CSIRO PUBLISHING

Functional Plant Biology, 2013, **40**, 1–13
http://dx.doi.org/10.1071/FP12180

A multisite managed environment facility for targeted trait and germplasm phenotyping

Greg J. Rebetzke[A,G], *Karine Chenu*[B], *Ben Biddulph*[C], *Carina Moeller*[D], *Dave M. Deery*[A], *Allan R. Rattey*[A], *Dion Bennett*[E], *Ed G. Barrett-Lennard*[C] *and Jorge E. Mayer*[F]

[A]CSIRO Plant Industry, PO Box 1600, Canberra, ACT 2601, Australia.
[B]Queensland Alliance for Agriculture and Food Innovation, PO Box 102, Toowoomba, Qld 4350, Australia.
[C]Department of Agriculture and Food, Western Australia Locked Bag 4 Bentley Delivery Centre, Perth, WA 6983, Australia.
[D]University of Tasmania, Tasmanian Institute of Agriculture, Private Bag 98, Hobart ,Tas. 7001, Australia.
[E]Australian Grains Technology, Perkins Building, Roseworthy Campus, Roseworthy, SA 5371, Australia.
[F]Grains Research and Development Corporation, 40 Blackall Street, Barton, ACT 2600, Australia.
[G]Corresponding author. Email: greg.rebetzke@csiro.au

Abstract. Field evaluation of germplasm for performance under water and heat stress is challenging. Field environments are variable and unpredictable, and genotype × environment interactions are difficult to interpret if environments are not well characterised. Numerous traits, genes and quantitative trait loci have been proposed for improving performance but few have been used in variety development. This reflects the limited capacity of commercial breeding companies to screen for these traits and the absence of validation in field environments relevant to breeding companies, and because little is known about the economic benefit of selecting one particular trait over another. The value of the proposed traits or genes is commonly not demonstrated in genetic backgrounds of value to breeding companies. To overcome this disconnection between physiological trait breeding and uptake by breeding companies, three field sites representing the main environment types encountered across the Australian wheatbelt were selected to form a set of managed environment facilities (MEFs). Each MEF manages soil moisture stress through irrigation, and the effects of heat stress through variable sowing dates. Field trials are monitored continuously for weather variables and changes in soil water and canopy temperature in selected probe genotypes, which aids in decisions guiding irrigation scheduling and sampling times. Protocols have been standardised for an essential core set of measurements so that phenotyping yield and other traits are consistent across sites and seasons. MEFs enable assessment of a large number of traits across multiple genetic backgrounds in relevant environments, determine relative trait value, and facilitate delivery of promising germplasm and high value traits into commercial breeding programs.

Additional keywords: commercial breeding, field experiments, heat stress, water stress, wheat.

Received 22 June 2012, accepted 18 September 2012, published online 23 November 2012

Introduction

Improving crop productivity is a priority to ensure global food security. Drought and high temperature are major limitations to yield in rainfed environments. Both can also affect grain quality while lowering returns through losses from costs associated with crop management (Cattivelli *et al*. 2008). Breeding for improved average performance is contributing to higher-yielding varieties under drought but at slow rates of progress: studies in some regions show wheat (*Triticum aestivum* L.) productivity under rainfed conditions have improved little in the past 30 years (e.g. Graybosch and Peterson 2010). One reason might be that breeders typically focus on selection for disease resistance, quality and then yield, with few available reliable methods for selecting tolerance to abiotic stresses, particularly drought. Indeed, because yield is genetically complex and subject to large genotype × environment interactions, selection for combinations of favourable alleles for performance under water limitation makes yield a difficult target (Cooper *et al*. 1997).

Several traits and genes have been reported with potential for improving yield under drought (e.g. Cattivelli *et al*. 2008). A more recent focus aims to identify traits contributing to improved water use efficiency (WUE) (i.e. biomass produced per unit water used), a major contributor to yield for crops experiencing water limitation (Richards *et al*. 2010). Commercial breeders have a strong belief in many of these traits and actively request elite germplasm carrying these traits for efficient water use (e.g. high early vigour, alternative dwarfing genes, reduced tillering, high stem carbohydrate content, greater leaf-level transpiration efficiency) that would perform better in rainfed environments and reduce the water requirement in irrigated

fields. Yet despite this interest, few traits have been adopted in the development of a commercial variety. This is partly because of the limited capacity of commercial breeding companies to screen for these traits, partly because of the absence of validation in field environments and genotypic backgrounds that are relevant to breeding companies, and partly because little is known about the economic benefit of selection for one trait over another.

Trait and gene adoption, and the critical need for validation

The link between basic plant science (genetics and plant physiology) and delivery to growers through breeding of new crop varieties can be schematised as in Fig. 1. Plant scientists span the continuum but most operate only in one of the first two boxes. There is a large gap between research that is mostly developed in the laboratory and glasshouse, and the validation and delivery to growers via the commercial breeding companies (Passioura 2012). Identification of a candidate gene or quantitative trait loci (QTL) underpinning a trait in a laboratory or glasshouse study does not necessarily translate to trait expression or value in the field, and so fails when passed to breeders wishing to select for improved performance in their breeding programs. The translation from a cell- or leaf-based phenotype to performance when scaled up to a canopy level is commonly very poor (Passioura 2010). For example, genotype × environment interactions can contribute to changes in genotype ranking for traits such as specific leaf area and stem carbohydrates when assessed in controlled environments and the field (van Herwaarden and Richards 1999; Rebetzke *et al*. 2004).

Commercial breeders are financially constrained in their capacity to take up the efforts of public research. Without data demonstrating the value of the research in a commercial context, they are unlikely to replace populations developed specifically in their own breeding programs with populations varying for untested traits or genes developed elsewhere (Chris Moore, InterGrain breeder, pers. comm.). There is a significant cost in population development and selection, with durations of up to 12 years from the initial cross to the release of a commercial wheat variety. The entire duration from trait inception to the release of the first high transpiration efficiency and low carbon isotope discrimination wheat varieties, Drysdale and Rees, was over 20 years (Rebetzke *et al*. 2002). Hence commercial breeders require very convincing evidence that elements of their breeding programs should be scaled back in order to resource the capture and use of specific traits and genes.

