

BIOINFORMATICS FOR ASSESSMENT OF REPRODUCTIVE COMPATIBILITY BETWEEN CROPS AND THEIR RELATIVES. Richard G. FitzJohn, Aaron D. Wilton, Tristan T. Armstrong and Linda E. Newstrom, Researchers, Landcare Research, Lincoln 8152, New Zealand.

Risk assessments of transgenic plants often rely on analyses of existing bodies of data; as the quantity of available data continues to expand, a major challenge is bringing information together in a format that can be conveniently updated and then reanalysed in future. Data may come in a variety of formats, and may be reported with varying levels of detail and accuracy. Moreover, depending on the required analysis, different pools of data may be appropriate to use.

We have designed an attributes database to address these challenges. Our primary design principles were that information should be entered with the least modification possible, and that information be clearly linked to its source. In particular, information is captured in a way that is entirely separate from its analysis and presentation, allowing for completely alternative use of the data in the future. Central to our database is an entity management system, which allows differentiation of taxon names at any taxonomic rank, common name and/or cultivar. In comparison with conventional databases, variables are easy to add, ordered hierarchically, and associated with metadata to facilitate information organisation.

We have applied our database to collecting and analysing published information on hybridisation between crops and their relatives, focusing on *Brassica* and allied genera. We distinguished between experimental hybridisation (manual cross-pollination), spontaneous hybridisation (unassisted pollination), and studies using *in vitro* methods (e.g. embryo rescue), and between different generations (F₁, F₂ and BC₁). In total, we databased 322 papers documenting hybridisation among species in *Brassica* and allied genera, covering 216 species combinations. We have also applied our approach to eight other crop groups, the largest of which was onion and related crops in *Allium* (34 papers). Because data from hybridisation experiments are reported in a wide variety of formats, substantial effort was required to standardise them. To deal with this, we developed a framework for extracting, analysing and presenting hybridisation data. Such transformations are required to present complex data regardless of method of storage, but our approach makes these transformations explicit.

A key advantage of our approach is that by retaining as much information as possible, decisions that affect the quantity of data available can be made at analysis rather than collection, and be based on the needs of the analysis. This increases transparency and repeatability of analyses, and allows future analyses to use alternative transformations. Using a database approach increases the ease of reusing data, and with more data available in this form some of the international duplication of effort in assembling and updating information could be avoided.