T3/Oat: Integrating public genotyping efforts and historical research

Clare Saied[1], David Matthews[2], Clay Birkett[2], Calvin Lietzow[2], James Clohessy[1], Shiaoman Chao[3], Nicholas Tinker[4], Jean-Luc Jannink*[2]

[1]Cornell University, Ithaca NY, USA; [2]USDA-ARS, Ithaca NY, USA; [3]USDA-ARS, Fargo ND, USA; [4]Agriculture and Agri-Food Canada, Ottawa ON.

*JeanLuc.Jannink@ars.usda.gov

About T3/Oat

T3/Oat (https://triticeaetoolbox.org/oat/) is the repository of international oat phenotype and genotype data for the Oat Global project. Data stored in T3/Oat is open access and all users are able to submit data, establishing the database as a tool to support coordinated or individual research efforts. T3/Oat provides flexible queries for extracting desired datasets for analysis as well as integrated visualization tools for comparing trials, and tools for genomic association and prediction (figure 1). Applying modern computational techniques to large-scale, shared data in this manner will support scientific publications and variety releases, and will maximize the returns on the increasing volume of data produced by the oat community.

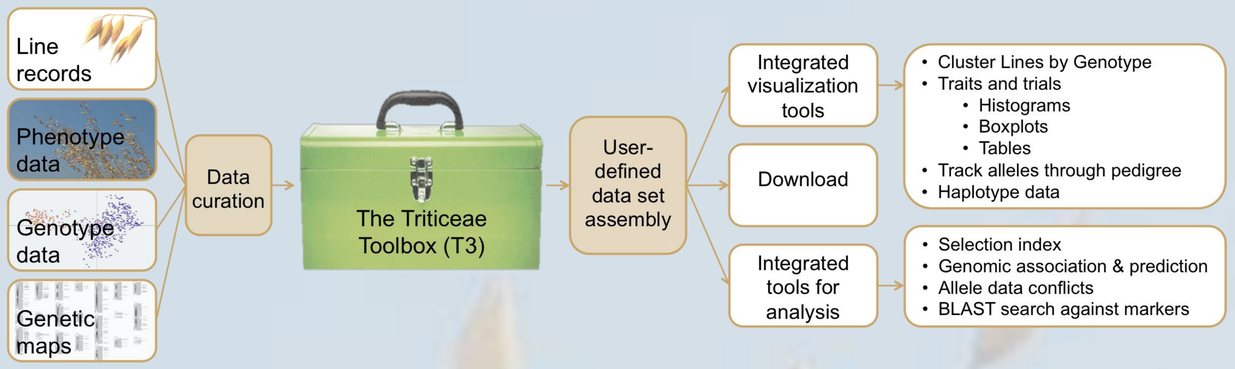

Figure 1. *Curated line information, phenotype and genotype data, and genetic maps can be uploaded to T3/Oat. Flexible data selection queries are available to assemble a data set, and this data can then be downloaded or analyzed using one of the integrated tools for data visualization or analysis.*

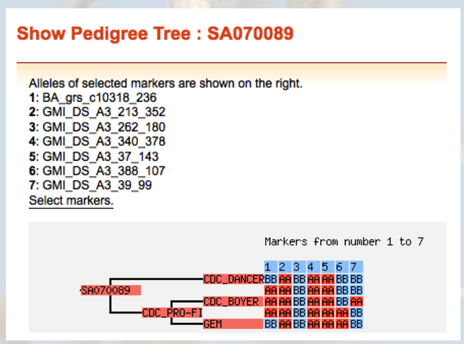

Figure 2. *The T3 graphical pedigree can also display marker values and is reached through a T3 line record or using the "Track Alleles Through Pedigree" tool.*

Line records

There are 16,394 line records stored in T3/Oat. Line records hold information on the breeding program, generation, and pedigree of a line. A graphical tree of the pedigree can be reached from the line record (figure 2), and there are links to GRIN and/or a record in the Pedigree Of Oat Lines (POOL) database when possible. Available phenotype and genotype data can be accessed from the line record. Finally, information may be supplied on a range of line characters, which can be used to select lines using the "Select Lines by Properties" tool.

Table 1. *The source of T3/Oat phenotype trials.*

Data Source	Trials
Collaborative Oat Research Enterprise	82
National Small Grains Collection	12
Uniform Early Oat Performance Nursery	273
Uniform Midseason Oat Performance Nursery	514
Uniform Oat Winter Hardiness Nurseries	61
Uniform Winter Oat Yield Trials	185
Total number of phenotype trials	**1,127**

Phenotype trials

T3/Oat holds data from 1,127 publicly funded phenotype trials (table 1). The T3 "Selection Wizard" allows users to select data from phenotype trials based on the program that submitted the lines, the program that collected the data, the location of the trial, or by the traits that were measured. The "Select Lines by Phenotype" tool allows users to select lines that performed within a phenotypic range for a given trait and trial combination.