The substantial cost in population development and selection commonly necessitates a trait validation step in collaboration with breeders before acceptance and use in breeding programs (Fig. 1). Furthermore, given the many hundreds of new traits and genes being recommended in the literature for use, the trait validation must be coupled with the capacity for valuing the relative merit of individual traits. Parameters under consideration here include ease in selection (cost and genetic complexity), robustness of expression across multiple genetic backgrounds and benefit for improving yield in targeted (specific) or all (broad) potential environments. Although modelling tools can help in this process (e.g. Chenu *et al*. 2009), there is a need for a coordinated delivery mechanism that allows the assessment and valuing of traits for use in breeding for specific target environments (e.g. the difficult nature of screening under drought). Being able to mimic specific target environments without the vagaries of natural rainfall should enable prebreeders and breeders to assign accurate values to the individual traits being analysed, and thereby better coordinate pre-breeding and commercial breeding activities to deliver research more effectively (Fig. 1).

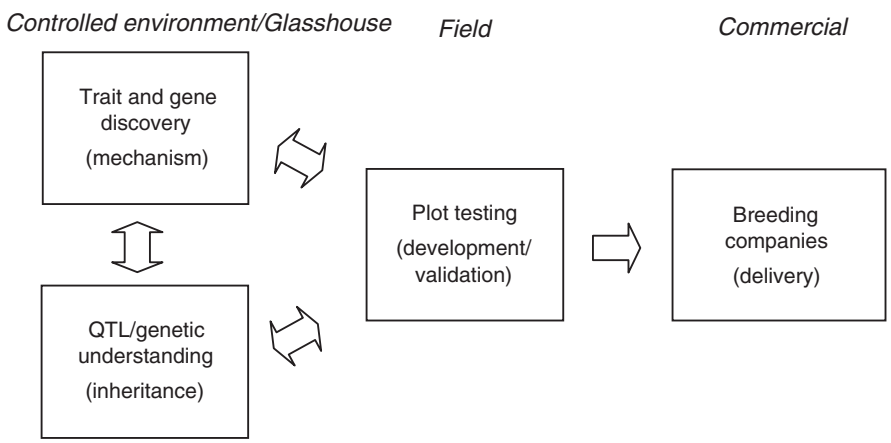

Fig. 1. Schematic showing the activities and corresponding scale from trait and gene discovery in prebreeding to uptake and deployment in a commercial breeding program. The arrows indicate the transfer or feedback of knowledge gained at one level to modify or invigorate the research activity at a different level. For example, results of plot testing in the field might refine phenotyping in controlled environment or lead to further research into gene discovery. Prebreeding is defined as all activities designed to identify desirable characteristics or genes from unadapted materials that cannot be used directly in breeding populations, and to transfer these traits to an intermediate set of materials that breeders can use further in producing new varieties for farmers. It is a necessary first step in the use of diversity arising from other materials (see Passioura 2012). The worth of the new breeding populations needs to be well demonstrated in the field if they are to attract the attention of commercial breeders.

A multisite field-based system for robust trait assessment

Reliable screening for performance under drought has been particularly challenging because of the inherent unpredictability of weather. Climate change is likely to increase weather variability to further challenge breeders in the identification of meaningful experiments for advanced line selection and breeding for performance in rainfed environments. Use of managed environments is becoming increasingly more important in both public and private organisations, permitting selection under controlled stress (e.g. Campos *et al*. 2004; Kirigwi *et al*. 2004; Trethowan *et al*. 2005; Bänziger *et al*. 2006). Yet despite their potential in screening traits and germplasm, there is little understanding of how best to efficiently implement such managed environments where the numbers of lines are large (e.g. in the order of 5000–10 000 lines, as occurs with many research agencies) and there is a critical need for accuracy in assessing sometimes small but important physiological and genetic effects. Large rainout shelters have the capacity to controlling water availability and thereby reduce the impact of year-to-year changes in the amount and timing of water availability. Such a system has been established using drip irrigation for large-scale screening of rice (*Oryza sativa* L.) in China (Liu *et al*. 2006). However, these and other rainout facilities are generally limited in their size and assessment capacity to bordered plots, reducing their potential to evaluate germplasm, traits and genes. Their smaller size also limits capacity to screen elite breeding lines, restricting their opportunities for prebreeders and breeders to assess traits meaningfully together under the same controlled conditions.

This paper reports on the development of three nationally coordinated management environment facilities (MEFs) targeting assessment of wheat germplasm and traits (and the genes contained therein) for improved performance under water-limited and high-temperature conditions. The development of these MEFs focuses on Australian conditions, but their understanding and implementation is presented in a way that can be extended to any country or major growing region. The paper focuses on a strong understanding of field phenotyping for performance under stress through considerations of: (1) identifying MEF locations representative of environment types in the cropping zone, (2) remote monitoring of the climate at each MEF site, (3) controlling field-based spatial variability, (4) selecting suitable germplasm for trait assessment, (5) assessing traits contributing to water-limited and high-temperature adaptation, and (6) phenotyping reliably with standardised measurement protocols.

Identifying MEF locations representative of the cropping zone

Selection for improved performance under drought is challenging owing to changes in line ranking (i.e. genotype × environment interaction) across environments (Cooper *et al*. 1997). These interactions reflect genotype responses to environment-based changes, particularly changes in temperature and the amount, timing and frequency of rainfall relative to the crop's growth cycle (Fig. 2). Characterising environments over the longer term identifies general patterns of water availability and leads to the identification of sites appropriate for establishment of an MEF. In turn, development of an MEF with known severity, timing and frequency of water limitation reflective of the type of drought experienced in the target population of environments increases confidence in repeatable screening of germplasm (Chenu *et al*. 2011).

Rainfall and temperature variability across target cropping zones

Climate records from 1889–2011 (Bureau of Meteorology 2011; Jeffrey *et al*. 2001) were used to analyse monthly temperature and rainfall, and to simulate water stress patterns and wheat yields for three experimental sites (defined by combinations of location and soil type) across the Australian wheatbelt. In this cropping zone, wheat is produced in diverse climatic regions with rainfall being winter-dominant in the south-western parts of the continent, evenly distributed in the south-east and summer dominant in north-eastern regions of the wheatbelt (Fig. 2; details for the three MEF sites are given in Appendix 2). Variability in rainfall and temperature between seasons is high and crops are cultivated on contrasting soil types (Appendices 1 and 2, Supplementary Table 1, available as Supplementary Material to this paper; Potgieter *et al*. 2002). As a result, wheat crops experience a broad range of water deficits and temperature events (great variability in the timing and intensity of stress events) spatially and across seasons (e.g. Chenu *et al*. 2011; Zheng *et al*. 2012).

Characterising drought patterns across the wheatbelt

Using a crop simulation modelling approach, drought patterns were simulated over long-term periods (123 years) for the three MEF locations representative of the wheatbelt using the Agricultural Production Systems SIMulator (APSIM) crop model (Keating *et al*. 2003; Fig. 2; Appendix 1). In this approach, seasonal water deficit patterns are described for combinations of location × soil type × year × sowing date × amounts of soil water at sowing for the commercial variety 'Hartog' grown with optimal nutrient supply (Appendix 1, Supplementary Table S1). The water deficit patterns are described by a water supply to crop demand ratio (the ratio of water available to the crop to the water that the crop could use if enough water were provided) simulated over the cropping season and is centred at zero at flowering. The patterns of all cropping seasons are clustered into distinct response patterns corresponding to four main drought environment types (ETs) experienced at each site (Fig. 2). Further information on the approach and applications are given in Appendix 1 and Chenu *et al*. (2011).

The environment types of the different sites presented some similarities: overall, the first environment types (ET1) comprised simulations with no or only short-term water deficits. The second environment types (ET2) represented mild stresses, which occurred mainly after flowering and eased during grain filling at some locations. The third environment types (ET3) were characterised by intermediate stress, beginning during the vegetative period, which was relieved during the grain filling period. Finally, the fourth environment types (ET4) corresponded to severe stresses beginning earlier than in all other environments types and lasting often up to maturity. These general patterns

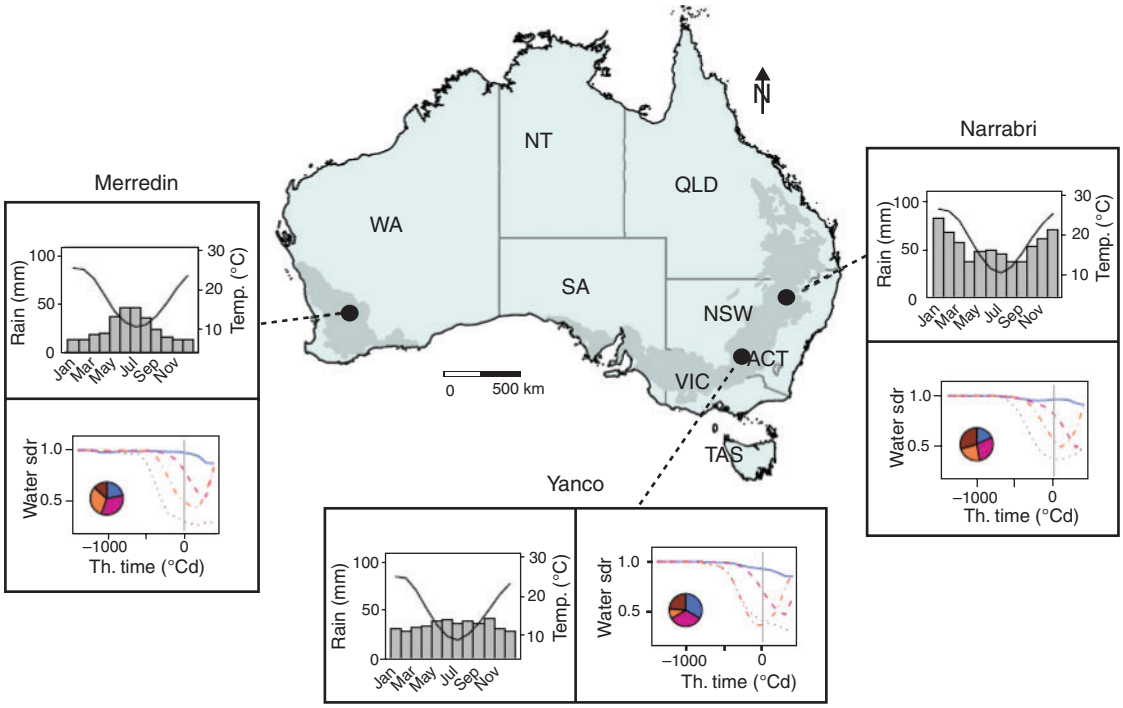

Fig. 2. Climate characteristics of the managed environment facility sites of Merredin, Narrabri and Yanco in the Australian wheatbelt (shaded). Mean monthly rainfall and temperature are given for data collected for 1889–2011 (Jeffrey *et al*. 2001). The sowing window for wheat is commonly from late April to mid-June. Water stress patterns for four environment types (ETs) have been modelled for the three MEF locations ET1 (blue, plain line), ET2 (purple, dash), ET3 (orange, dash-dot), and ET4 (brown, dot) (see text for explanation). The stress index (shown here as the supply : demand ratio, sdr) is represented as a function of thermal time relative to flowering and ranges from 1 (no water stress) to 0 (full stress, no water available to the crop). The inserted pie chart contains the frequency of the four environment types at each location.

differed slightly between Narrabri, Merredin and Yanco (Fig. 2), and their frequency of occurrence varied greatly across locations (Fig. 2, insets) and seasons (data not shown). Crops at all MEFs often experienced drought around flowering, and substantial frequency of severe terminal drought also occurred at all sites, in particular Narrabri and Yanco (Fig. 2). The simulation analysis of water-stress patterns showed that the drought environment types at the experimental sites Narrabri, Merredin and Yanco are representative of those experienced elsewhere in the wheatbelt (unpubl. data). The severity with which the ETs reduced the water-limited yield potential of the simulated wheat crop increased on average in the order ET1 > ET2 > ET3 > ET4 (Fig. 3). The relative effect of drought severity on grain yield was reasonably consistent across all MEFs, with average yields being greatest for Narrabri and smallest for Merredin (Fig. 3).

Managing irrigation requirements at the MEF

Two significant challenges with any drought imposition experiment are the identification and implementation of the appropriate level of water stress under which performance can be monitored. Passioura (2007) highlighted the danger of the commonly used extreme water stress imposition in pots and the potential relevance to plant and tissue survival but not necessarily productivity. This has been demonstrated empirically in field studies where line ranking is zero and sometimes negative for performance in droughted and irrigated environments (e.g. Ceccarelli 1994; Cooper *et al*. 1997). This result is not uncommon in the literature and highlights that performance under extremes of water stress are likely to compromise performance in average or more favourable environments where growers make their greatest income.

Selection for improved productivity must be undertaken in appropriate target environments. In rainfed environments, these are commonly represented by seasons with median rainfall and representative soil types. With this view, each MEF site maintains two irrigation treatments under which performance is assessed. The first is the development of a median-type season where rainfall and soil water (using neutron moisture meters in representative genotypes) are monitored, and supplementary irrigation is supplied to achieve a season with a drought pattern close to the median of the target environment (Fig. 2). Initial soil water conditions typical of those anticipated in the region (e.g. little stored water after dry Mediterranean summers) are targeted through monitoring of crop rotation and irrigation at sowing. The second irrigation regime is aimed at assessing genotype and trait performance when water is less limited. This treatment is to ensure there is no loss in productivity with traits or germplasm in more favourable seasons. Here, additional irrigation is supplied to target a Decile 8 rainfall (wettest 20% of years).

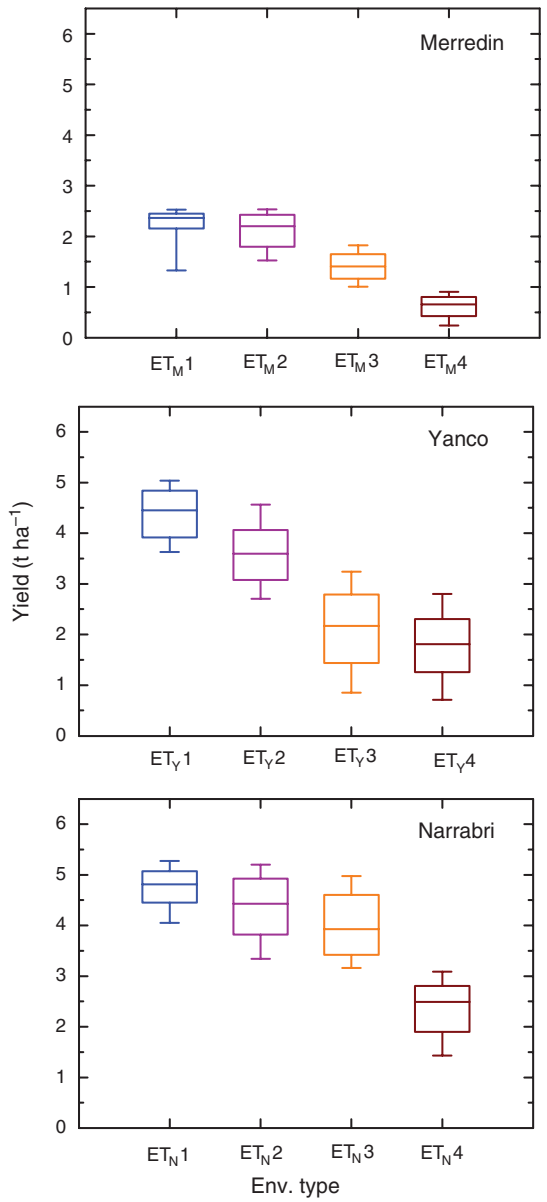

Fig. 3. Box plots of simulated yield for four water stress environment types (ET) at three managed environment facility (MEF) locations (Merredin, Yanco and Narrabri). The line in each box is the median and the whiskers define the 5% and 95% percentiles. Yields are simulated from 123 years of historical climate data.

Despite the knowledge of representative target rainfall scenarios, managing irrigation requirements is challenging, given the uncertainty of in-season rainfall. Managing representative drought scenarios in some seasons may commence with little starting soil moisture (dry summer) whereas others may commence with a full soil profile (wet summers), and the timing of rainfall throughout the season is highly variable (e.g. Fig. 2). Crop management software such as APSIM (Keating *et al*. 2003) permits assessment in near real-time of the timing and intensity of different irrigation events based on recorded rainfall, soil water data, current management and plausible future climatic scenarios for the remainder of the season. Decisions as to the need for and amount of irrigation may then be determined accurately for a growing crop and its predicted water use, and probabilities can then be derived as to the likelihood of rainfall given historic climate data for the MEF.

Remote climate monitoring at each MEF

To monitor genotype performance and environmental variables at each MEF site, wireless (distributed) sensor networks remotely monitor a core set of 10 representative wheat genotypes for changes in soil water use and transpiration. The local weather conditions are continuously monitored and the data are transmitted wirelessly in near real-time via the mobile phone network for viewing on the internet.

Phenonet wireless sensor network

The 'Phenonet' is a wireless sensor network and data visualisation analysis tool for field-based crop phenotyping. The Phenonet comprises three subsystems: (1) a weather station recording variables such as incoming solar radiation and photosynthetically active radiation (PAR), relative humidity, atmospheric pressure, air temperature, rainfall and wind speed; (2) a soil moisture status monitoring system; and (3) crop canopy (leaf) temperature sensors. The measurements are made at 10-s intervals and a 15-min average is transmitted to a web service via a third generation (3G) modem every hour. The sensors are radio-linked to a base station for data upload to a server at a remote location via the mobile phone network. The technology is mature, reliable, cost-effective and simple to deploy. The system has undergone a considered design process and does not interfere with other measurements and agronomic practices.

