

SORS: Language and Genetics

Abstract

Research of biology and genomics inspired from linguistics is a dangerous territory, for several reasons. Once these are identified, we can then distill the shared concerns and common features between linguistics and genetics. This will be the introduction to my talk. I will address the question of what is the nature of human language to finish with well a defined notion of interest in the generative modeling of the regulation of gene expression, including a brief discussion on tokenization of large language models (LLMs) of genomes. I will close my talk with what I consider a remarkable symbolic discovery.



Short Bio

Dr. Julio Collado completed his undergraduate, master's and doctoral studies at the National Autonomous University of Mexico (UNAM). After a three-year postdoctoral stay at the Massachusetts Institute of Technology (MIT) in Boston, he joined UNAM as a researcher at the current Center for Genomic Sciences. He is currently Distinguished Professor at UNAM and at the National System of Researchers. In his postdoc, Dr. Collado made the mathematical demonstration that justifies the use of grammatical models in genetic regulation and implemented the first generative grammar for the regulation of genetic expression, specifically of promoters of the sigma 70 family in *E. coli*. He made the first annotation and prediction of regulatory elements of transcription in complete genomes with the publication in *Science* of the complete genome of *E. coli* K-12 (1997). His group has been a pioneer in modeling and curation with RegulonDB and regulation editor in EcoCyc for almost 3 decades. He implemented predictions of operons, transcription factors and their DNA binding sites. In 2020, he directed her PhD student with a teams of experts and published a redefinition of the basic concepts of microbial regulation in *Nature Reviews Genetics*. Dr. Collado was a Guggenheim Foundation Fellow (1992); he was the Robert F. Kennedy Visiting Professorship at Harvard University (2007), and was awarded the National University Award in Natural Sciences (2004) and the National Award in Sciences and Arts in Physical-Mathematical and Natural Sciences (2011). He has been a driving force behind bioinformatics and genomics in Mexico, with the EMBNET-Mexico node; He served as director of the Center for Genomic Sciences (CCG) at UNAM (2005-2009) and was a member of the team that created the Bachelor's Degree in Genomic Sciences at UNAM, Cuernavaca. He was recognized as "ISCB Distinguished Fellow Class 2015" (the only Latin American to date). His research has been

supported by NIH grants for more than 20 years. Since 2018 he is an Adjunct Professor in the Department of Bioengineering at Boston University and since 2020 he is a Senior Visiting Scientist at the Center for Genomic Regulation, in Barcelona, Spain. He was founding President of the Ibero- American Society of Bioinformatics in 2009. Graduates of his laboratory are or have been researchers in Canada, France, Belgium, Spain, the United States and Cuba.

Speakers

Speaker: Julio Collado Vides, Professor of Computational Genomics, CCG, UNAM

Host: Salvador Capella, Group Leader-INB - Life Sciences , BSC

Barcelona Supercomputing Center - Centro Nacional de Supercomputación

Source URL (retrieved on 3 Abr 2025 - 14:37): <https://www.bsc.es/es/research-and-development/research-seminars/sors-language-and-genetics>