

# Supplementary Materials for

## **Epistasis, core-genome disharmony, and adaptation in recombining bacteria**

Aidan J. Taylor<sup>1</sup>, Koji Yahara<sup>2</sup>, Ben Pascoe<sup>3</sup>, Seungwon Ko<sup>3</sup>, Leonardos Mageiros<sup>4,13</sup>, Evangelos Mourkas<sup>3</sup>, Jessica K Calland<sup>5</sup>, Santeri Puranen<sup>6</sup>, Matthew D. Hitchings<sup>4</sup>, Keith A. Jolley<sup>3</sup>, Carolin M. Kobras<sup>7</sup>, Sion Bayliss<sup>8</sup>, Nicola J. Williams<sup>9</sup>, Arnoud H. M. van Vliet<sup>10</sup>, Julian Parkhill<sup>11</sup>, Martin C. J. Maiden<sup>3</sup>, Jukka Corander<sup>6,7,12</sup>, Laurence D Hurst<sup>13</sup>, Daniel Falush<sup>14</sup>, Paul Keim<sup>3,15</sup>, Xavier Didelot<sup>16</sup>, David J. Kelly<sup>17\*</sup>, Samuel K. Sheppard<sup>3\*</sup>

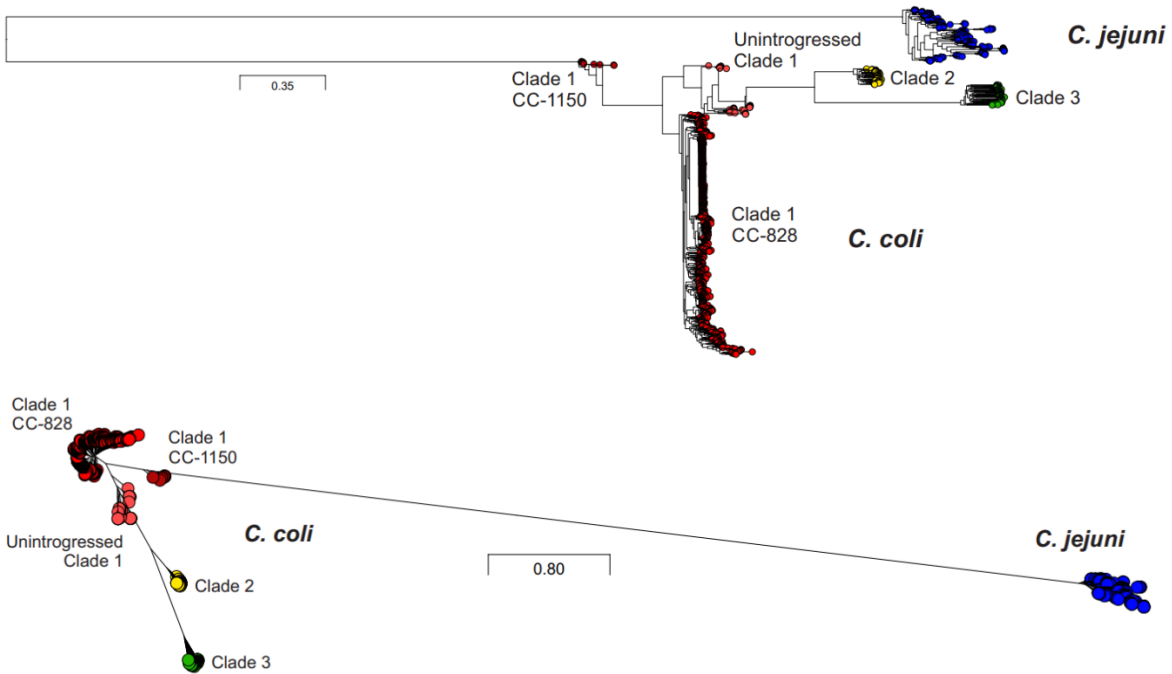
\*Corresponding author(s): David J. Kelly, [d.kelly@sheffield.ac.uk](mailto:d.kelly@sheffield.ac.uk); Samuel K. Sheppard, [s.k.sheppard@bath.ac.uk](mailto:s.k.sheppard@bath.ac.uk);