The weather station comprises: (1) a Vaisala WXT520 weather transmitter (http://www.vaisala.com/en/products/multiweathersensors/Pages/WXT520.aspx), which measures wind speed and direction, rainfall (duration and intensity), air temperature, relative humidity and barometric pressure; and (2) two radiation sensors to measure shortwave radiation (shortwave silicon cell pyranometer (Model SP-110 http://www.apogeeinstruments.com/pyranometer/) and PAR (Model SQ-100 http://www.apogeeinstruments.com/quantum/index1.html). The weather station sensors are located at a standard nominal height of ~2 m above the ground to ensure that measurements of the wind speed and other climatic parameters are reliable. Data are then transmitted electronically via very high frequency (VHF) radio to the 3G modem.

The soil moisture status is measured with gypsum block soil moisture sensors MEA (GBHeavy) at depths of 0.1–1.2 m (Fig. S1). The output of the GBHeavy sensors is logged and transmitted via VHF radio to a 3G modem that uploads the data to a server every 2 h. The logger has eight channels and is typically interfaced with six gypsum block sensors and two thermistors for soil temperature measurement at depths of 0.1–1.2 m. The crop canopy temperature is measured with a noncontact infrared thermometer similar in design to that described by O'Shaughnessy *et al*. (2011) (Fig. S2*a*). Measurements are

transmitted via ultra high frequency (UHF) radio to the 3G modem.

A portable vehicle has also been designed with the aim of monitoring multiple plant and canopy traits across multiple field plots in the MEF (Fig. S3). The purpose-built crop monitoring buggy has been fitted with four red–green–blue cameras for measurement of ground cover and plant establishment, light detection and ranging (LiDAR) sensors to measure plant height and biovolume, a spectral radiometer from 300 nm to 2500 nm to measure normalized difference vegetation index (NDVI) and various spectral vegetation indices, and three infrared temperature sensors for crop canopy temperature. A 6-m long field plot can be surveyed in less than 10 s, enabling a typically 400-plot experiment containing several genotypes to be sampled for multiple traits in less than an hour.

Data visualisation and analysis web interface

Phenonet allows environmental and crop data to be gathered at a far higher resolution than conventional methods. The platform enables the comparison of genotype performance and assists in identifying the optimal sampling time for target traits. An online data analysis and visualisation platform is under development (Fig. S2*b*), which can assist crop management decisions (e.g. timing of irrigation) and has the capacity for extension to real-time modelling.

The data platform links closely with the capture of other phenotypic and genotypic data from all three sites into a web-based database permitting ready access to all prebreeding and breeding users of the MEF.

Controlling spatial variability

Ideally, managed environments should be spatially uniform and have perfectly characterised environment conditions that allow clear and repeatable differentiation of genotypes as well as the assessment of mechanisms, traits and putative QTL or candidate genes associated with performance under drought. However, the reality is that such sites rarely exist. Site uniformity and control of water availability should aid in reducing the error variance and reduce the potential for genotype × season interactions to increase repeatability and genetic variance.

A first step is the use of appropriate experimental designs (e.g. partial replication; Smith *et al*. 2006) and statistical tools for analysing complex data such as a mixed linear statistical methodology (Smith *et al*. 2001). Second, the site should be managed appropriately to reduce the potential for uneven conditions (e.g. soil constraints that affect crop growth). For instance, biotic constraints such as soilborne diseases commonly limit root growth to reduce water and nutrient uptake. The use of a fallow or non-host 'break' crops to reduce pathogen growth and soil inoculum levels are important considerations in the management of rotations in which the trial sites are grown (e.g. Kirkegaard *et al*. 2004). Slower growth, and limited water and nutrient uptake, may also result from several abiotic constraints including high and low soil pH, micronutrient deficiencies and toxicities, soil waterlogging and high ion concentrations that reduce the potential of the soil solution.

Extensive soil testing of a field can help identify locations to be used for managing water availability, but there is a real question as to the intensity of measurement that is required to characterise the site. An inadequate sampling program may not produce data of sufficient intensity to interpret site effects on yield and how to interpret the crop data; on the other hand, sampling that is too intensive wastes resources. A simple method for determining the scale of variation in a site that may appear quite uniform at the soil surface is through electromagnetic induction (Bennett *et al*. 2009). The EM38 meter (Geonics Ltd, Ontario, Canada) estimates the apparent electrical conductivity of the soil. Variation in the signal is affected by the soil's clay content, moisture and ionic strength. As an example, Fig. 4 shows EM38 readings for one of the fields at the Merredin MEF site. The EM38 measurements (the apparent electrical soil readings) at this site range between 56 and 161 mS m^{-1}. Understanding this range across the field enabled experiments to be blocked to minimise sampling across different conductivities, thereby increasing heritability and confidence in detecting treatment differences (e.g. Slavich *et al*. 1990). Similarly, an EM38 meter can be used in scouting and identifying suitable sites for use in a controlled field experiment.

Nature of germplasm for trait assessment

The germplasm used in assessing trait value may take numerous forms including:

(1) large, frequently diverse germplasm sets;
(2) near-isogenic lines (NILs) from a population, selected to be divergent for the trait of interest but phenotypically similar for other attributes (e.g. Mathews *et al*. 2007);
(3) large multiparent populations including nested association mapping and multi-parent advanced generation inter-cross containing doubled-haploid lines (DH; Rebetzke *et al*.

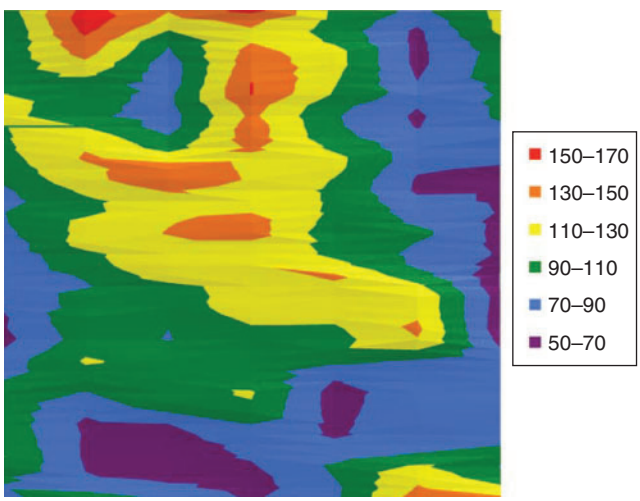

Fig. 4. Map of electrical conductivity readings (mSm^{-1}) at the MEF site of Merredin. Data measured immediately before wheat sowing in 2011, with a EM38 in the vertical orientation. The site was 56 m (*y*-axis) × 60 m (*x*-axis) in area. The EM38 readings were taken on a grid of 10 m (*y* axis) × 1.2 m (*x* axis).

