Using high-throughput phenotyping and genome wide association study (GWAS) techniques to identify molecular markers for transpiration efficiency in wheat

Andrew Fletcher[1], Alison Kelly[2,3], Jack Christopher[2], Valeria Paccapelo[3], Karine Chenu[1]

[1]The University of Queensland, Queensland Alliance for Agriculture and Food Innovation (QAAFI), 203 Tor Street, Toowoomba, Queensland 4350, Australia. Email: karine.chenu@uq.edu.au
[2]The University of Queensland, QAAFI, Leslie Research Facility, PO Box 2282, Toowoomba, QLD 4350, Australia
[3]Department of Agriculture and Fisheries, Leslie Research Facility, Toowoomba, QLD, Australia

Abstract
While drought is typically limiting for wheat productivity in rain fed cropping systems around the world, climate models predict more variable rainfall as well as increased temperature and evaporative demand in major production regions. Breeding for greater transpiration efficiency (TE) has been proposed as a means to improve crop yield in water-limited environments, especially where crops rely on stored soil moisture. Using a recently developed high-throughput lysimeter platform and a genetically-structured mapping population, two experiments were conducted to explore genetic variability of TE in wheat. Using GWAS techniques, 17 QTL for TE were identified. These QTL were detected using a nested association mapping (NAM) population, in which a high yielding cultivar in the northern region of the Australian wheatbelt (Suntop) was crossed to 10 donor lines each having adaptive traits including some with adaptation for heat and drought stress. Thus, identified QTL have potential to assist breeders to improve TE and yield in wheat for drought-prone environments, particularly in the northern region of the Australian wheatbelt.

Key Words
Transpiration efficiency, water-use efficiency, water use, genomics, phenotyping, lysimeter.

Introduction
Bread wheat (*Triticum aestivum* L.) is one of the most significant staple food crops, providing up to 18.3% of daily calorie intake and 19.5% of daily protein intake globally (FAO 2018). In the past century, global population has more than doubled, reaching 7.5 billion people in 2017 and is predicted to reach between 9.3-9.8 billion by the year 2050 (Fischer *et al.* 2014, FAO 2018). Within the same time frame, global climate change is predicted to have unfavourable impacts in major food producing regions of the world (IPCC 2014), including Australia (Lobell *et al.* 2015, Watson *et al.* 2017, Wang *et al.* 2018). Projections for greater variability in water availability for crops and higher evaporative demand make it increasingly necessary to improve water use efficiency of crops. Increasing transpiration efficiency (TE) has the potential to increase or maintain crop yields in drought environments, especially where crops rely on stored soil moisture (Condon *et al.* 2002, Fletcher and Chenu 2015, Chenu *et al.* 2018, Ababaei and Chenu 2019).

While plant genomics and gene editing are revolutionising plant breeding, the current limitation in the application of genomic tools is typically the ability to collect meaningful phenotypic data on large numbers of genotypes (Chenu *et al.* 2009, Araus and Cairns 2014). Here, we illustrate how the combination of a high-throughput TE phenotyping platform (Chenu *et al.* 2018, Fletcher *et al.* 2018) and genome wide association studies (GWAS) can be used to identify QTL for TE in wheat.

Material and Methods
Growing conditions and experimental design
Using a genetically structured nested association mapping (NAM) population with known interrelatedness and genetic diversity, two experiments were conducted in an 560-pot automated lysimetry platform, located inside a polyhouse at the University of Queensland, Gatton Campus (Chenu *et al.* 2018). Four plants of the same genotype were grown in each 4L pot. Plants had unrestricted access to water, and a layer of plastic wrap was applied to the soil surface of all pots to minimize evaporation of water from the soil surface and to control weed growth.

The NAM population consisted of 457 recombinant inbred lines (RILs) from a cross between Suntop and donor lines varying for drought adaptation: Drysdale (52 lines), Dharwar Dry (52), FAC10-16 (42), EGA Gregory (42), EGA Wylie (40), RIL114 (42), SB062 (52), Seri M82 (52), ZWB10-37 (41) ZWW10-128 (42). To maximise the throughput of the platform, two experiments were conducted within a single season, following

the methodology of Fletcher *et al.* (2018). The two experiments (Exp1 and Exp2) were fully randomized. All genotypes were replicated at least twice across both experiments (2.32 replicates per NAM line). The Drysdale and Suntop parents were replicated 9 and 10 times, respectively, and all other parental lines were replicated four times. For each experiment, an additional pot with no plants was used to measure any background moisture loss.

