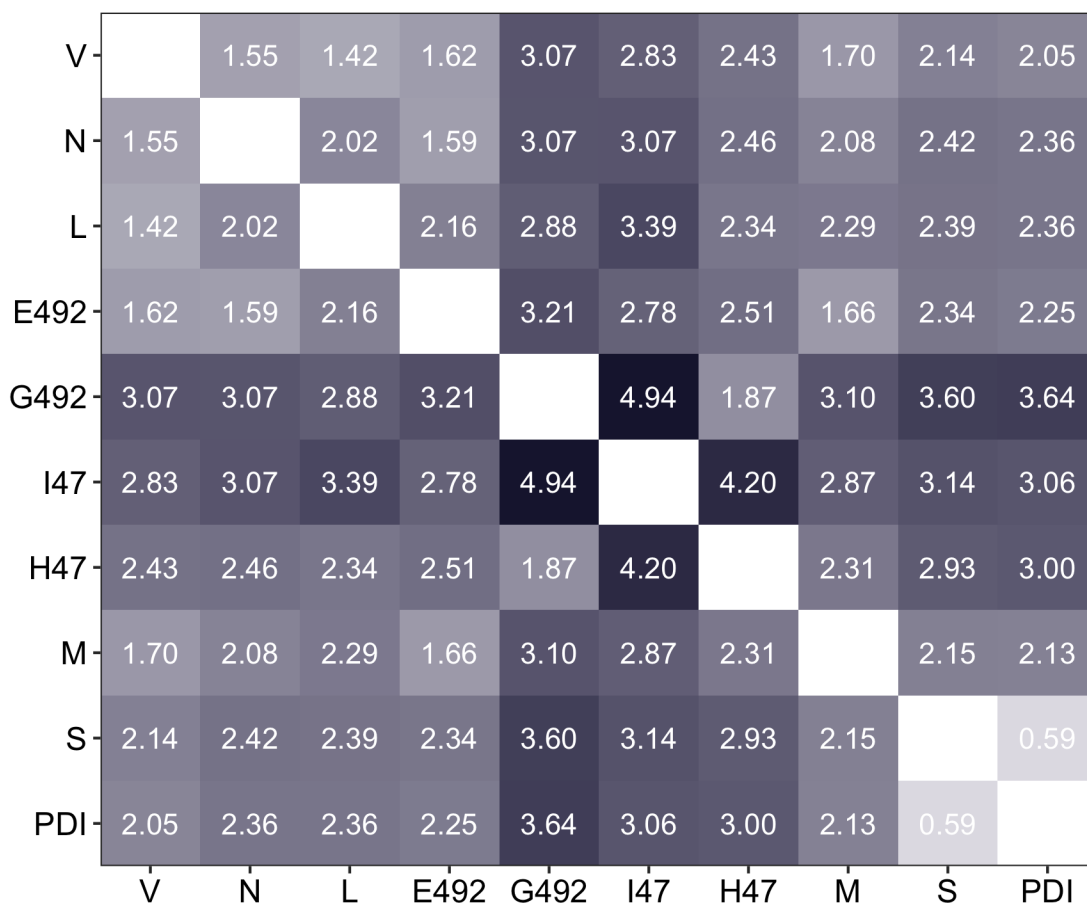


## Supplementary Figures and Files



**Supplementary Figure S1:** Embedding distance between all pairs of the ten known microcins.

V		72.82	50.00	70.43	63.46	70.48	69.52	63.55	71.20	68.50
N	72.82		66.67	49.02	67.02	73.40	69.23	71.57	72.13	70.49
L	50.00	66.67		65.77	66.67	72.38	68.87	71.03	72.36	72.36
E492	70.43	49.02	65.77		64.42	66.67	60.40	63.73	65.35	69.84
G492	63.46	67.02	66.67	64.42		71.43	58.24	65.59	71.67	73.33
I47	70.48	73.40	72.38	66.67	71.43		64.63	64.13	72.50	72.50
H47	69.52	69.23	68.87	60.40	58.24	64.63		59.79	69.17	68.33
M	63.55	71.57	71.03	63.73	65.59	64.13	59.79		70.00	68.55
S	71.20	72.13	72.36	65.35	71.67	72.50	69.17	70.00		20.00
PDI	68.50	70.49	72.36	69.84	73.33	72.50	68.33	68.55	20.00	
	V	N	L	E492	G492	I47	H47	M	S	PDI

**Supplementary Figure S2:** Percent sequence divergence between all pairs of the ten known microcins.

**Supplementary File S1:** Information about the dataset of 25 *E. coli* genomes. Contains the columns `name`, `strain_name`, `accession_number`, `strain_category`, `phylogroup`, `Genome_ID`, `microcin_hit_count`, `CvaB_hit_count`.

**Supplementary File S2:** Information about the dataset of *Enterobacter* genomes. Contains the columns `accession`, `ncbi_organism_name`, `cinful_CvaB_found` (indicates whether a CvaB was previously found by cinful—the value is TRUE for all genomes in this set), `cinful_microcins_found`.

**Supplementary File S3:** Information about the dataset of *Klebsiella* genomes. Contains the columns `accession`, `ncbi_organism_name`, `cinful_microcins_found` (number of microcins found by cinful), `cinful_CvaB_found` (indicates whether a CvaB was previously found by cinful—the value is TRUE for all genomes in this set).

**Supplementary File S4:** Cinful microcin and CvaB hits from the Touchon dataset. Contains the columns `name`, `strain_name`, `accession_number`, `strain_category` (origin of the bacteria), `phylogroup`, `Genome_ID`, `microcin_hit_count` (how many microcins were detected by Cinful), `CvaB_hit_count` (the number of microcin exporter proteins Cinful detected), `Assembly` (accession number for retrieval from NCBI).

**Supplementary File S5:** Cinful results for 40 water-sourced and extra-intestinal genomes from the Touchon dataset. Contains the columns `name`, `strain_name`, `accession_number`, `strain_category` (origin of the bacteria), `phylogroup`, `Genome_ID`, `microcin_hit_count` (how many microcins were detected by Cinful), `CvaB_hit_count` (if Cinful detected a microcin exporter protein), `Assembly` (accession number for retrieval from NCBI).

**Supplementary File S6:** Amino acid sequences of the ten known microcins, provided in FASTA format.

**Supplementary File S7:** Alignment of embedding hits collected from 25 *E. coli* genomes.

**Supplementary File S8:** Putative microcins found in the *E. coli*, *Enterobacter*, and *Klebsiella* datasets.

**Supplementary File S9:** Alignment of embedding hits collected from 44 *Enterobacter* genomes.

**Supplementary File S10:** Alignment of embedding hits collected from 46 *Klebsiella* genomes.