### **This file includes:**

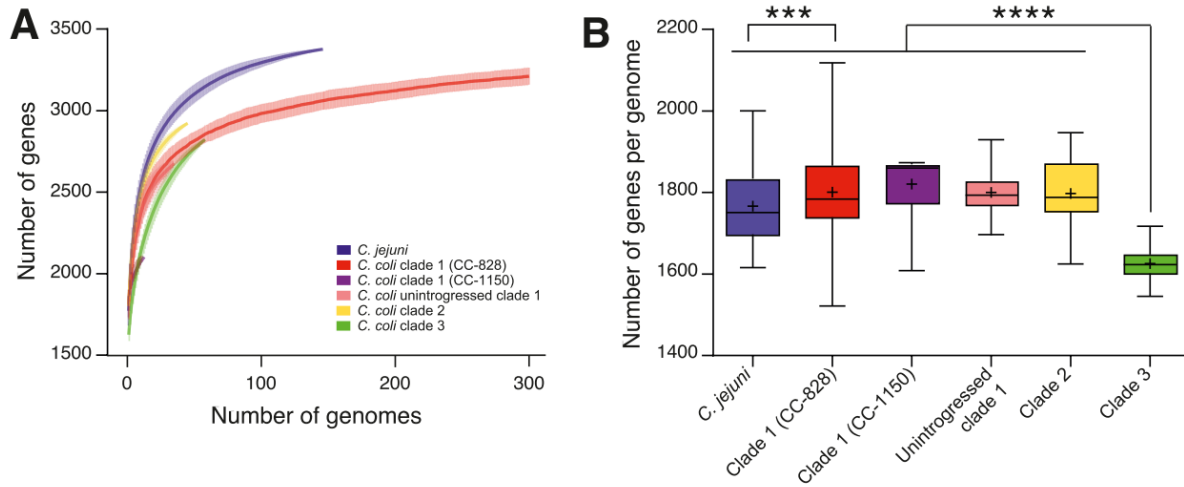
Figs. S1 to S8

### **Other Supplementary Materials for this manuscript include the following:**

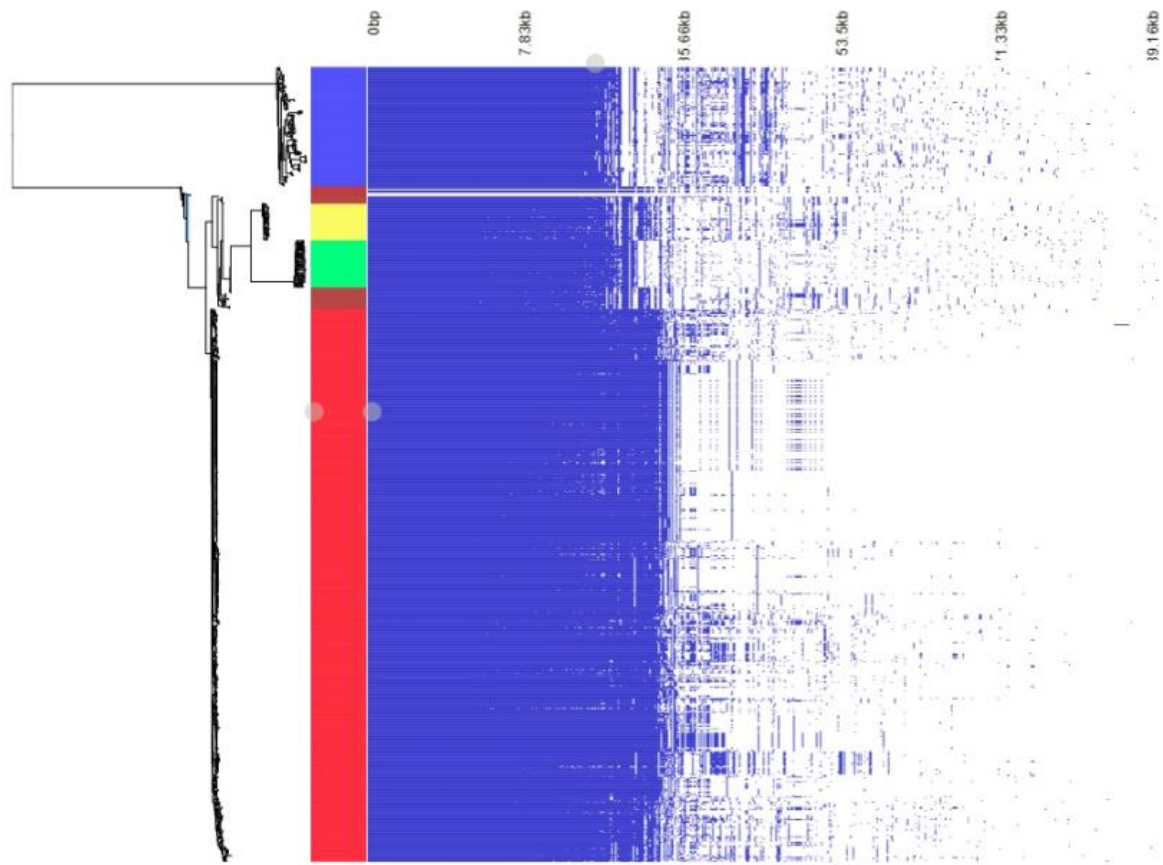
Data S1 to S5



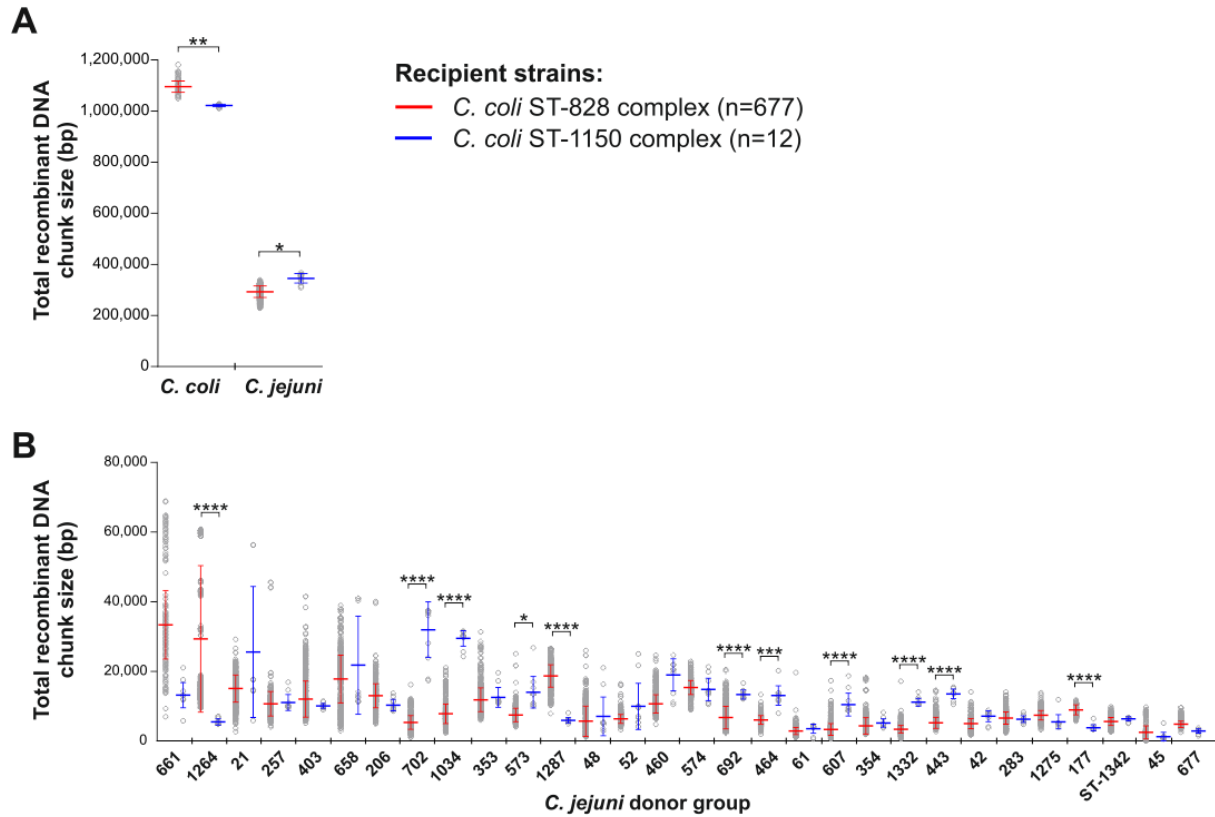
**Fig S1. RAxML phylogeny of 973 *C. jejuni* and *C. coli* clades.** The phylogenetic tree was constructed using an approximation of the maximum-likelihood (ML) method implemented in RAxML using the GTRGAMMA model, from an imputed whole-genome alignment of 973 genomes to the *C. coli* CVM N29710 reference genome (ST-828 clonal complex) and a total of 286,393 variable sites. Three clades of *C. coli* are included: 689 isolates from introgressed clade 1 (677 from ST-828 complex and 12 from ST-1150 complex) and 35 isolates from unintrogressed clade 1 (red), 45 isolates from clade 2 (yellow) and 58 isolates from clade 3 (green). A selection of 146 *C. jejuni* from a representation of 30 different clonal complexes is included (blue). The scale represents the number of substitutions per site.



