Lifestyle, gene repertory and base composition bias in spirochetes
Natalia Rego¹, Martín Graña¹, Guillermo Lamolle¹², Fernando Alvarez-Valin², Hugo Naya¹
¹Unidad de Bioinformática, Institut Pasteur de Montevideo, Montevideo, Uruguay
²Sección Biomatemática, Facultad de Ciencias, UdelaR, Montevideo, Uruguay

Background
Extensive data support the notion that bacterial lineages undergo a process of gene repertory change when they shift from free-living or facultative parasitic lifestyle to an obligate host-associated one. Due to relaxation in selective pressures and increase in genetic drift effects related to changes in population structure, acquisition of novel genes by horizontal transference do not suffice to counteract gene decay and deletion events and this gene repertory dynamic ends in a final balance of gene loss and genomic reduction. Another signature of the new lifestyle seems to be a concomitant enrichment in the A+T genomic content, yet in this case, controversy between neutralist and selectionist explanations about its causes remains. Spirochetes represent an attractive opportunity to shed light on the forces shaping these genomic changes, since for genus as *Borrelia*, *Leptospira* and *Treponema*, different lifestyles and pathogenic features can be found at the intragenus level. In addition, reduced host range species do not show the parallel A+T base bias above mentioned, while harboring sharp differences in G+C content. For instance, *L. borgpetersenii* with a host to host transmission cycle is 17% smaller in genome size but 5% higher in G+C content than *L. interrogans*, which can survive more than 200 days in aquatic environments until infecting a new mammalian host.

Results
Due to the availability of an outgroup lineage with saprophytic lifestyle (*L. biflexa*), we focused on *Leptospira* comparative genome analyses. While *L. interrogans* presented the most biased base composition and despite its phylogenetic proximity to *L. borgpetersenii*, correspondence analysis of codon usage could split *L. borgpetersenii* from *L. interrogans* and *L. biflexa*. The pattern of preferred codons mainly correlates with G+C content but also with G+C skew. However, the studied strains cannot be split from an amino acid usage perspective. Searches for presence of structural domains as defined by SupFam [1] were also conducted. Strikingly, UmuC, a Y-family DNA polymerase involved in translesion bypass, with biased insertion of dAMP [2], seems absent from the *L. borgpetersenii* genome. Figure 1 shows UmuC presence among spirochetes.

![Figure 1](image)

**Figure 1.** Detected presence of UmuC in spirochetes (one or two gene copies) and correspondence with G+C content pattern.

Conclusion
Mutational bias related to UmuC presence-absence is a plausible mechanistic hypothesis for the uneven G+C content pattern observed within spirochetes. Further analyses regarding this result are currently being conducted.

References