Application of the international crop information system for retrieval and usage of pedigree and phenotypic data for use in durum research and breeding

Clarke F.R., Clarke J.M., Ruan Y., Lin X., N'Diaye A., Kthiri D., Pozniak C.J., Weibe K., Yates S.

in

Porceddu E. (ed.), Damania A.B. (ed.), Qualset C.O. (ed.).
Proceedings of the International Symposium on Genetics and breeding of durum wheat

Bari : CIHEAM
Options Méditerranéennes : Série A. Séminaires Méditerranéens; n. 110

2014
pages 315-318

Application of the international crop information system for retrieval and usage of pedigree and phenotypic data for use in durum research and breeding

Fran R. Clarke[1], John M. Clarke[2], Yuefeng Ruan[2], Xue Lin[2], Amidou N'Diaye[2], Dhouha Kthiri[2], Curtis J. Pozniak[2], Krystalee Weibe[2], Shawn Yates[1]

[1] Semiarid Prairie Agricultural Research Centre , Agriculture an Agri-Food Canada, Swift Current, Canada
[2] Department of Plant Sciences and Crop Development Centre, University of Saskatchewan, Canada

Abstract. The International Crop Information System (ICIS) links pedigrees to phenotypic and genotypic data that is made easily accessible. The Genealogy Management System (GMS) interfaces 'Central' public databases that can have various levels of access, and 'Local' private databases within institutions. This facilitates the global sharing of non-sensitive pedigrees, selection histories and other descriptors in the Central database while interfacing with Local databases which contain sensitive data. Canadian public-sector durum wheat researchers have set up databases containing extensive phenotypic data (agronomic, disease and end-use functionality) from cultivar registration and pre-registration trials conducted over the past 25 years. Pedigree information on the lines is uploaded to the globally-available Central database so that pedigrees of Canadian lines can be traced back to ancestors and easily visualized for presentation purposes. The power of this system was demonstrated by tracing the sources of the low grain cadmium concentration allele by selective phenotyping of the lineages of diverse durum cultivars, and management of data for association mapping studies for discovery of QTL for biotic stresses.

Keywords. *Triticum turgidum* L. var *durum* – Data management – Pedigree – Grain cadmium.

Application du système international d'information sur les cultures pour la récupération et l'utilisation des données de pédigrées et phénotypiques pour une utilisation dans la recherche du blé dur et la sélection

Résumé. *Le système international d'information sur les cultures (ICIS) fait le lien entre les pédigrées et les données phénotypiques et génotypiques qui sont rendues facilement accessibles. Le système de gestion des données généalogiques (GMS) fait l'interface entre des bases de données publiques « centrales », qui peuvent avoir différents niveaux d'accès, et des bases de données privées «locales», au sein des institutions. Cela facilite le partage mondial des pédigrées non-sensibles, des histoires de sélection et d'autres descripteurs présents dans la base de données centrale, tout en permettant l'interfaçage avec des bases de données locales qui contiennent des informations sensibles. Au Canada, les chercheurs du secteur public qui travaillent sur le blé ont mis en place des bases de données réunissant de nombreuses données phénotypiques (aspects agronomiques, maladies et fonctionnalité à l'utilisation finale) provenant des essais pour l'homologation et la pré-homologation des cultivars, réalisés ces 25 dernières années. Les informations sur le pédigrée des lignées sont chargées dans la base de données centrale, disponible à l'échelle mondiale, pour pouvoir remonter aux ancêtres et visualiser les pédigrées des lignées canadiennes à des fins de présentation. La puissance de ce système a été démontrée en traçant les sources des allèles de la faible concentration de cadmium dans le grain à travers un phénotypage sélectif des lignées de divers cultivars de blé dur, et la gestion des données pour des études de cartographie d'association visant la détermination de QTL pour des stress biotiques.*

Mots-clés. *Triticum turgidum L. var durum – Gestion des données – Pédigrée – Cadmium du grain.*

I – Introduction

The International Crop Information System (ICIS) links pedigrees to phenotypic and genotypic data for easy querying and use. The Genealogy Management System (GMS) permits various levels of access when interfacing 'Central' public databases (e.g., IWIS, International Wheat Information System), and 'Local' private databases within institutions. This facilitates the global sharing of, for example, non-sensitive pedigrees and selection histories in the Central database while interfacing with Local databases which may contain sensitive data (6.5M). This system has the advantages of accommodating input of different synonyms depending on the source of information, and allowing determination of the wheat pedigrees by many experts from around the world.