Measurements and analysis

Water use was measured automatically by the lysimetry platform every 10 minutes from sowing until harvest, and summed for the entire period (minus background water loss measured in unsown pots) to give the cumulated water used per pot. Weekly observations of Zadoks stage (Zadoks *et al.* 1974) were taken on two plants per pot for the parental lines, and for one plant per pot for all NAM lines at harvest. Both experiments had similar development rates and were harvested at a similar physiological stage from approximately flag-leaf expansion to mid-booting stage. At harvest, above-ground plant material was dried at 70^{o}C for 5 days and then weighed for dry biomass. TE was calculated as the ratio of above-ground dry biomass produced per cumulative gram of water transpired from sowing until harvest. All data analysis was performed using ASReml-R (Butler *et al.* 2009) in the R software environment (R Core Team 2018). As all studied traits were highly correlated between the two experiments, the data from both experiments were combined to calculate empirical Best Linear Unbiased Estimators (BLUEs) for each genotype. To account for the large degree of genetic diversity and the high level of relatedness of NAM lines, a multi-locus mixed-model (MLMM) GWAS technique (Segura *et al.* 2012) was used with the BLUEs to identify QTL for TE in the NAM population. All QTL discovered had a $-\log_{10}P$ greater than 2 ($P < 0.01$) (

the methodology of Fletcher *et al.* (2018). The two experiments (Exp1 and Exp2) were fully).

Results and Discussion

A total of 23 most significant markers for TE were identified on chromosomes 1D, 3A, 3B, 3D, 5A, 5B, 5D, 6A, 6B, 6D, 7A, 7B and 7D, with some of these being found in close proximity in several genomic regions (Table 1). Where multiple most significant markers were determined to be in close proximity with an overlapping range of nearby markers were also found as significant, they were aggregated into a single QTL. Thus, in total, 17 QTL for TE were identified.

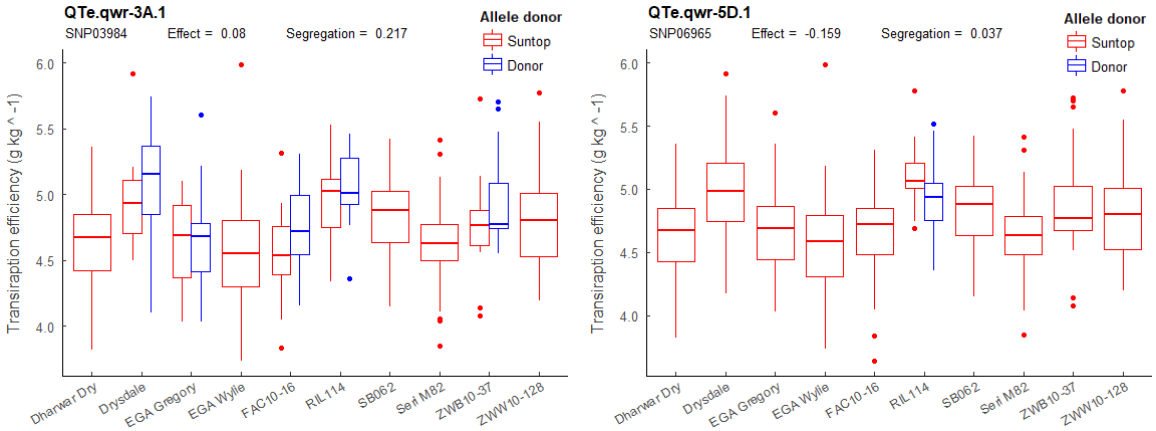

Figure 1. Boxplots of phenotypic variation in TE (BLUEs) for families with the reference (Suntop, red) or donor (blue) allele for two QTL for TE (*QTe.qwr-3A.1* and *QTe.qwr-5D.1*; see Table 1). QTL name, marker name, effect size (g kg^{-1}) and segregation (proportion NAM lines with the donor allele) within the Suntop population are presented above each graph.

Seven of the identified markers for TE had effects larger than 0.1 g kg^{-1}, while the other 16 had smaller effects (Table 1). Effect size for each marker ranged from 0.108 g kg^{-1} to -0.159 g kg^{-1} (Table 1). The number of families segregating for these markers and the marker effect across these segregating families varied greatly. For instance, *QTe.qwr-3A.1* segregated in much more of the Suntop population than *QTe.qwr-5D.1* (21.7% and 3.7%, respectively; Figure 1). In addition, differences in TE between lines of a family that did or did not have the donor allele (blue vs red boxplots in Figure 1, left) differed greatly across families. For instance, *QTe.qwr-3A.1* had a substantial positive impact in the Drysdale family, whereas in other segregating families its impact was typically smaller, so that, on average, this marker 'only' had an effect of 0.08 g kg^{-1} in the NAM population. On the other hand, *QTe.qwr-5D.1* had a much greater absolute

effect size (-0.159 g kg^{-1}) resulting from variation in only one family (RIL114). Overall, the average effect size of a marker (Table 1) is affected by how the marker segregates and its impact in the segregating families.

QTL of main interest with potential to improve the TE of Suntop included *QTe.qwr-1D.1*, *QTe.qwr-3B.1:3*, *QTe.qwr-7A.1*, *QTe.qwr-3A.1* and *QTe.qwr-5A.1:2*, given their high positive effect on TE and high level of significance. Interestingly, three of these QTL (*QTe.qwr-3B1:3*, *QTe.qwr-7A.1* and *QTe.qwr-5A.1:2*) collocated with QTL found in different field studies for canopy temperature, i.e. colocation with QTL for another integrated trait related to heat and drought adaptation that could be a consequence of differences in TE (Pinto *et al.* 2010, Bennett *et al.* 2012, Rebetzke *et al.* 2013). In addition, *QTe.qwr-5A.1:2* also collocated with QTL from this study for traits related to stomatal control (transpiration per unit of leaf area), leaf greenness and biomass (data not shown).

Several of the identified QTL, such as *QTe.qwr-3B.4* and *QTe.qwr-5B1:3*, collocate with QTL for carbon isotope discrimination (CID), a trait strongly linked to TE (Rebetzke *et al.* (2008). In both studies, the allele from Drysdale, a cultivar bred for its high TE (e.g. Richards *et al.* 2010), had a large positive effect for these two QTL, i.e. *QTe.qwr-3B.4* and *QTe.qwr-5B1:3* (data not shown).

Table 1. Most significant markers identified for TE. 'QTL range' corresponds to genomic region surrounding the molecular marker (identified by MLMM). QTL in bold represent a composite region including multiple markers (in italics) which are close and have overlapping ranges. For each QTL, the corresponding identified molecular marker(s), chromosome, position on the chromosome, effect, significance level, % of genetic variation explained, and the range of the QTL are given. Suntop NAM Seg. is the proportion of lines that have the donor allele. *P*-values: * = 0.05, ** = 0.01, * = 0.001**