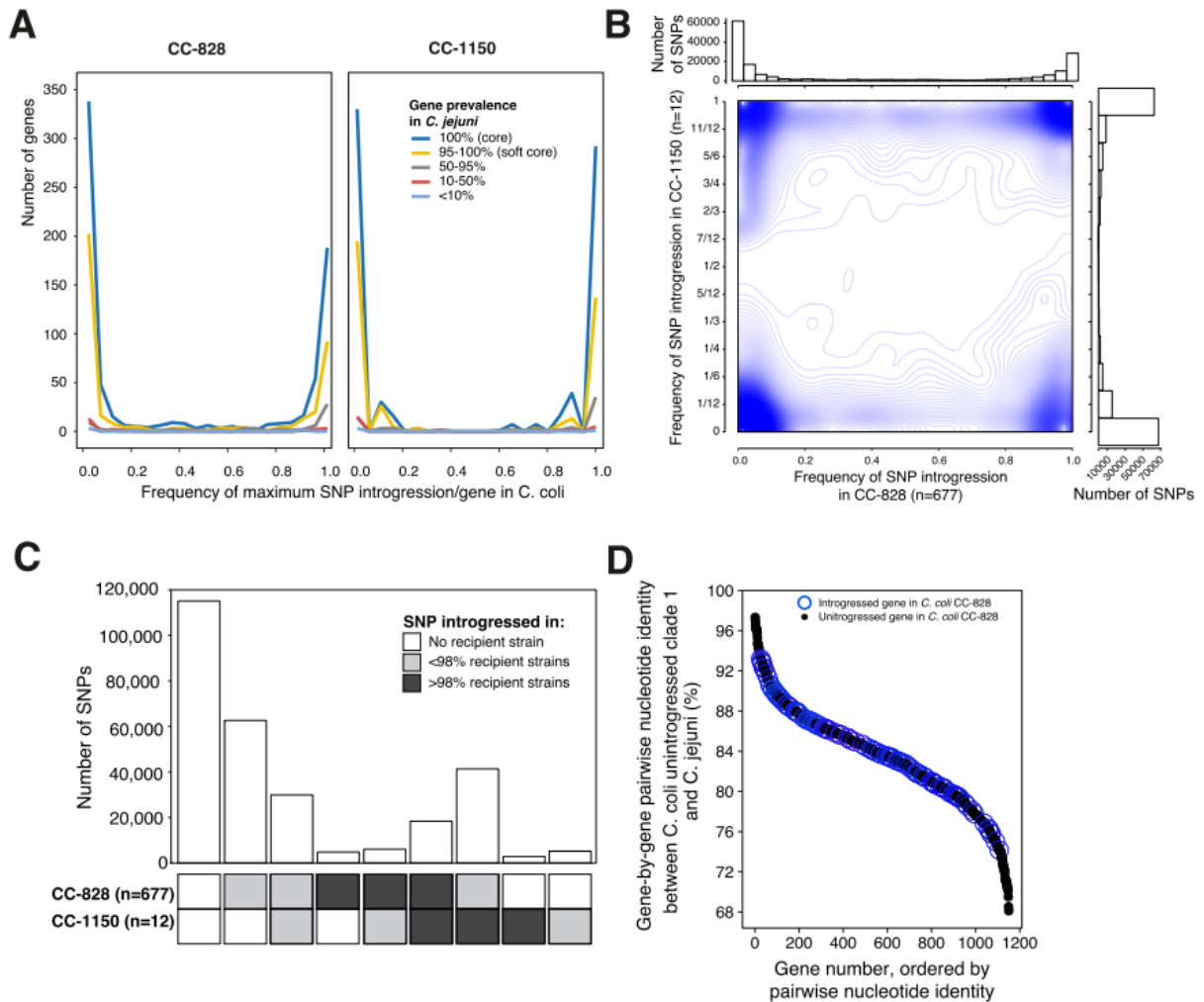
**Fig S2. Population genomics of *C. jejuni* and *C. coli*.** (A) Accumulation curves of the clade-specific pan-genome content of *Campylobacter* lineages, using the same colour code as panel A. Randomized genome sampling was used to obtain the average number of genes for each comparison (plain lines) and standard deviations (dotted lines). (B) The average number of genes/genome, identified by BLAST, is shown for *C. jejuni* and the different *C. coli* clades. Asterisks denote significant difference between distributions as inferred from a Dunn's multiple comparisons test after a Kruskal-Wallis test, as follows: \*\*\*:  $p < 0.001$ ; \*\*\*\*:  $p < 0.0001$ .



