

**Table 5. Unique genes in the UTI89 genome**

<u>Cutoff, %</u>	<u>Database</u>	<u>Unique genes in UTI89 (%)</u>	
		<u>Chromosome</u>	<u>Plasmid</u>
50	nr	134 (2.65%)	2 (1.38%)
	<i>E. coli</i>	351 (6.93%)	102 (70.3%)
80	nr	239 (4.72%)	10 (6.90%)
	<i>E. coli</i>	477 (9.42%)	112 (77.2%)

Protein sequences encoded by genes from the UTI89 chromosome and plasmid were compared with the GenBank nr database or the predicted proteins from the six published *E. coli* genomes (including two *Shigella* genomes) by using BLASTP. For each UTI89 protein, only the best BLASTP hit (based on E-value) was used. A gene was counted as “unique” at a certain percentage cutoff if (i) the alignment length (reported by BLASTP) divided by the length of the UTI89 protein sequence was less than the cutoff; or (ii) the percent identity (reported by BLASTP) was less than the cutoff. The number of unique genes at cutoffs of 50% and 80% is shown.