ICIS is a very useful repository of valuable data that were used, in addition to the primary function as a breeding tool, for estimation of genetic gain (Clarke *et al.,* 2010) and association genetics studies (Pozniak *et al*., 2012). In this report, we look at usage of ICIS pedigree and molecular databases to track sources of the low grain cadmium concentration allele in diverse durum lines, and the phenotypic database to manage data for association mapping studies.

Cadmium is a heavy metal potentially toxic to humans (McLaughlin *et al.,* 1999) that occurs naturally in trace quantities in almost all soils. North American durum wheat (*Triticum turgidum* L. var *durum*) was reported to accumulate higher Cd levels in grain than hexaploid wheat (*T. aestivum* L.) (Zook *et al*.,1970). Breeding for low grain cadmium concentration began in Canada in the early 1990s (Clarke *et al.,* 2010) due to observed genetic variation in durum wheat cadmium concentration (Penner *et al.,* 1995), which was simply inherited (Clarke *et al*., 1997). This led to the commercialization of the now widely-grown low cadmium cultivar Strongfield (Clarke *et al.,* 2005). However, the ancestral source of the low cadmium trait is not known.

Resistance to biotic stresses is a long-standing focus of Canadian durum breeding. The wheat stem sawfly (*Cephus cinctus* Norton) is a major insect pest of the Canadian durum growing area. Changes in North American races of leaf rust (*Puccinia triticina* Eriks.) and stripe rust (*Puccinia striiformis* f. sp. *Tritici*) are a concern due to widespread virulence on Canadian durum germplasm. Discovery of QTL for resistance to these biotic stresses is a high priority using bi-parental populations and association mapping panels, with data management facilitated by ICIS.

II – Material and methods

Canadian public-sector durum wheat researchers maintain databases containing extensive phenotypic data (agronomic, disease and end-use functionality) from cultivar registration and pre-registration trials conducted over the past 25 years. Pedigree information on the lines is uploaded to the globally-available Central database so that pedigrees of Canadian lines can be traced back to ancestors and easily visualized for research purposes at (https://www. integratedbreeding.net/crop-information/wheat).

The origin of the low cadmium allele was traced in the cultivar Biodur (Valdur//Wascana/Durtal), which originates from Germany. It was used as a donor of the low cadmium trait and stem solidness in Canadian durum breeding programs. Solid stems reduce damage by the wheat stem sawfly. The pedigree of Biodur in ICIS was updated or corrected where necessary following cross-checks with the European Wheat Database of the European Cooperative Programme for Plant Genetic Resources (http://genbank.vurv.cz/ewdb/), research publications and communication with durum wheat breeders. The Draw Tree option in ICIS was used to generate seven generations of pedigrees of Biodur to track the origin of the allele. As many key ancestors as possible were obtained for testing cadmium genotype. Multiple accessions of each ancestor were tested where possible to detect heterogeneity. Seed not already in our possession was obtained from the

USDA Small Grains Collection, Plant Gene Resources Canada, or directly from durum breeding colleagues. Lines and accessions were tested with the marker *XBF474090* (Weibe *et al*., 2010) to detect presence of the low cadmium allele.

Other capabilities of ICIS are demonstrated, such as calculation of coefficient of parentage among solid-stem lines in an association mapping panel and production of publication quality tables of germplasm lists with pedigrees and origins.

III – Results and discussion

Biodur, by its pedigree, could derive the low cadmium allele from hexaploid wheat, Durtal having come from the cross of a semidwarf hexaploid with the durum Sentry. Durtal, however, turned out to have high cadmium phenotype (Fig. 1). Valdur was therefore the donor of the low allele, possibly obtained from M'Rari via D117. However, Sterpe 131066, which we could not obtain, could be a co-donor. Biodur is the donor of the solid stem trait and co-donor of low cadmium in two new durum cultivars CDC Fortitude (Pozniak *et al*., unpublished) and AAC Raymore (Singh *et al,.* unpublished).

In other tested lineages (not shown), we were able to demonstrate that the low cadmium allele was in some cases obtained from hexaploid wheat parents. Hexaploid wheats were the sources of Rht-B1b dwarfing genes in durum, so the low cadmium allele in many CIMMYT durum lineages may derive from those crosses. We are conducting further testing to confirm our preliminary observations. This demonstrated the utility of ICIS to generate the pedigrees of low cadmium durum cultivars. Combined with molecular information available in ICIS, these data were a powerful tool to determine the likely source of the allele conferring low grain cadmium concentration.

