraburkholderia kirstenboschensis* and *Par. caledonica* as well as the between members of *E. coli sensu stricto* and *Escherichia* Clades III, IV and V are shown below.



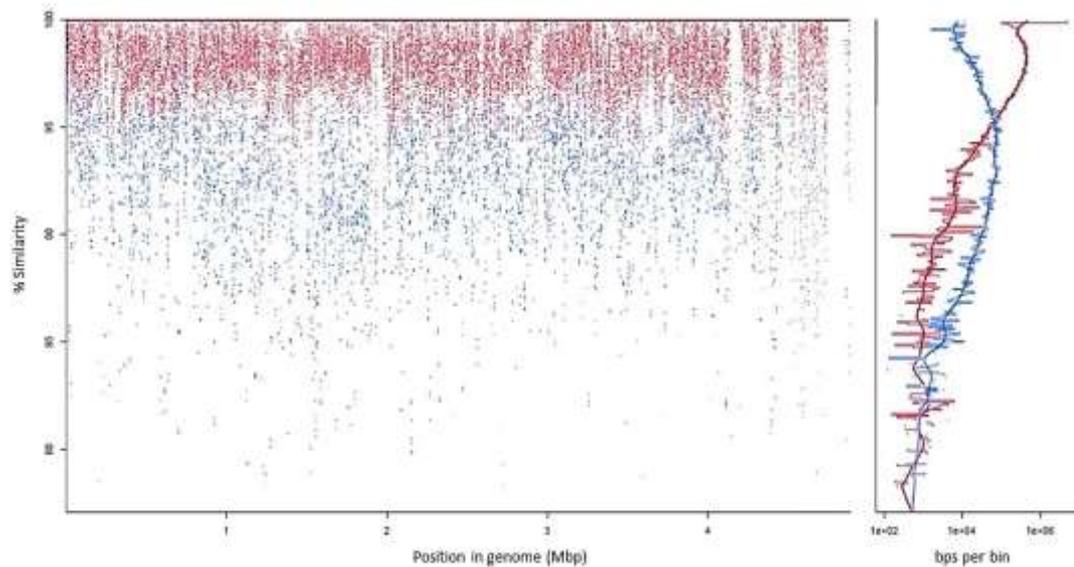
Recruitment plot comparing the genome sequences of *Paraburkholderia kirstenboschensis* (RAU2d2) and *Paraburkholderia caledonica* (NBRC 102488) with the *Paraburkholderia kirstenboschensis* (Kb15^T) genome as reference. Genes belonging to the *Par. kirstenboschensis* strains are indicated in red and *Par. caledonica* genes are indicated in blue. Genes are ordered on the x-axis according to their position on the reference genome. The corresponding position of the gene on the y-axis indicates the gene's similarity with the corresponding gene on the reference genome. The side graph indicates the distribution of genes according to their similarity values.



Recruitment plot comparing the genome sequences of an *Escherichia* Clade V strains (E1118) and five *Escherichia coli sensu stricto* strains (E2348/69, EDL933, MG1655, CFT073 and 2457T (*Shigella flexneri*)) with the *E. coli* (042) genome as reference. Genes belonging to the *E. coli sensu stricto* strains are indicated in red and those belonging to the *Escherichia* Clade V strain are indicated in blue. Genes are ordered on the x-axis according to their position on the reference genome. The corresponding position of the gene on the y-axis indicates the gene's similarity with the corresponding gene on the reference genome. The side graph indicates the distribution of genes according to their similarity values.



Recruitment plot comparing the genome sequences of an *Escherichia* Clade III strains (TW09231 and TW09276) and five *Escherichia coli sensu stricto* strains (E2348/69, EDL933, MG1655, CFT073 and 2457T (*Shigella flexneri*)) with the *E. coli* (042) genome as reference. Genes belonging to the *E. coli sensu stricto* strains are indicated in red and those belonging to the *Escherichia* Clade III strains are indicated in blue. Genes are ordered on the x-axis according to their position on the reference genome. The corresponding position of the gene on the y-axis indicates the gene's similarity with the corresponding gene on the reference genome. The side graph indicates the distribution of genes according to their similarity values.



Recruitment plot comparing the genome sequences of an *Escherichia* Clade IV strains (TW14182 and H6056) and five *Escherichia coli sensu stricto* strains (E2348/69, EDL933, MG1655, CFT073 and 2457T(*Shigella flexneri*)) with the *E. coli* (042) genome as reference. Genes belonging to the *E. coli sensu stricto* strains are indicated in red and those belonging to the *Escherichia* Clade IV strains are indicated in blue. Genes are ordered on the x-axis according to their position on the reference genome. The corresponding position of the gene on the y-axis indicates the gene's similarity with the corresponding gene on the reference genome. The side graph indicates the overall distribution of genes according to their similarity values.

Ghai R, Martin-Cuadrado A-B, Molto AG, Heredia IG, Cabrera R, Martin J, Verdu M, Deschamps P, Moreira D, Lopez-Garcia P, Mira A, Rodriguez-Valera F (2010) Metagenome of the Mediterranean deep chlorophyll maximum studied by direct and fosmid library 454 pyrosequencing. *ISME J* 4:1154-1166.