QTL	Marker	Chr.	Position (cM)	Effect g kg^{-1}	*P*-value	Sig.	% Var.	QTL range	Suntop NAM Seg.
QTe.qwr-1A.1	SNP00444	1A	174.33	-0.135	1.52E-04	***	5.7%	174.33 – 191.82	0.125
QTe.qwr-1A.2	SNP14048	1A	200.33	0.067	1.71E-02	*	2.9%	200 – 206.59	0.208
QTe.qwr-1D.1	SNP01330	1D	15.21	0.108	4.22E-04	***	4.6%	15.21 – 25.42	0.084
QTe.qwr-3A.1	SNP03984	3A	147.59	0.080	5.36E-05	***	3.4%	143.73 – 151.45	0.217
QTe.qwr-3B.1:3		3B						14.27 – 26.79	
QTe.qwr-3B.1	SNP04683	3B	14.27	0.066	1.67E-03	**	2.8%	14.27 – 26.79	0.178
QTe.qwr-3B.2	SNP04068	3B	17.36	0.042	3.58E-03	**	1.8%	14.27 – 26.79	0.479
QTe.qwr-3B.3	SNP04716	3B	24.98	0.103	1.85E-04	***	4.4%	14.27 – 26.79	0.109
QTe.qwr-3B.4	SNP04371	3B	65.49	0.055	9.52E-03	**	2.3%	60.19 – 67.1	0.174
QTe.qwr-3D.1	SNP04870	3D	81.76	0.057	1.87E-03	**	2.4%	81 – 82	0.261
QTe.qwr-5A.1:2		5A						60.22 – 91.64	
QTe.qwr-5A.1	SNP06040	5A	71.17	0.064	6.64E-04	***	2.7%	60.22 – 85.51	0.222
QTe.qwr-5A.2	SNP06089	5A	85.51	-0.151	1.15E-16	***	6.4%	71.17 – 91.64	0.449
QTe.qwr-5B.1:3		5B						6.01 – 43.08	
QTe.qwr-5B.1	SNP06279	5B	24.88	-0.065	4.79E-03	**	2.8%	6.01 – 43.08	0.134
QTe.qwr-5B.2	SNP06325	5B	29.72	0.067	2.30E-03	**	2.9%	24.88 – 43.08	0.17
QTe.qwr-5B.3	SNP06434	5B	43.08	0.050	9.57E-03	**	2.1%	24.88 – 43.08	0.225
QTe.qwr-5B.4	SNP06787	5B	123.85	-0.157	2.49E-04	***	6.7%	118.28 – 132	0.034
QTe.qwr-5D.1:2		5D						130.22 – 152.06	
QTe.qwr-5D.1	SNP06965	5D	141.90	-0.159	3.19E-04	***	6.8%	130.22 – 145.06	0.037
QTe.qwr-5D.2	SNP07003	5D	152.06	0.097	2.91E-02	*	4.1%	134.1 – 152.06	0.034
QTe.qwr-6A.1	SNP07134	6A	27.53	-0.051	1.78E-03	**	2.2%	25.6 – 43.27	0.384
QTe.qwr-6B.1	SNP16080	6B	66.36	-0.088	2.08E-05	***	3.7%	55.93 – 67.6	0.198
QTe.qwr-6D.1	SNP08293	6D	96.42	-0.114	2.52E-04	***	4.9%	91.05 – 107.98	0.069
QTe.qwr-7A.1	SNP08441	7A	19.77	0.093	3.16E-04	***	4.0%	17.51 – 23.83	0.121
QTe.qwr-7B.1	SNP09332	7B	57.67	-0.045	8.96E-03	**	1.9%	52.5 – 59.54	0.342
QTe.qwr-7D.1	SNP09690	7D	43.57	0.051	1.63E-04	***	2.2%	43.57 – 57.86	0.486

Conclusion

By combining a high throughput methodology (Chenu *et al.* 2018, Fletcher *et al.* 2018) with the MLMM GWAS technique (Segura *et al.* 2012), 17 QTL were detected for TE in a genetically structured NAM population of wheat. Several of these QTL collocate with QTL reported in the literature for drought related traits, such as CID and canopy temperature. The method also enabled the detection of new QTL with strong effects on TE. Alleles of interest from QTL identified in this study could be stacked together to produce new wheat germplasm with improved TE and drought tolerance.

References

Ababaei, B. and Chenu, K. 2019. Impact of genotypic variations in transpiration rate on Australian wheat productivity. *Australian Agronomy Conference.* Wagga Wagga, Australia.

Araus, J. L. and Cairns, J. E. 2014. Field high-throughput phenotyping: the new crop breeding frontier. *Trends in Plant Science,* 19**,** 52-61.

Bennett, D., Reynolds, M., Mullan, D., Izanloo, A., Kuchel, H., Langridge, P. and Schnurbusch, T. 2012. Detection of two major grain yield QTL in bread wheat (Triticum aestivum L.) under heat, drought and high yield potential environments. *Theor Appl Genet,* 125.

Butler, D., Cullis, B. R., Gilmour, R. R. and Gobel, B. J. 2009. ASReml-R reference manual. Technical Report 3, Queensland Department of Agriculture, Fisheries and Forestry.

