Bipyrimidine Signatures as a Photoprotective Genome Strategy in G+C-rich Halophilic Archaea

Daniel L. Jones and Bonnie K. Baxter, Ph.D.

Background

Halophilic Archaea
- Experience high levels of ultraviolet (UV) radiation in their environments
- Demonstrate high resistance to UV
- Are protected by pigmentation and efficient DNA repair
- Have high genomic G+C content

UV-induced DNA Damage
- The predominant forms of UV-induced DNA damage are cyclobutane pyrimidine dimers (CPDs)
- These form between adjacent pyrimidines
- Bipyrimidine photoreactivity is in the descending order of: TC > TT > CT > CC
- Limiting of the most photoreactive sequences should reduce overall genomic photoreactivity

Results

Mean bipyrimidine incidences for each sample group. Error bars represent +/- 1.96 standard errors. Intergroup differences were assessed via one-way ANOVA and post-hoc Tukey contrasts. Halophilic archaea and enterobacteriaceae have significantly smaller P_g than archaea and cyanobacteria (p = 10^-4).

Conclusions

1. There is a strong, negative correlation between P_g and G+C content (Figure 6)
   - This may be explained by the fact that the most photoreactive sequences are T-containing
2. We found no evidence that UV exposure is a selective pressure for low P_g
   - Enterobacteriaceae have similar P_g to halophilic archaea
   - Cyanobacteria have significantly higher P_g than both
3. The UV-resistance observed in halophilic archaea can be attributed in part to a genomic strategy

Methods

Genome Sampling
- Sequences were obtained from the NCBI database
- Four our G+C content analysis, one representative genome for each prokaryotic species presently available was sampled at random
- For all other analyses, we randomly sampled 1 halophilic archaea strain per species, 1 (non-halophilic) archaea, cyanobacteria, and enterobacteriaceae strain per genus, and 101 bacterial strains of unique genus

Determining Bipyrimidine Incidences
- We wrote a word-counting script in R to determine bipyrimidine frequencies within sampled genomes
- Bipyrimidine incidences (TC, TT, CT, CC) were computed by dividing frequency by genome size in bases

Determining Theoretical Genomic Photoreactivity (P_g)
- P_g corresponds to the weighted sum of a genome’s bipyrimidine incidences:
  \[ P_g = 1.73(TC) + 1.99(TT) + 0.61(CT) + 0.39(CC) \]
- The weighting coefficients represent the intrinsic photoreactivity of each bipyrimidine, as determined by Matallana-Surget et al. (2008)

Acknowledgments

We would like to thank the Utah NASA Space Grant Consortium and the Lawrence T. Dee-Janet T. Dee Foundation for funding. In addition, we express much gratitude to Jaimi Butler, Kendall Tate, and Chrono Nu for their assistance in the development of this project.

References