Genotype data

Data from 19 genotype experiments are currently available in T3/Oat (table 2). Genotypic data can be selected in T3 using the "Select Lines by Genotyping Experiment" tool.

Table 2. *The source of T3/Oat genotype trials.*

Data Source	T3 experiments	
	Infinium	GBS
Collaborative Oat Research Enterprise (CORE)	1	1
National Small Grains Collection (NSGC)	1	-
Bi-parental mapping populations from Oliver *et al* (2013)	6	-
Public Oat Genotyping Initiative (POGI) trials	-	10
Total number of genotype experiments	**8**	**11**

Genetic maps

There are two oat consensus maps currently available in T3/Oat (Table 4). The first is composed of the framework markers from the consensus map developed by Chaffin *et al*. (2016). The twelve component maps used to construct this map are also available in T3/Oat.

Additional marker calls in mapping parents and breeding populations have been made using the new software 'Haplotag' (Tinker *et al*, 2016) and placed on an expanded version of the oat consensus map (Chaffin *et al*, 2016). This expanded map is available in T3/Oat and will be critically examined in a forthcoming manuscript (Bekele *et al*, unpublished).

Genetic maps can be downloaded directly from T3, or downloaded along with genotype data, and can be selected for use with the "Genomic Association and Prediction" T3 tool.

Table 4. *The genetic maps currently available in T3/Oat.*

Genetic map		Number of markers
Framework Oat Consensus Map (Chaffin *et al*, 2016)		7,202
Expanded Oat Consensus Map (Bekele *et al*, unpublished)		50,668
12 component maps published in Chaffin *et al*. (2016)	Kanota x Ogle	1,914
	CDC Sol-Fi x HiFi	888
	Hurdal x Z-597	1,508
	Ogle x TAMO-301	2,257
	CDC Boyer x 94197A1-9-2-2-5	660
	Otana x PI269616	1,166
	Provena x 94197A1-9-2-2-5	1,821
	IL86-1156 x Clintland 64	623
	Provena x CDC Boyer	598
	Dal x Exeter	895
	AC Assiniboia x MN841801	1,366
	IL86-6404 x Clintland 64	608
Total number of mapped markers		**52,680**

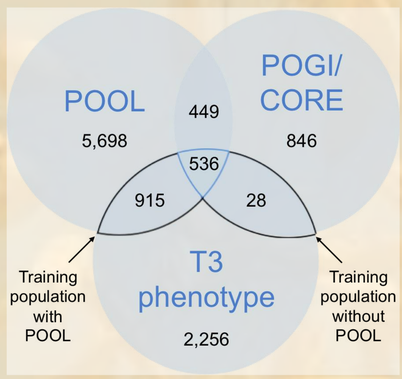

Figure 3. *The number of lines with T3/Oat phenotype data, POGI or CORE genotype data, and pedigree information in the POOL database.*

The Public Oat Genotyping Initiative

The Public Oat Genotyping Initiative (POGI) is an ongoing project to carry out high-density genotyping of oat breeding lines submitted by public sector breeding programs. Fourteen North American breeding programs submitted a total of 1,427 lines to the 2015 project. Library construction and sequencing was carried out in USDA-ARS Fargo. The Haplotag pipeline, developed by Tinker *et al.* (2016), was used for GBS SNP calling. Pedigree information and genotype data from the 2015 POGI are now publicly available through T3/Oat.

Preliminary analyses have been carried out using the genotype data generated by POGI and by the CORE project, combined with phenotype data stored in T3/Oat and POOL pedigree information.

Figure 4 shows the imputation accuracy that was achieved when GBS data generated by POGI and the CORE project (1881 lines, 68294 markers: 36.2% missing, 1.1% heterozygous) was used to impute CORE Illumina Array SNPs (635 lines, 1930 markers: 0.1% missing, 0.2% heterozygous).

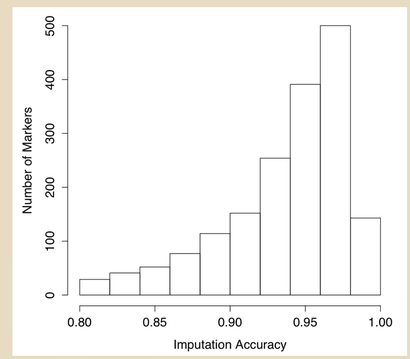

Figure 4. *The distribution of imputation accuracy when estimating Illumina Array SNPs generated by the CORE project using GBS data from POGI and the CORE project.*

Figure 5 shows the correlation between relationship matrices that were constructed using POGI GBS data, CORE Array SNPs or POOL pedigrees. A low correlation score indicates disagreement between the two sources of information that are being compared. The two sources of marker information have a stronger positive correlation to each other than with the pedigree information.