Chenu, K., Chapman, S. C., Tardieu, F., Mclean, G., Welcker, C. and Hammer, G. L. 2009. Simulating the yield impacts of organ-level quantitative trait loci associated with drought response in maize: a "gene-to-phenotype" modeling approach. *Genetics,* 183**,** 1507-23.

Chenu, K., Van Oosterom, E. J., Mclean, G., Deifel, K. S., Fletcher, A., Geetika, G., Tirfessa, A., Mace, E. S., Jordan, D. R., Sulman, R. and Hammer, G. L. 2018. Integrating modelling and phenotyping approaches to identify and screen complex traits: transpiration efficiency in cereals. *Journal of Experimental Botany,* 69**,** 3181-3194.

Condon, A. G., Richards, R. A., Rebetzke, G. J. and Farquhar, G. D. 2002. Improving intrinsic water-use efficiency and crop yield. (Crop Physiology & Metabolism).(Statistical Data Included). *Crop Science,* 42**,** 122.

Fao 2018. FAOSTAT statistics database. FAO.

Fischer, R. A., Byerlee, D. and Edmeades, G. O. 2014. *Crop yields and global food security: will yield increase continue to feed the world?,* Canberra, Australian Centre for International Agricultural Research.

Fletcher, A. and Chenu, K. 2015. Change in biomass partitioning and transpiration efficiency in Australian wheat varieties over the last decades. *17th Australian Agronomy Conference.* Hobart, Australian.

Fletcher, A., Christopher, J., Hunter, M., Rebetzke, G. and Chenu, K. 2018. A low-cost method to rapidly and accurately screen for transpiration efficiency in wheat. *Plant Methods,* 14**,** 77.

Ipcc 2014. *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change,* Geneva, Switzerland, Intergovernmental Panel on Climate Change (IPCC).

Lobell, D. B., Hammer, G. L., Chenu, K., Zheng, B., Mclean, G. and Chapman, S. C. 2015. The shifting influence of drought and heat stress for crops in northeast Australia. *Glob Chang Biol,* 21**,** 4115-27.

Pinto, R. S., Reynolds, M. P., Mathews, K. L., Mcintyre, C. L., Olivares-Villegas, J. J. and Chapman, S. C. 2010. Heat and drought adaptive QTL in a wheat population designed to minimize confounding agronomic effects. *Theoretical and Applied Genetics,* 121**,** 1001-1021.

R Core Team 2018. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.

Rebetzke, G. J., Condon, A. G., Farquhar, G. D., Appels, R. and Richards, R. A. 2008. Quantitative trait loci for carbon isotope discrimination are repeatable across environments and wheat mapping populations. *Theor Appl Genet,* 118**,** 123-37.

Rebetzke, G. J., Rattey, A. R., Farquhar, G. D., Richards, R. A. and Condon, A. G. 2013. Genomic regions for canopy temperature and their genetic association with stomatal conductance and grain yield in wheat. *Functional Plant Biology,* 40**,** 14-33.

Richards, R. A., Rebetzke, G. J., Watt, M., Condon, A. G., Spielmeyer, W. and Dolferus, R. 2010. Breeding for improved water productivity in temperate cereals: phenotyping, quantitative trait loci, markers and the selection environment. *Functional Plant Biology,* 37**,** 85-97.

Segura, V., Vilhjálmsson, B. J., Platt, A., Korte, A., Seren, Ü., Long, Q. and Nordborg, M. 2012. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature Genetics,* 44**,** 825.

Wang, B., Liu, D. L., O'leary, G. J., Asseng, S., Macadam, I., Lines-Kelly, R., Yang, X., Clark, A., Crean, J., Sides, T., Xing, H., Mi, C. and Yu, Q. 2018. Australian wheat production expected to decrease by the late 21st century. 24**,** 2403-2415.

Watson, J., Zheng, B., Chapman, S. and Chenu, K. 2017. Projected impact of future climate on water-stress patterns across the Australian wheatbelt. *Journal of Experimental Botany***,** 368-368.

Zadoks, J. C., Chang, T. T. and Konzak, C. F. 1974. A decimal code for the growth stages of cereals. *Weed Res,* 14.