Evolution of complexity metrics of a genome under the effect of a substitution model

Román-Escrivá, P.[a,b], Arnau, V.[a], Díaz-Villanueva, W.[a], Moya, A.[a,b,c]

1. Institute for Integrative Systems Biology(I2SysBio), Universitat de València(UV) and Consejo Superior de Investigaciones Científicas (CSIC), Valencia, Spain
2. Genomic and Health Area, Foundation for the Promotion of Sanitary and Biomedical Research of the Valencia Region(FISABIO), Valencia, Spain
3. Centro de Investigación Biomédica en Red en Epidemiología y Salud Pública(CIBEResp), Madrid, Spain

INTRODUCTION

In the search for the assessment of a genome's complexity and its comparison between organisms, several complexity metrics have been proposed by the literature in recent years.

Genomic Signature (GS) is a k-mer-based metric, the value corresponding to the k-mer that maximizes the difference between observed and expected equifrequent classes of mers. This metric is based on the relative abundances of short oligonucleotides and chaos game representation applied to genomes.

BioBit is also a k-mer-based metric based on the difference between the maximum entropy for a k-mer of a random genome of the same length as the genome under consideration and the entropy of that genome for such a k-mer.

Sequence Compositional Complexity (SCC) is another metric that increases with the number of parts (i.e., compositional domains) and the length and compositional differences found in a genome sequence by a segmentation algorithm, paralleling the concept of 'pure complexity' of McShea and Brandon.

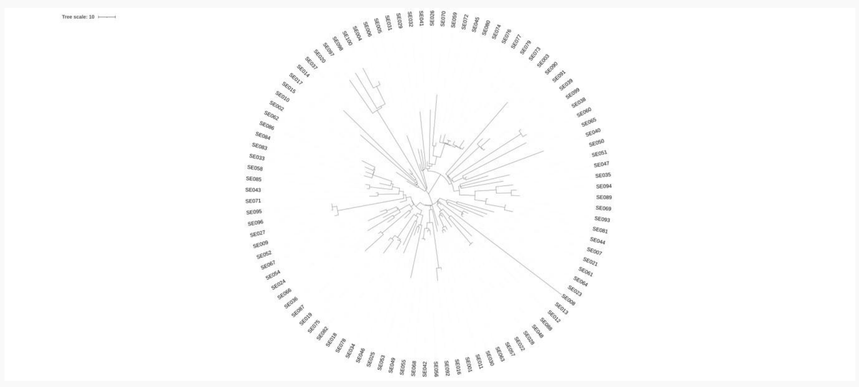
OBJECTIVES

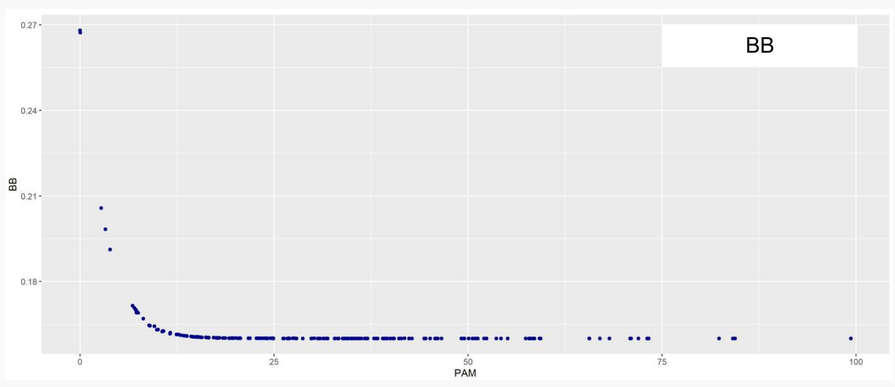
By applying random changes to an initial genome under a substitution model, its complexity should decrease with evolutionary time. This complexity will be measured with the three proposed metrics and a decrease in all metrics should be expected.

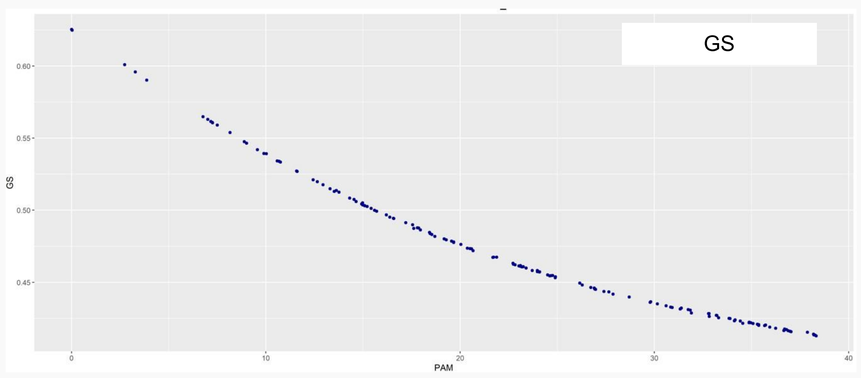
MATERIALS & METHODS

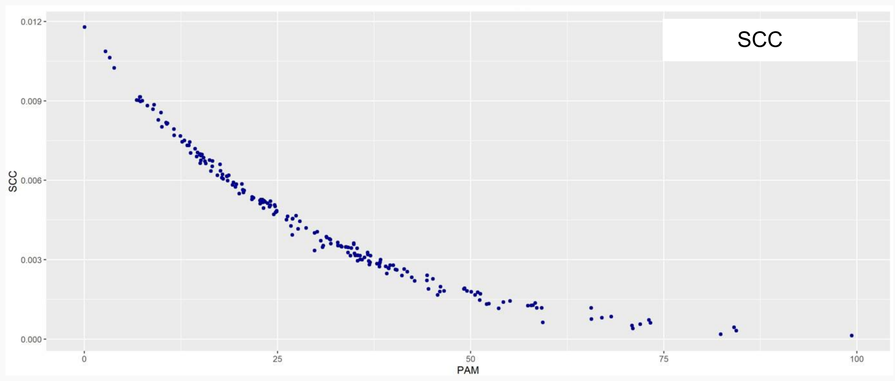
The initial genome is Escherichia coli K-12 and the substitution model is applied using ALF (A Simulation Framework for Genome Evolution). The substitution model used is GTR (R_{AC} = 0.3, R_{AG} = 0.8, R_{AT} = 0.4, R_{CG} = 0.5, R_{CT} = 0.7, R_{GT} = 0.3; $\pi_A = \pi_C = \pi_G = \pi_T = 0.25$) during 100 PAM.

RESULTS









CONCLUSIONS & FUTURE RESEARCH

We have observed that all the metrics decrease with evolutionary time. If they would be really measuring complexity, we could conclude that genome complexity also falls in the absence of natural selection. More research on these metrics is needed to ensure that they are actually measuring complexity and a future line of investigation could be the use of these metrics within certain evolutionary transitions associated with a theoretical gain or loss of complexity.

REFERENCES

1. Moya, A., Oliver, J.L., Verdú, M. et al. Driven progressive evolution of genome sequence complexity in Cyanobacteria. Sci Rep 10, 19073 (2020).

2. Daniel A. Dalquen, Maria Anisimova, Gaston H. Gonnet, Christophe Dessimoz, ALF—A Simulation Framework for Genome Evolution, Molecular Biology and Evolution, Volume 29, Issue 4, April 2012, Pages 1115–1123