2008*b*) and recombinant inbred lines (RILs; Rattey *et al*. 2009); and

(4) selected tails.

In all cases, control of confounding through differences in plant development, height, disease reaction and so forth is critical. Although all of the above options are valid, it is easy to understand that as the trait(s) of interest become more complex and thus controlled by multiple genes (polygenic traits), it is desirable to have as much as possible of the 'background genetics' constant, so that the 'foreground genetics' that are of greatest importance for the target trait(s) can be easily dissected. Furthermore, as trait(s) of economic importance are more frequently polygenic, it is unlikely that a single pair of NILs would contain the necessary contrast in the relevant genes underpinning the trait(s). For example, a trait controlled by 10 independent genes would require assessment of over 3000 backcross F_1 individuals to be 95% confident of identifying at least one F_1 individual heterozygous for all 10 alleles (Bonnett *et al*. 2005). Thus, in many instances, Options 1 and 2 have been replaced by Option 3, as in the research of Rebetzke *et al*. (2008*a*, *b*) and many others in detecting QTL for polygenic trait(s).

Population sizes required for investigating of polygenic traits are very large (>200 DHs or RILs are necessary) and therefore resource-intensive to phenotype confidently. If this is extended to multiple genetic backgrounds and multiple traits, and several hectares are soon required to compare and value confidently among traits. Thus, a novel approach for phenotyping is briefly described here: Option 4, namely 'selected tails'. Initially, lines from the tails of populations are grown under conditions that are favourable for maximising trait expression and heritability (see Richards *et al*. 2010). This enables the effective selection of high or low extremes of the population for a target trait that are phenotypically similar for other traits including phenology, height and disease reaction (Fig. 5). This approach is somewhat similar to a bulk segregant approach where selected bulks reflect the presence or absence of a target phenotype (Michelmore *et al*. 1991). These high and low tails are also selected to be phenotypically similar for other 'fundamental' nontarget attributes (e.g. phenology) and are analogous to a NIL pair described in Option 2, as both tails of the population should have a similar 'genetic background' when considered together but diverge for the chromosomal regions of interest. Multiple lines (5–10% of the population) may be included in each tail, compared with only a single line in a NIL pairing or hundreds of lines for a population. The representation of multiple lines in a tail minimises the likelihood of the sampling of background genetics having a significant impact on the confounding of trait value.

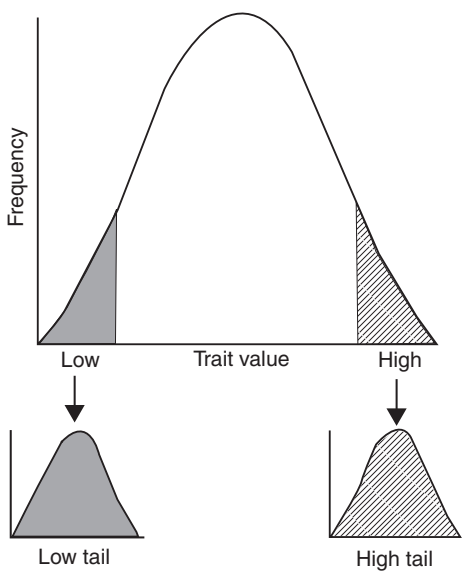

Fig. 5. Identification of high and low 'tails' from a biparental population for a target trait.

Subsequently, these selected tails are grown in the target environment types and for multiple populations. By growing only a selected set of lines from a population that are divergent for the trait of interest but, on average, otherwise genetically and phenotypically similar, resources can be more efficiently used to increase the number of environments or populations and traits under examination. This is crucial knowledge to enable deployment of the trait to growers' fields, as breeders must know if a target trait is of economic importance across all or only some genetic backgrounds, as well as in all or only some environments.

Owing to the large reduction in numbers of lines comprising each of the selected tails (5–10%), the tail approach (Option 4) is vastly more resource-efficient than Option 3. Also, assuming there is suitable genetic variation, the tail option enables the same genetic population to be utilised for multiple traits (through development of multiple tails). Populations are routinely phenotyped for multiple traits, allowing one to add value for specific traits from relevant populations for little additional cost. In comparison, development of NILs specific for a target trait allows one to identify trait-specific lines with little value beyond the target traits. This is very costly and becomes even more costly as the numbers of traits and genetic backgrounds to be assessed increases. Importantly, owing to gains in resource use efficiency, the selected tails approach allows numerous populations to be tested (e.g. across prebreeding groups) side-by-side with a reduced risk of spatial variability and other environmental factors that can adversely affect very large field trials impacting on findings. This considerably reduces the cost and increases precision in assessing traits, and adds considerable value to the development of trait-based populations.

Breeder germplasm and the need for 'benchmarking'

The phenotyping and assessment of germplasm representing multiple traits or genes is an essential component of the MEFs. This germplasm should be derived from parental sources that are commercially relevant but also include exotic germplasm established as high yielding in drought conditions elsewhere, including overseas. As before, the germplasm should have the appropriate phenology, height and other adaptations, including disease resistance required for the target environments. Breeders should be encouraged to nominate elite germplasm from their own programs to allow assessment of target traits within existing cultivars and elite breeding material, as well as potentially

comparing yield performance. This comparison would provide a benchmark to value target traits and genes with what is available within contemporary elite germplasm.

Contrasting performance and subsequent value among different traits would be considerably more useful if combined with trait assessment in all relevant commercial germplasm growing nearby. These 'probe' genotypes would permit understanding of whether the trait value was specific to experimental populations or if the extent of variation for the assessed traits in experimental populations was comparable to that present within the existing commercial gene pool. If trait value was high for adaptation but the extent of trait variation was small within breeding programs, it would then be simple to identify germplasm combining high trait value with improved adaptation for use by breeders.

