Without high quality gene models, scientists won’t be able to design experiments targeting their genes of interest. Traditionally validating those gene models require manual curation. This is a labor-intensive and time-consuming process in which one or a few individuals evaluate and correct the computational predictions by using all the available evidence they can find.

Before the manual curation takes place, automated gene finders identify the parts of a sequence that encodes genes. These pipelines are becoming faster as the algorithms improve and more accurate as they are able to incorporate more biological data to use in their predictions. But they still make errors.

We tested two distinct approaches to identify mispredictions from automated gene finders: the quality values generated from a gene annotation pipeline: MAKER-P, and the alignments between the translated protein sequences and its homologs across species.

We selected a subset of genes from the most recent maize reference genome annotation that was analyzed by a group of students. Our results showed that all the gene models analyzed using these two methods had errors, and gave the students the opportunity to learn and support the community curation of a very relevant eukaryotic genome. Ideally this method could be used as a means to let students and even citizen scientists participate in the annotation of any sequenced eukaryotic genome.