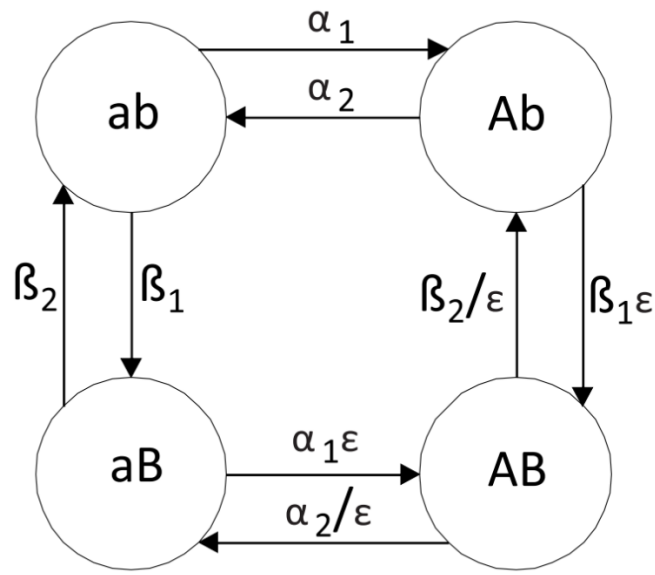
**Fig S3. Pan genome variation among *C. jejuni* and *C. coli* isolates**



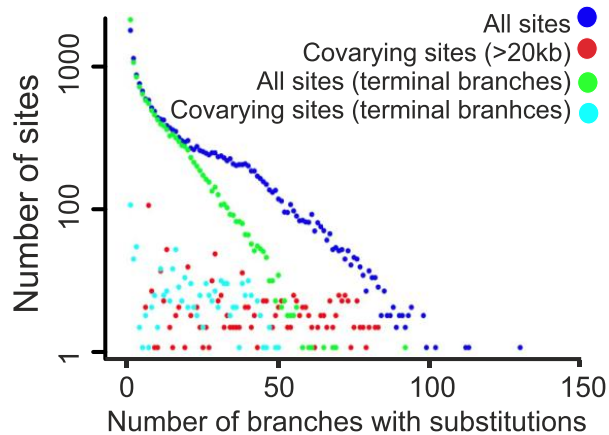
**Fig S4. Chromosome painting inference of recombination-derived *C. jejuni* sequence in recipient *C. coli* genomes.** (A) Length of DNA chunks recombining from 30 *C. jejuni* donor groups into *C. coli* ST-828 and ST-1150 clonal complexes and (B) the relative importance of different *C. jejuni* clonal complexes as donors.