The deployment of germplasm and its assessment is illustrated in Fig. S4 for a prebreeding (i.e. precommercial germplasm) and benchmarking–breeding line stream. By running the two streams in parallel, prebreeders can assess the value of traits for improving productivity in prebred, almost commercially ready germplasm with performance in existing and new commercial varieties. Further, the extent of variation for proposed traits (such as carbon isotope discrimination) can be compared with diversity for the same traits in the breeding populations. Fig. S4 also shows how traits or populations can be maintained or replaced as knowledge about a trait or gene is acquired. This knowledge would also include engagement with the breeders regarding the value of the trait and the means for deployment of parental germplasm and trait selection in their breeding programs. For example, the value of indirect selection for carbon isotope discrimination and improved transpiration efficiency is readily assessed through screening of surrogate traits such as canopy temperature in the tail sets. The first 2 years permit trait assessment in prebreeding and breeding germplasm: trait robustness across backgrounds and environments (sites, years, irrigation regimes and sowing dates), the extent of variation, and the relationship with biomass and grain yield. By the third season, the value of this trait should be well understood and the need for transfer of parental germplasm ascertained. Subsequent years would focus on the introduction of new germplasm and populations, and the assessment and validation of phenotyping techniques developed in the MEFs.

Assessing traits contributing to water-limited and high-temperature adaptation

Water limitation

Empirical selection for yield in dry environments has led to the development of well-adapted wheats, primarily through selection for earlier flowering and reduced height (Richards *et al*. 2010). With many of the major alleles for appropriate phenology and height now fixed in commercial breeding programs, there is reliance on new alleles underpinning traits for improved productivity under drought. Assessment of performance under drought is challenging, owing to the multifaceted nature of drought as reflected in variability in the timing and nature of water availability. In turn, it can be argued that performance in a water-limited environment can be more simply thought of through a requirement for greater water productivity. Increases in the biomass or yield per unit of water used should permit a broader consideration of multiple traits that account for the nature, timing and variability of water limitation in the target environment.

The challenge of genetic improvement of grain yield (GY) under water-limited conditions can be expressed in terms of three components, namely to breed crops that:

(1) transpire more of the limiting water supply (e.g. limited soil evaporation; better root system),
(2) exchange transpired water for CO_2 more effectively in producing biomass (TE), and
(3) convert more of the total biomass into grain (harvest index; HI).

Several characteristics have been identified that affect transpiration, WUE and HI, and thereby yield improvement under drought. Many of these traits are physiologically independent, allowing genetic effects to be accumulated through selection (e.g. early vigour to reduce soil evaporation and carbon isotope discrimination to increase TE; Condon *et al*. 2004). Suitable genetic variation has been identified in wheat for many of these traits, and breeding has been undertaken to

Table 1. Different traits affecting performance under drought targeted at each of the different managed environment facilities (MEFs). The population type and number of genetic backgrounds under assessment are included as well as references concerning the effects of each trait

T, water transpired; WUE, water-use efficiency; HI, harvest index; M, Mediterranean environments; NIL, near-isogenic lines

Traits	Effect	Population type	Number of genetic backgrounds	Regions	Reference
Awn presence	WUE and HI	NILs	4	All	Evans *et al*. (1972)
Canopy staygreen	HI	Tails	2	All	Sirault *et al*. (2004)
Canopy temperature	T and WUE	Tails	2	All	Rattey *et al*. (2011)
Carbon isotope discrimination	WUE	Tails	6	All	Rebetzke *et al*. (2002)
Early vigour	T and WUE	Tails	6	M	Botwright *et al*. (2002)
Grain fertility	HI	Tails	3	All	Dolferus *et al*. (2011)
Leaf glaucousness	HI and WUE	Tails	2	All	Richards *et al*. (1986)
Plant development	T and HI	NILs	2	All	Trevaskis *et al*. (2003)
Reduced tillering	T and HI	NILs	6	All	Duggan *et al*. (2005)
Root vigour	T	Tails	2	All	Lopes and Reynolds (2010)
Stem carbohydrates	HI	Tails	2	All	Rebetzke *et al*. (2008*a*)

introgress these into a range of commercial genetic backgrounds (Table 1). Validation across genetic backgrounds and growing regions via the MEF enables assessment in populations selected to be relevant nationally.

Trait value is assessed for lines within and across genetic backgrounds. Consideration towards trait use in breeding is then assessed according to breeding-based criteria: (1) variation for the traits themselves, and (2) their influence on biomass and grain yield through correlated genetic response. In our case, the additive genetic correlation (r_A) between two traits (X and Y), and their narrow-sense heritabilities h_x^2 and h_y^2 are calculated, and the correlated response of trait Y to selection on trait X ($\Delta_{GY.X}$) predicted by:

$$\Delta_{GY.X} = k\sigma_{py}h_y h_x r_A, \quad (1)$$

where k is the standardised selection differential, and σ_{py} is the phenotypic standard deviation for Trait Y (Falconer and Mackay 1996).

Heat stress

Elevated temperatures during the reproductive stages of the growth cycle can lead to premature leaf senescence, reduced photosynthetic rate, reduced seed set, reduced duration of grain fill, reduced grain size and ultimately reduced grain yield. This also results in poorer WUE, with less grain produced per unit of water. Many wheat-growing countries experience average maximum temperatures well above the optimum for reproductive growth, and in Australia, extreme heat events often occur during the grain forming and filling periods of September through to November. Conditions such as these are believed to substantially reduce grain yields up to half of what would be achieved in comparable but cooler environments (Midmore *et al*. 1984; Shpiler and Blum 1986; Gibson and Paulsen 1999). Expected increases in average and extreme temperature in coming decades are a further incentive to improve wheat tolerance to elevated temperature (e.g. Intergovernmental Panel on Climate Change 2007; Zheng *et al*. 2012).