We made extensive use of ICIS for retrieval of phenotypic data from Canadian durum registration trials for association mapping studies (Pozniak *et al*., 2012) using DArT markers. The phenotypic information comprising agronomic, end-use quality and disease resistance data for 14 years of the registration trial were retrieved. The data consisted of individual replicated plot data for agronomic and disease traits, and within location or among location composite data for end-use quality. The 'SETGEN' feature of ICIS was used to create a publication-ready list of the lines in the study, together with pedigrees and origin of the material (Table 1). The same dataset is being used to evaluate the recently-available genotypic data from the Infinium 90K iSelect array.

ICIS was also used for management of data and pedigrees for an association mapping panel used for study of linkage disequilibrium (Somers *et al*., 2007) and identification of QTL associated with semolina yellow pigment (Reimer *et al*., 2008). The same panel of lines is currently being used for association mapping of stripe rust resistance with the Infinium 90K iSelect array and for validation of markers for stem solidness. Phenotyping of the panel for stem solidness identified several solid stem accessions from Italy. The coefficient of parentage (COP) was calculated among these lines using ICIS (Table 2), and with the recently released Canadian cultivars CDC Fortitude and AAC Raymore, and their ancestor Biodur. The COP showed that Biodur is not closely related to the Italian lines. In contrast, Biodur showed a larger COP with all three Canadian lines because it is the source of solidness, and Biodur also has a Canadian ancestor (Wascana). The COP of the Italian cultivars Fortore, Lesina and Mongibello was greater than 0.3 as would be expected given their similar pedigrees.

In summary, ICIS is a useful tool in our durum breeding and research activities. Use of ICIS offers time efficiencies over maintenance of phenotypic, molecular and pedigree data in spreadsheets or text files. Local curation of our database enables correction of errors as they are found by users, thus providing users with access to the most up to date version of all data sets. Use of ICIS for data management also ensures access to properly annotated datasets by future users,

overcoming the all too common situation where data are lost when a researcher moves to a new job or retires.

Acknowledgments

We gratefully acknowledge funding provided by Genome Canada to the CTAG project, as well as the by the Western Grains Research Foundation, the Province of Saskatchewan, and Agriculture and Agri-Food Canada.

References

Clarke J.M., Leisle D., Kopytko G.L., 1997. Inheritance of cadmium concentration in five durum wheat crosses. *Crop Sci*., 37, pp. 1722-1726.

Clarke J.M., McCaig T.N., DePauw R.M., Knox R.E., Clarke F.R., Fernandez M.R., Ames N.P., 2005. Strongfield durum wheat. *Can. J. Plant Sci*., 85, pp. 651-654.

Clarke J.M., Clarke F.R., Pozniak C.J., 2010. Forty-six years of genetic improvement in Canadian durum wheat cultivars. *Can. J. Plant Sci*., 90, pp. 791-809.

McLaughlin M.J., Parker D.R., Clarke J.M., 1999. Metals and micronutrients - Food safety issues. *Field Crops Res*., 60, pp. 143-163.

Penner G.A., Clarke J., Bezte L.J., Leisle D., 1995. Identification of RAPD markers linked to a gene governing cadmium uptake in durum wheat. *Genome,* 38, pp. 543-547.

Pozniak C.J., Clarke J.M., Clarke F.R., 2012. Potential for detection of marker-trait associations in durum wheat using unbalanced, historical phenotypic datasets. *Mol. Breed.,* 30, pp. 1537–1550.

Reimer S., Pozniak C.J., Clarke F.R., Clarke J.M., Somers D.J., Knox R.E., Singh A.K., 2008. Association mapping of yellow pigment in an elite collection of durum wheat cultivars and breeding lines. *Genome,* 51, pp. 1016-1025.

Somers D.J., Banks T., DePauw R., Fox S., Clarke J., Pozniak C., McCartney C., 2007. Genome-wide linkage disequilibrium analysis in bread wheat and durum wheat. *Genome,* 50, pp. 557-567.

Wiebe K., Pozniak C.J., Harris N., Faris J., Clarke J.M., Knox R.E., Taylor G., 2010. Saturation mapping of *Cdu1*, a major locus for grain cadmium concentration in durum wheat (*Triticum turgidum* L. var *durum*). *Theor. Appl. Genet*., 121, pp. 1047–1058.

Zook E.G., Greene F.E., Morris E.R., 1970. Nutrient composition of selected wheats and wheat products. VI. Distribution of manganese, copper, nickel, zinc, magnesium, lead, tin, cadmium, chromium, and selenium as determined by atomic absorption spectroscopy and colorimetry. *Cereal Chem*., 47, pp. 720-731.