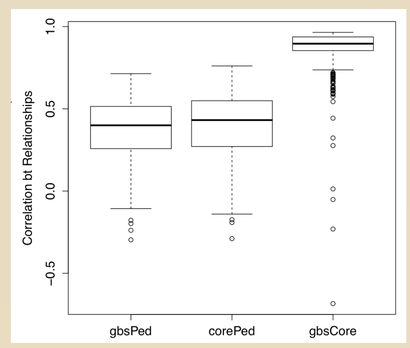

Figure 5. *Three relationship matrices were created using POGI GBS markers, CORE Array SNPs and POOL pedigrees. Each box plot shows the correlation in the results from two of these data sources.*

Finally, it was possible to carry out preliminary genomic prediction of yield, beta-glucan, heading date and height on spring oat. A core training population of 536 lines, lines that have both phenotype and genotype data available in T3/Oat and pedigree information in POOL (figure 3), was split into five folds and sequentially each fold was set to missing and their values imputed. The remaining lines in the overall training population (i.e. the core set - fold + 28 lines without POOL data; the core set - fold + 915 lines with POOL data) were used to fit the prediction model. The prediction accuracy was calculated as the correlation between the 536 predictions and the 536 observations (table 3).

Table 3. *Genomic prediction of oat traits using pedigree data from POOL, and genotype and phenotype data stored in T3/Oat.*

	Yield	Beta-Glucan	Heading Date	Height
Pedigree	0.48	0.38	0.55	0.41
Genotype	0.66	0.53	0.69	0.58
Both	0.67	0.54	0.68	0.56

These preliminary analyses demonstrate how new genotyping methods can be used to add further utility to historical phenotyping efforts in oat.

Marker records

There are 862,438 markers stored in T3/Oat. Marker records contain sequence information as well as map locations when available. The line records of GBS markers that were called using HaploTag also contain a link to the HaploTag passport file (figure 6).

T3 lines can be selected by haplotype, based on up to five markers of interest, using the "Select Lines by Haplotype" tool.

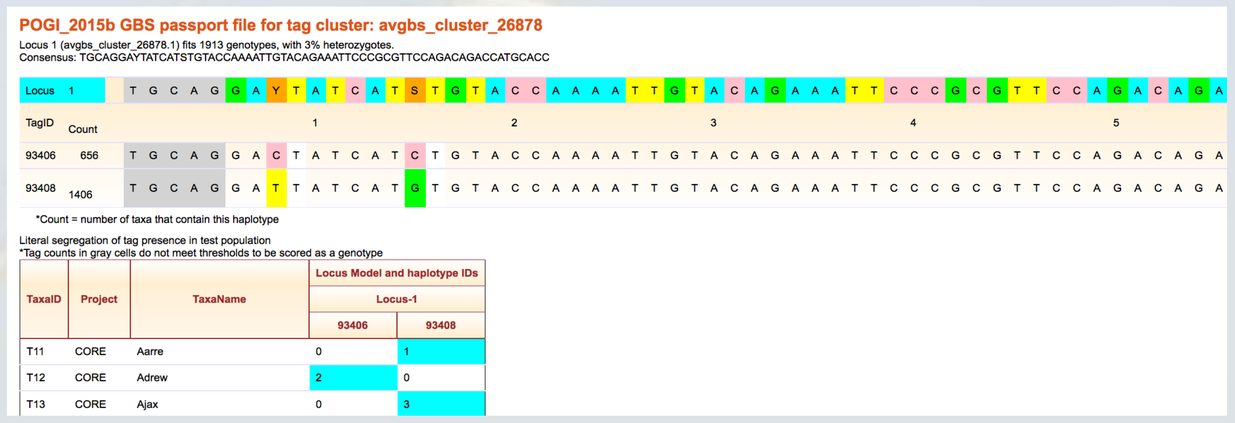

Figure 6. *An example of an HaploTag passport file. Here, two tags (potential haplotypes) are identified. The position of the two SNPs (Y and S) are identified by color. The table shows the tag counts at the presumed haplotypes within the locus. Counts greater than or equal to one are shaded i.e. scored as "present".*

References

Chaffin, A.S., Huang, Y.-F., Smith, C.A.S., Bekele, W.A., Babiker, E., Gnanesh, B.N., Foresman, B., Blanchard, S., Jay, J., Reid, R.W., Wight, C.P., Chao, S., Oliver, R.E., Islamovic, E., Kolb, F.L., McCartney, C., Mitchell Fetch, J.W., Beattie, A.D., Bjørnstad, A., Bonman, J.M., Langdon, T., Howarth, C.J., Brouwer, C., Jellen, E.N., Klos, K.E., Poland, J., Hseih, T.-F., Brown, R., Jackson, E.W., Schlueter, J.A., and Tinker, N.A. (2016) A consensus map in cultivated hexaploid oat reveals conserved grass synteny with substantial sub-genome rearrangement. *Plant Genome*, pp. 1-35. doi : 10.3835/plantgenome2015.10.0102

Tinker, N.A., Bekele, W.A., and Hattori, J. (2016) Haplotag: Software for Haplotype-Based Genotyping-by-Sequencing Analysis. G3: Genes, Genomes, Genetics. **6**:857-863.