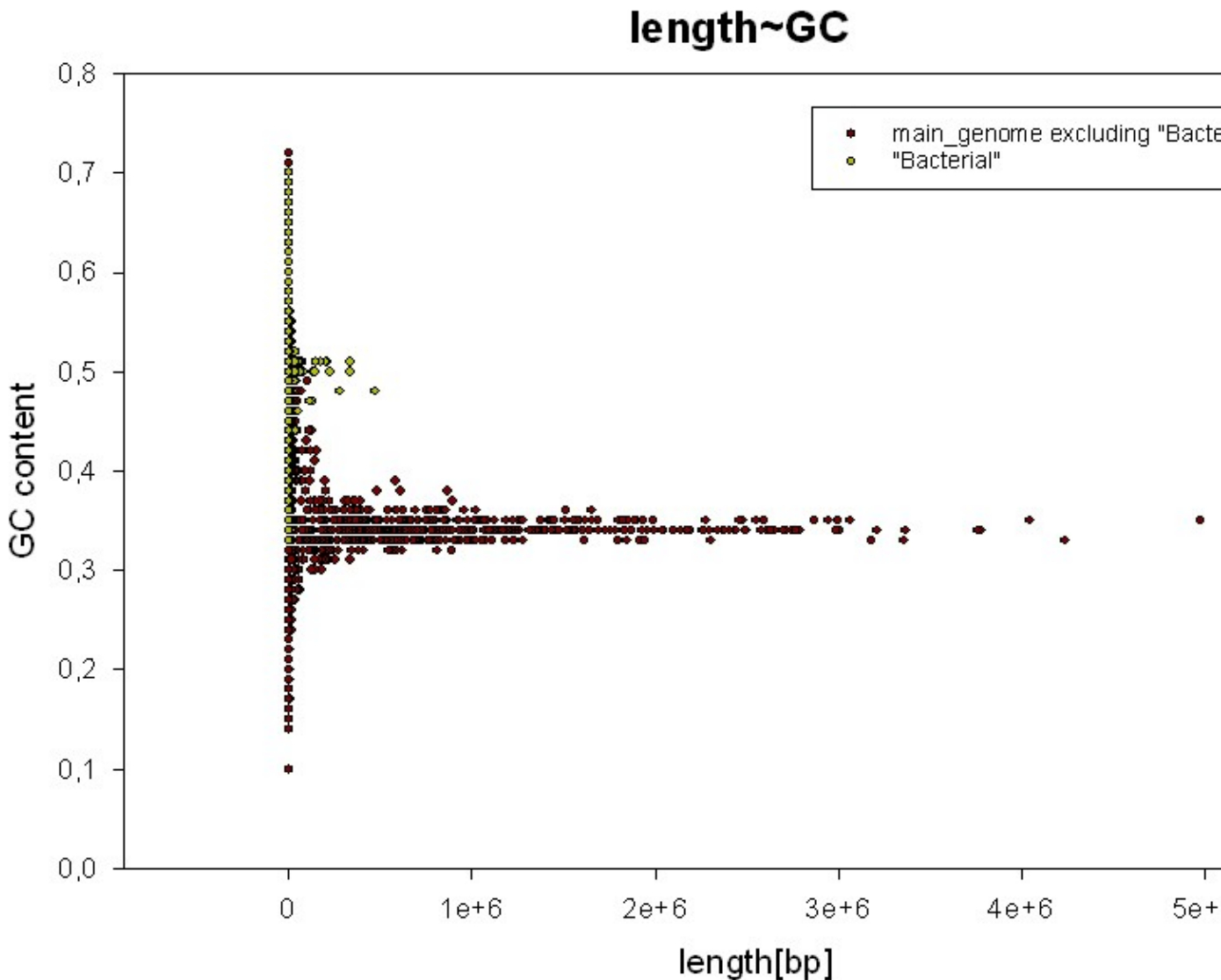


Looking at the GC/Length distributions of different scaffold populations

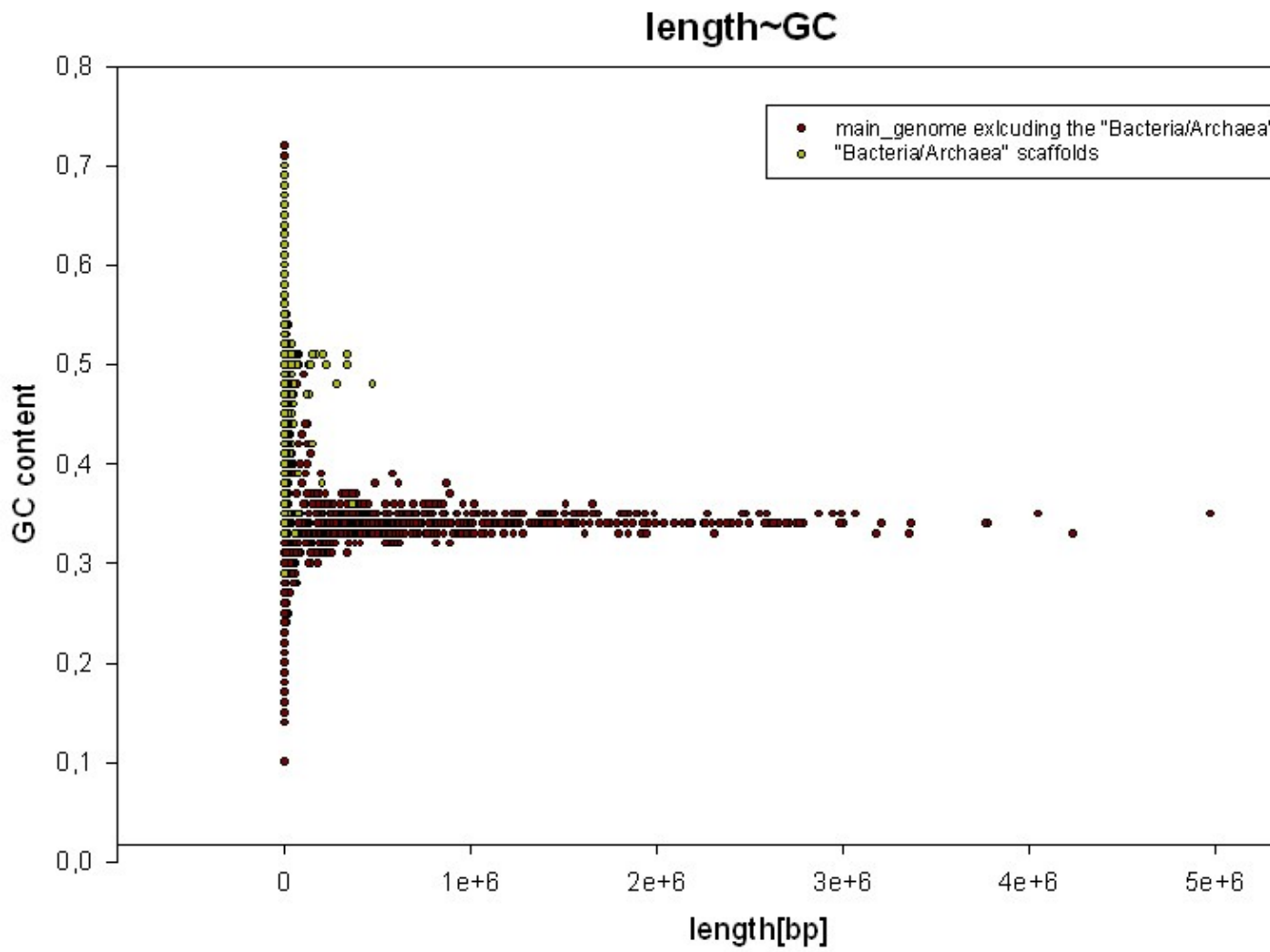
--Lang 14:14, 28 September 2006 (CEST)

Using the results from Jeffreys and Stefans analysis, I've had a look at the GC content and scaffold length and compared the following populations:

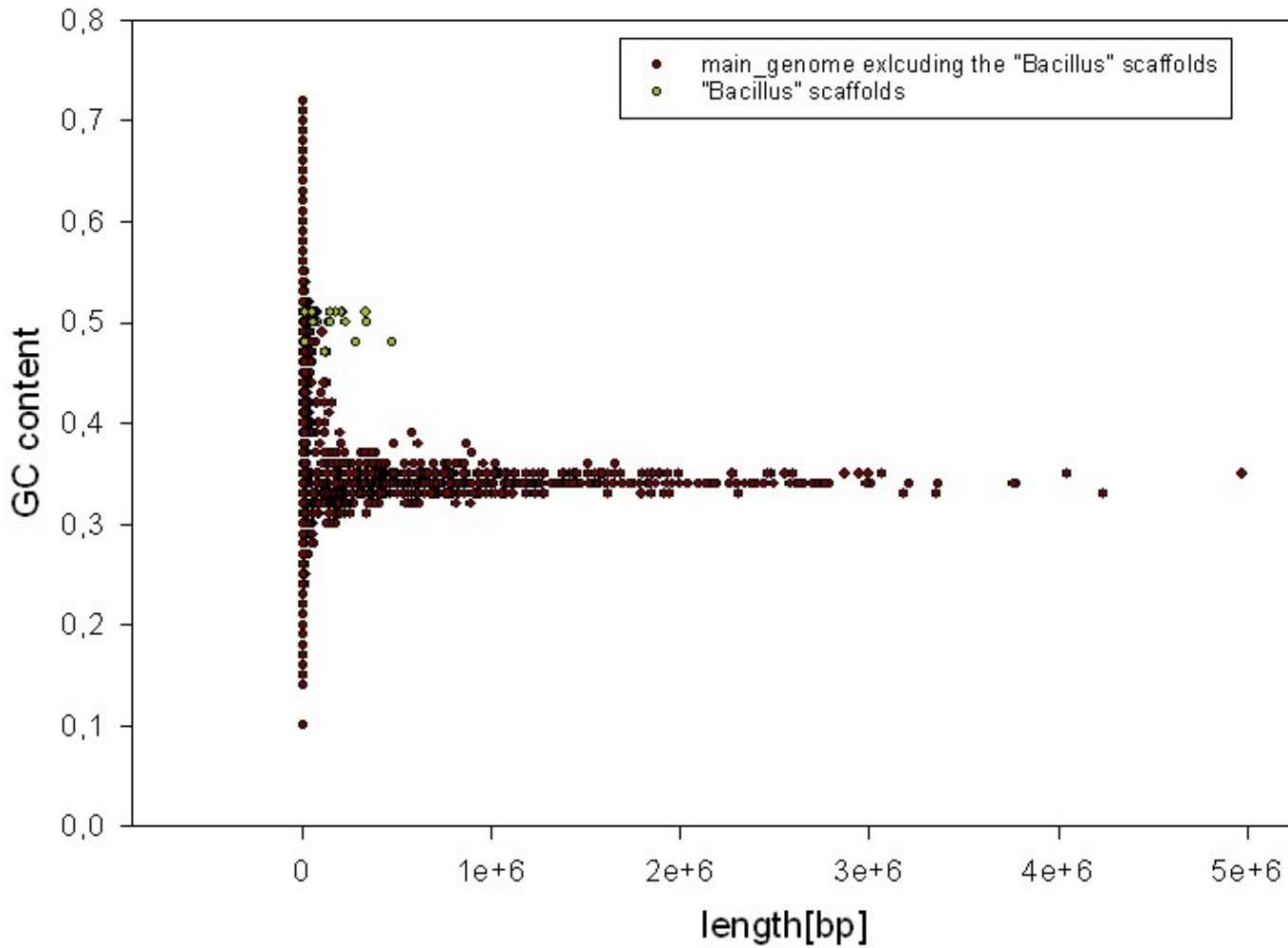
- stringent Bacteria (see below) [827 scaffolds]



- Archaea/Bacteria (more than eukaryotes) [950 scaffolds]



- Bacillus spec more than five hits [29 scaffolds]

length~GC

against the respective remainder of the main_genome scaffolds.

Looking at these plots, its quite obvious that we have definitely a set of scaffolds in the main_genome with small up to medium length, which have higher GC than the majority of scaffolds. We have to check wether these are contaminants.