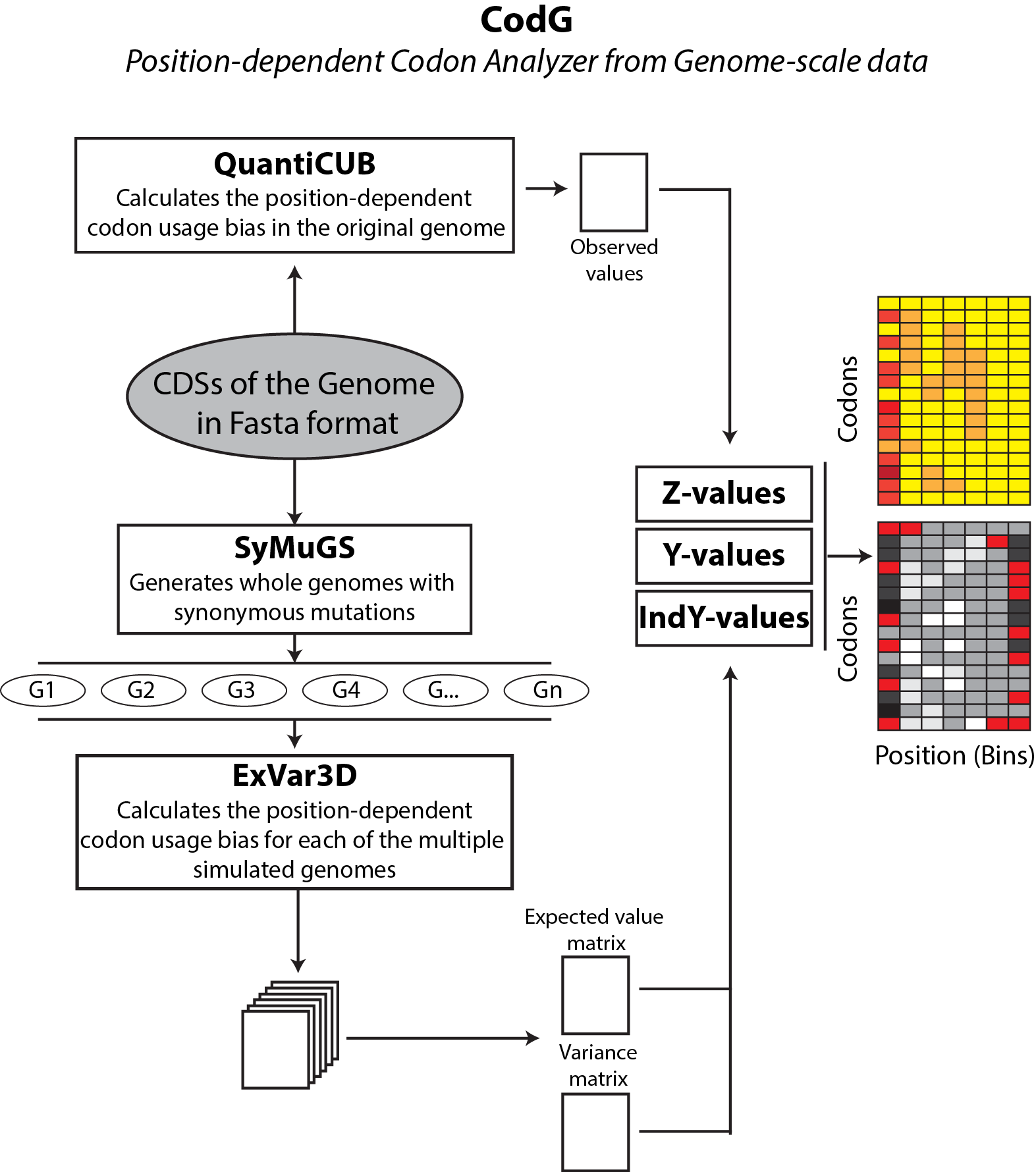
**CodG: *Position-dependent Codon Analyzer from Genome-scale data***

*Code availability:* [*https://github.com/juanvillada/CodG*](https://github.com/juanvillada/CodG)



***CodG*** is a program developed in R language that automates the testing procedures of the preferred codons usage as a function of the position within genes using genome-scale data.

This program is based on SeqinR and BioStrings packages. The program uses six algorithms: ***SyMuGS*** (Synonymous Mutated Genome Simulator), ***QuantiCUB*** (Position-dependent Codon Usage Bias Quantifier), ***ExVar3D*** (Expected Value and Variance calculator), ***Z-values***, ***Y-values*** and ***IndY-values***.

The algorithm ***SyMuGS*** simulates complete genomes introducing random synonymous mutations in each of the coding sequences. The preferred codon usage bias of each of the sequences is conserved. This algorithm uses the file containing coding sequences in FASTA format as input and generates as output a different file in FASTA format for each simulated genome.

***QuantiCUB*** quantifies the codon usage bias as a function of the position an according to the length of each coding sequence in the genome, using as input the FASTA file format containing the coding sequences of the genome to be analyzed. It generates the “Observed\_pdCUB” array of 10 positions for 59 codons in comma separated values ​​(CSV) format.

The ***ExVar3D*** algorithm calculates the position-dependent codon usage bias for each of the multiple simulated genomes, having as input the sequences of the simulated genomes in FASTA format which are obtained with the ***SyMuGS*** algorithm. The program is based on ***QuantiCUB*** algorithm, but other than this, the ***ExVar3D*** generates a virtual tridimensional matrix with the quantification of codons of each genome, and from this generates other two arrays: one expected value of the matrix of each codon by position and an array of variance codons for each position. The output format of these two matrices is CSV.

***Z-values***, ***Y-values*** and ***IndY-values*** are algorithms implemented to create comprehensive figures in order to illustrate the complex results obtained in the whole previous experimentation.

Therefore, ***CodG*** compares the codons distribution differences of the original genome with a large number of simulated genomes, allowing the analysis of the evolutionary selective forces acting on individual codons. This program is a useful tool for the analysis of genes and genomes from microorganisms by different approaches such as evolutionary research, expression of recombinant proteins and synthetic biology.