Amino Acid Usage Is Asymmetrically Biased in AT- and GC-Rich Microbial Genomes.

**Introduction:** Genomic base composition ranges from less than 25% AT to more than 85% AT in prokaryotes. Since only a small fraction of prokaryotic genomes is not protein coding even a minor change in genomic base composition will induce profound protein changes. We examined how amino acid and codon frequencies were distributed in over 2000 microbial genomes and how these distributions were affected by base compositional changes. In addition, we wanted to know how genome-wide amino acid usage was biased in the different genomes and how changes to base composition and mutations affected this bias. To carry this out, we used a Generalized Additive Mixed-effects Model (GAMM) to explore non-linear associations and strong data dependences in closely related microbes; principal component analysis (PCA) was used to examine genomic amino acid- and codon frequencies, while the concept of relative entropy was used to analyze genomic mutation rates. Results: We found that genomic amino acid frequencies carried a stronger phylogenetic signal than codon frequencies, but that this signal was weak compared to that of genomic %AT. Further, in contrast to codon usage bias (CUB), amino acid usage bias (AAUB) was differently distributed in AT- and GC-rich genomes in the sense that AT-rich genomes did not prefer specific amino acids over others to the same extent as GC-rich genomes. AAUB was also associated with relative entropy; genomes with low AAUB contained more random mutations as a consequence of relaxed purifying selection than genomes with higher AAUB. Conclusion: Genomic base composition has a substantial effect on both amino acid- and codon frequencies in bacterial genomes. While phylogeny influenced amino acid usage more in GC-rich genomes, AT-content was driving amino acid usage in AT-rich genomes. We found the GAMM model to be an excellent tool to analyze the genomic data used in this study.

**General Information**

State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis, Norwegian School of Veterinary Science
Contributors: Bohlin, J., Brynildsrud, O. B., Vesth, T. C., Skjrve, E., Ussery, D.
Publication date: 2013
Peer-reviewed: Yes

**Publication Information**

Journal: PLoS ONE
Volume: 8
Issue number: 7
Article number: e69878
ISSN (Print): 1932-6203
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): CiteScore 3.01 SJR 1.164 SNIP 1.111
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 3.11 SJR 1.236 SNIP 1.101
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): CiteScore 3.32 SJR 1.427 SNIP 1.136
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): CiteScore 3.54 SJR 1.559 SNIP 1.148
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 1
Scopus rating (2013): CiteScore 3.94 SJR 1.772 SNIP 1.153
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): CiteScore 4.15 SJR 1.982 SNIP 1.156
Web of Science (2012): Impact factor 3.73
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes