Supplemental Tables

Table S1: Strains and metadata
Table S2: Antibiotics and heavy metals used
Table S3: Phenotype data
Table S4: Correlation data
Table S5: Plasmids identified in the strains
Table S6: Resfinder data

Supplemental Figures

Fig S1: **Broad sense heritability** for each A) antibiotic condition and B) heavy metal condition. Data estimates the proportion of phenotypic variance attributable to genetic variation. Control data on MH (A) and NB (B) are shown for comparison.

Fig S2 **Correlation between Generation time and yield.** A: correlation between generation time (G) and yield (Y) is shown for the control conditions with no stressor. Frequency distributions G vs Y on no antibiotic is also shown for the two different media. B: Correlation between G and Y on stressors. All three concentrations are included (low, medium and high).

Fig S3 **Phenotypic analysis by isolation source.** A) PCA, B) box plot and C) heat maps showing median generation time and yield by source of isolation.

Fig S4 **Phenotypic analysis by decade of isolation.** Shown is a heat map of median phenotypes separated by decade of isolation; left, generation time and right, yield. Note that strains with no known date of isolation are not included.

Fig S5 **Correlation matrixes separated by source.** Animals (A and B), Environment (C and D) and Human (E and F). A, C and E) generation time, B, D, F) yield

Fig S6 **Correlation Matrixes of unnormalized ratios and without the ESBL strains** Correlation matrices for data not normalized to no stress condition A) generation time and B) yield. Correlation matrixes of data excluding the ESBL strains C1) generation time and C2) yield.

Fig S7: **Frequency distribution coverage.** The distribution of the coverages reveals that only 20 strains had coverages below 20, while the majority of the strains had a suitable high coverage with a mean of 50 and >30 for the 1944 strains.

Fig S8: **Missing rate distribution.** The missing rate of the reference sites' frequency distribution exhibits a multimodal distribution. This is to be expected because the reference strain is placed at the edge of the phylogenetic tree and strains that are farther away from it are thought to have undergone more sequential divergence and, as a result, have greater frequencies of missing sites. The strains with a greater rate of missing data were disregarded from the study using a threshold of 0.2 for the missing rate. However, most of the removed strains had a missing rate 0.6 or higher and were suspected as non-E. Coli species.

Fig S9: **Frequency distribution of de novo assembly sizes.** Assembly size represents the total number of assembled bases. The distribution of assembly sizes peaked at around 5 million

bases which is the assembly size of the reference strain. The assembly size varies between 4-6.5 * 10^6 bases with the distribution being skewed towards the higher values with ~1300 strains with assembly size > 5 * 10^6. The existence of a significant number of plasmids in *E. coli* could be one explanation for this.

Fig S10: **Coverage vs de novo assembly sizes.** Coverage vs De novo assembly size shows a very small negative correlation with a significant p-value (2.2 e-16). The sequencing quality seems to be robust to the size of the genome and a high coverage is observed even at the tail of the de novo assembly size distribution.

Fig S11: **Frequency distribution of reads quality**. The distribution of reads quality scores for the collection shows that all the strains had a significantly high quality score. The Reads Quality score (or confidence score) is the measure of probability of making an erroneous base call.

Fig S12 **Phylogenetic tree and PCA analysis of collection.** On the left side is the data using only SNPs and on the right Presence-absence. Groups are indicated by colors and are defined as shown.

Fig S13 **Copy number variation.** A) The figure shows the segmental duplication in various regions of the chromosome identified using a density based clustering algorithm called DBSCAN. The segmental duplications are present in 63 strains which are arranged on the y-axis according to their position in the phylogenetic tree and are shown in blue, green and yellow with copy numbers two, three and four respectively. B) The figure shows the segmental duplication in various regions of the chromosome identified using a tool called Control- FREEC. The segmental duplications are present in 66 strains which are arranged on the y-axis according to their position in the phylogenetic tree and are shown in blue, green, yellow, and red with copy numbers two, three, four, and five respectively. We observe some regions more prone to duplications compared to the rest and shows duplications in multiple strains.

Fig S14: **Principal component analysis for each stressor.** Colors indicate the phylogenetic group

Fig S15: **Heat map of D and Y for phylogenetic groups.** The median generation time (GT) and Yield is plotted for each phylogenetic group and condition

Fig S16: **Variation in phenotype for strains with known antibiotic resistance genes.** A) Phenotype on relevant antibiotics of strains carrying known antibiotic resistance genes is compared to the phenotype of those lacking this gene for A) aminoglycoside resistance genes, B) beta-lactamases and C) sulfonamide resistance genes.