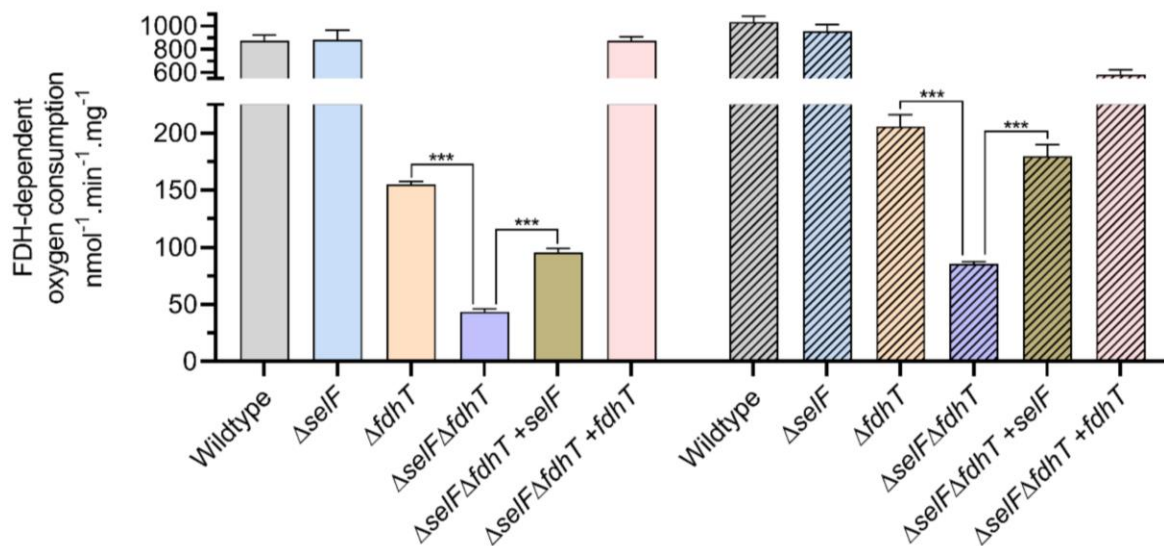
**Fig S5. Genome-wide introgression from *C. jejuni* to *C. coli*.** (A) The number of genes with different frequencies of maximum SNP introgression/gene in *C. coli* as a function of gene frequency in *C. jejuni*. Highly introgressed genes in CC-828 and CC-1150 tend to be core in *C. jejuni*. (B) Density plot (n=1000 bins) of specific and shared introgression events in CC-828 (x-axis) and CC-1150 (y-axis). The frequency of SNP introgression/gene is shown for both lineages. Close blue lines denote a high density of points. (C) Shared introgression between *C. coli* CC-828 and CC-1150. The number of SNPs being shared between the two lineages at various frequencies is shown in y-axis. (D) Pairwise nucleotide identity between *C. jejuni* and ancestral (unintrogresed) clade 1 *C. coli* core genes (black circles). Genes found to be introgressed in clade 1 CC-828 are highlighted in blue.



**Fig S6. Schematic of the continuous time Markov chain (CTMC) model used to analyse covariation in the *Campylobacter* genomes.** Covariation was assessed for pairs of biallelic sites separated by at least 20kb along the genome while accounting for the effect of population structure and recombination.

