

An attempt to create the animal genomic data as a Linked Data

KAWASHIMA, TAKESHI^{1*}, KOYANAGI, Ryo¹, GYOJYA, Fuki¹

¹Okinawa Institute of Science and Technology

The pace of DNA decoding is growing faster and faster in the recent years according to improvement of the DNA sequencers. The animal genome which has relatively large size of DNA are not exceptional in this aspect. Prediction of genes, prediction of proteins and assignment of gene expression data are required tasks for more analysis of genome sequences after the genome decoding. But under the present circumstances, the pace of such predictions and annotations are not catching up to the pace of the data creation. Especially in the recent animal genome projects, the quality of gene predictions and annotations show the deteriorative because only small number of researchers can be involved.

The accuracy of genemodels depends on the accuracy of the previous gene models, because each genemodel is created by the comparative analysis with the known predicted genemodel. So in sometimes, the low quality genemodel create the next low quality genemodels, as it is called "Junk makes Junks". Such low quality genemodels could prompt a deadly mistakes for the comparative genome analyses although the comparative extant animal genomes are required to know the evolution of animals.

For getting away from such issues, we are now re-curating the genemodels for the variety of published animal genomes from the aspect of evolutionary biology. Additionally, we start an attempt to create the Linked data from the animal genomic data for making them convenient to the variety of comparative genomics.

Keywords: animals, genome, annotation, ontology, linked data, semantic web