Long repeats in bacterial genomes reveal alternative evolutionary mechanisms in *Bacillus subtilis* and other competent prokaryotes

Eduardo P.C. Rocha[1,2], Antoine Danchin[2] and Alain Viari[1]

[1]Atelier de BioInformatique, Université Paris VI, 12 Rue Cuvier, 75005 Paris, France and [2]Unité de Regulation de l'Expression Génétique, Institut Pasteur, 28 Rue Dr. Roux, 75724 Paris, France.

E-mail: erocha@abi.snv.jussieu.fr.

Prokaryotic genomes seem to be optimised towards compactness, and have therefore been thought to lack long redundant DNA sequences. To test this hypothesis we have put forward a methodology to analyse long strict repeats in bacterial genomes. Statistical significance of long repeats was calculated using an extreme statistics, considering a model where the nucleotide bias and the size of the genomes is taken into account. Long strict repeats were searched through the iterative use of an in house implementation of the Karp-Miller-Rosenberg (KMR) algorithm. Through this method we obtained the largest possible non-redundant repeats in the genome. In spite of the compactness of bacterial genomes, we have identified a large number of long strict repeats in eight complete genomes and have found that their density is anti-correlated with genome size. A detailed analysis of the long repeats present in the genome of *B. subtilis,* revealed a very strict constraint on the spatial distribution of repeats in this genome. We interpret this as the hallmark of selection processes leading to the addition of new genetic information. Such addition is independent of insertion sequences and relies on the non-specific DNA uptake by the competent cell and its subsequent integration in the chromosome in a circular form through a Campbell-like mechanism. Similar patterns are found in other competent genomes, suggesting a similar evolutionary mechanism.