

International Conference on

COMPUTATIONAL BIOLOGY AND BIOINFORMATICS

September 05-06, 2018 Tokyo, Japan

ETC: A toolkit for converting phenotype descriptions into computable data**Hong Cui¹, Thomas Rodenhause¹, Bertram Ludäscher², James Macklin³ and Nico Franz⁴**¹University of Arizona, USA²University of Illinois at Urbana-Champaign, USA³Agriculture and Agri-Food, Canada⁴Arizona State University, USA

The explorer of taxon concepts project has produced a web application that consists a set of five tools unlocking phenotype data from text narratives often found as taxonomic descriptions or character descriptions. Aside from the tools described below, the site supports division of labor by allowing users share their tasks. Text capture tool parses textual taxonomic descriptions and marks up anatomical entities and characters. The tool is powered by CharaParser and MicroPIE (for microbial taxonomic descriptions). Ontology building tool enables experts without ontology knowledge to organize a set of phenotypic terms (e.g. those discovered by using is a part of or synonym relationships). Resulting ontology can be used in the other tools to improve data quality. Matrix generation tool takes the output and assembles a raw taxon-character matrix for the user to edit and refine, with or without ontology. Key generation tool employs a novel algorithm that directly takes a taxon-character matrix with polymorphic characters as input and computes the information entropy scores for each character. The result is a multi-access key that order the characters based on their discrimination power within the pool of the taxa. Taxonomy comparison tool supports the task of taxon concept analysis by supporting both expert asserted RCC-5 (congruent, narrower/broader than, disjoint and overlap) relationships among taxa, as well as providing character data related to the taxa in question to aid expert decisions. These five tools can be used to construct several pipelines that generate a taxon-by-character matrix, create a multi-access identification tool or facilitate taxon concept comparisons. A related software application, matrix converter is public available for users to convert a raw matrix to a scored matrix for phylogenetic analysis.

hongcui@email.arizona.edu