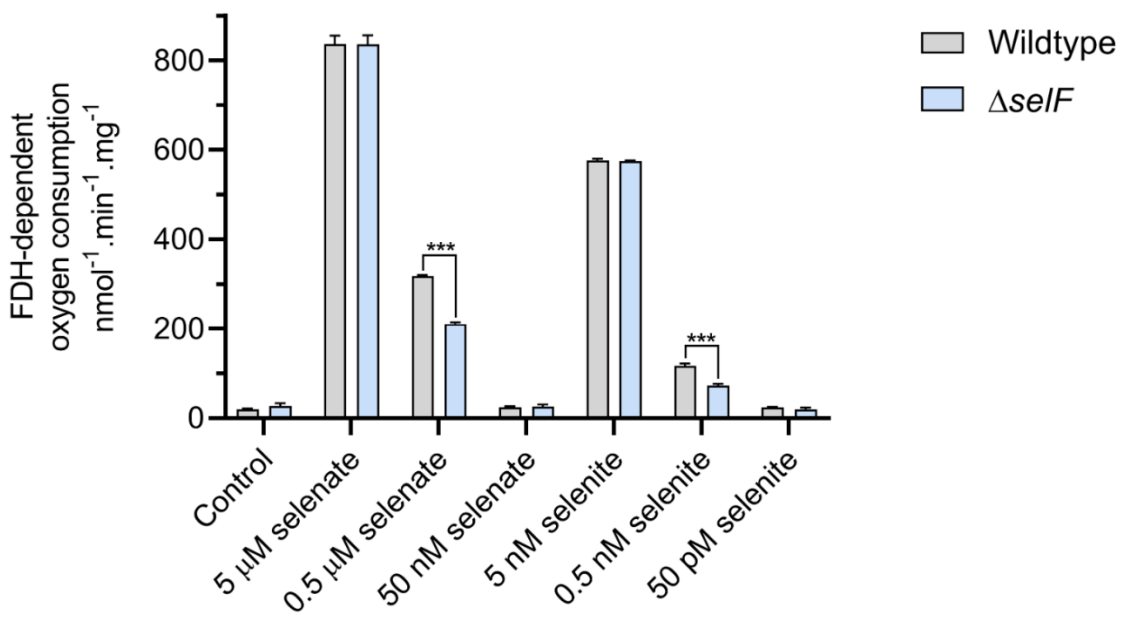


**Fig S7. The CTMC model identifying joint evolution of pairs of biallelic sites in *C. coli* CC-828 and CC-1150 genomes.** Covariation was assessed for pairs of biallelic sites separated by at least 20kb along the genome while accounting for the effect of population structure and recombination. There are many biallelic sites that do not change often on the tree and few that do. Putative epistatic sites change more frequently than average with biallelic pairs found together on multiple branches.

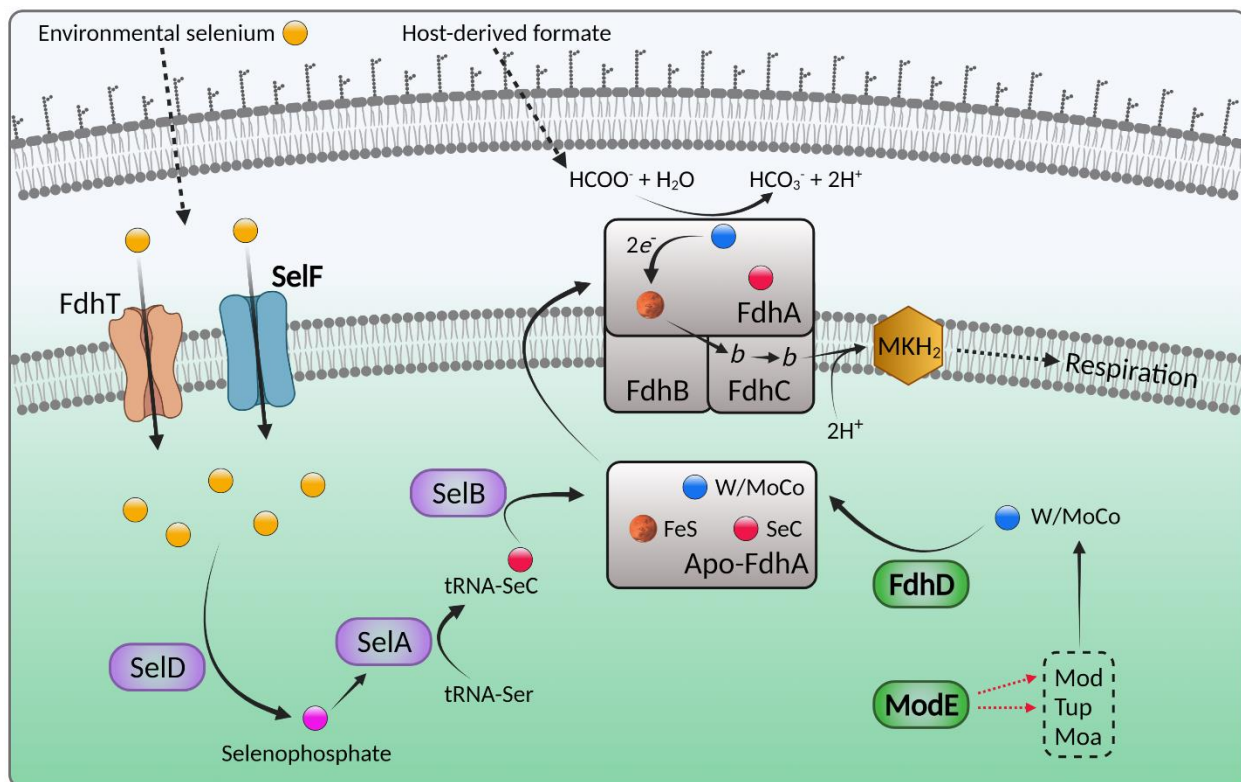


**Fig S8. SelF and the previously described FdhT both influence FDH activity.** Cells grown in minimal media with 0.5 nM sodium selenite, and (D) cells grown in minimal media with either 5 nM sodium selenite (left, open bars) or 5 μM sodium selenate (right, hashed bars). All data are means of at least 4 independent determinations and error bars are SD. \*\*\* denotes  $p$  value of  $<0.001$  by students  $t$ -test. Assays suggest that both FdhT and SelF facilitate selenium acquisition, possibly representing low and high affinity transporters, respectively.





**Fig S9. Effect of selenite or selenate concentration on the FDH activity of the *selF* mutant and parental wildtype grown in minimal media.**



**Fig S10. Model for epistatically linked genes involving FDH biogenesis and activity.** Host derived formate is converted to bicarbonate in the periplasm by the FDH complex to release electrons which are transferred from the iron sulfur (FeS) cluster of FdhA to the *b*-type hemes of FdhC and into the menaquinone (MK) pool where they can ultimately be used to reduce molecular oxygen via terminal oxidases. FdhA contains a selenocysteine residue and Mo/W-pterin cofactor (W/MoCo) at its active site, both of which are essential for catalysis. ModE is a DNA binding regulator which represses the Mo and W transporters Mod and Tup to regulate the cellular pool of Mo/W. W/MoCo is generated by the Moa pathway and inserted into apo-FdhA by the sulphur-transferase FdhD. Environmental selenite (the most abundant oxyanion) or selenate diffuses into the periplasm where it must be actively imported to the cytoplasm by FdhT and SelF. Putatively, FdhT is low affinity and functions efficiently when ample selenium is available. When selenium is limited, SelF can import sufficient selenium to maintain FDH production and activity. Cytoplasmic selenium is converted to selenophosphate by SelD, which is used by SelA to generate tRNA-SeC from tRNA-Ser. During translation of FdhA, tRNA-SeC is incorporated by the specific elongation factor SelB. Apo-FdhA with W/MoCo and SeC inserted is then transported to the periplasm by the TAT translocation system and incorporated into the FDH complex.

**Data S1. (separate file) Maximum SNP frequency and covarying p-value per gene.**

**Data S2. (separate file) Source data.**

**Data S3. (separate file) Genes with covarying SNPs identified in this study.**

**Data S4. (separate file) Isolates and genomes used in this study.**

**Data S5. Primers used in this study.**