Previously, field characterisation of germplasm for heat stress tolerance has been under conditions where other factors, such as water availability, may be inconsistent, contributing error to the phenotypic assessment. Assessment of germplasm in well organised MEFs can make significant progress towards overcoming these issues. By sowing experiments later than is conventional, the chance of experiencing heat stress during reproductive growth is increased. Meanwhile, the ability to irrigate allows water availability patterns to be matched to the conventionally sown benchmarking experiments. This ensures that growing season water availability patterns are similar between the two experiments, enabling performance of germplasm in the hotter late sowing to be more comparable to the cooler conventional sowing. This is particularly important in environments with winter-dominant rainfalls, when rainfall becomes more infrequent through spring and into summer, where water availability may confound identification of the true heat tolerance phenotype.

A further strength of such facilities is the availability of a common pool of specialist equipment, with technicians trained in their use. Equipment such as SPAD meters to assess changes in leaf chlorophyll content before and after heat events (Ristic *et al*. 2007), and multiple fixed and handheld infrared temperature probes to identify lines that maintain cooler canopies under warmer temperatures permit screening of germplasm during the growing season. End-of-season measurement of yield components such as grain size and spike fertility, which are known to be sensitive to heat stress conditions in susceptible genotypes (Stone and Nicolas 1995; Gibson and Paulsen 1999), further aids in the identification of germplasm with greater levels of field-valid heat stress tolerance. This also improves our understanding of the heat tolerance phenotype and the traits to target, and the variation for these traits in experimental germplasm can be directly compared with varieties and advanced breeding germplasm.

Reliable phenotyping and the need for standardised protocols

The plant's physiological and agronomic measurements need to be reliable, robust and repeatable to elucidate the differential responses of genotypes to diverse stress environments. However, the sampling of the 'true' genotypic variation can be varied depending on the types of measurements, methods and practices employed by individual workers and different organisations across multiple experiments. This makes interpretation across sites and environments difficult. Hence, there is a strong need for standardisation to ensure that the phenotyping of yield and other traits is unconfounded by diverging experimental routines.

Table 2. Essential and desirable measurements for phenotyping of yield and other traits across multiple experiments in water-limited and high-temperature environments
DC, decimal code

	Timing[A] (frequency)
Essential (core)	
Plant establishment counts (plants m^{-2})	DC 12–13 (1×)
Ground cover (%)	DC 12–37 (3×)
Anthesis date	DC 45–70.2 (every 3 days)
Canopy temperature (°C)	DC 35–70 (2×)
Harvest index	DC 90
Spike number (spikes m^{-2})	DC 90
Plant height (cm)	DC 90
Thousand grain weight (g)	DC 90
Grain yield (g m^{-2})	DC 90
Observations and scores (e.g. incomplete plots, temperature damage, disease, lodging, and shattering)	As required
Desirable	
Canopy light interception (μmol m^{-2} s^{-1})	DC 35–60 (2×)
Normalised difference vegetation index (NDVI)	DC 12–60 (4×)
Carbon isotope discrimination	DC 30–32
Anthesis biomass (g m^{-2})	DC 60–65
Water soluble carbohydrates	DC 65–70 (1×)
Canopy temperature during grain filling (°C)	DC 70+ (4×)

[A]Timing according to the Zadoks decimal code for scoring stages of cereal development (Zadoks *et al*. 1974).

To address this issue, we standardised the protocols for assessing line performance under abiotic stresses for consistency across all experiments. As water and temperature stresses are the focus of the research, the field sites need to be free of physical, chemical and biological constraints affecting root growth. The plant nutritional status should be consistent with best local farming practice and crop stands need to be free of pests (weeds, insects and diseases). Representative measurements and samples are taken from the plot centre only (i.e. border rows are excluded) to minimise edge effects resulting from differences in growth conditions (briefly, the availability of resources light, water and nutrients) between the outer and central plot area (Rebetzke *et al*. 2012). Among the environmental variables that are routinely collected at each site are the daily weather (e.g. rainfall, temperature and solar radiation) and changes in soil moisture in a set of probe check lines. Soil characteristics (physical and chemical) are also being measured at each site.

A core set of measurements that we believe are critical to characterise genotype performance in water-limited environments are summarised in Table 2. These measurements are simple, reliable and repeatable, and allow screening of large numbers of lines for traits underpinning performance. The protocols of the measurements and their methodology constitute the standards used at each of the experimental sites. The protocols are freely available as a 41-page PDF titled 'Protocols for experimental plot sampling, handling and processing of cereal experiments' (Rebetzke *et al*. 2012) on PrometheusWiki (http://prometheuswiki.publish.csiro.au/tiki-index.php) using the search term 'field experiments'.

Concluding comments

The time from identification of a new trait or gene to its release in a commercial variety can take up to 20 years. Understanding the robustness of the trait value across different environments and genetic backgrounds is critical but very time-consuming. Convincing a breeder that one or a combination of traits is better than another requires a thorough knowledge of the trait based on robust phenotyping in relevant environments. Field validation in relevant environments will hasten uptake and delivery by breeding companies. Field-based facilities targeting the assessment of germplasm under rainfed and high temperature conditions, with the capacity to screen thousands of elite lines against commercial varieties, should provide information on trait value, and improve engagement and delivery to commercial breeding programs. We have detailed those considerations important to develop resource-efficient, nationally coordinated MEFs. Although the focus has been on genotypic adaptation to drought and high temperature, it would be possible to extend these to other abiotic and potentially biotic constraints.

Acknowledgements

The authors thank the MEF coordinators and technical staff who have contributed to the on-site development of the three facilities. These include Alan Harrod, Rick Graham, Graeme Rapp and Andy Hundt. Thanks also to Drs Rana Munns, John Kirkegaard and John Passioura for comments on the manuscript; to Dr Chris Moore, wheat breeder with InterGrain Pty Ltd for discussions; and to the Australian commercial wheat breeders who provided support in the MEFs for the assessment of germplasm under water limitation. We also thank the Grains Research and Development Corporation for their strong support in the development and funding of activities contained in the MEF.

References

Bänziger M, Setimela PS, Hodson D, Vivek B (2006) Breeding for improved abiotic stress tolerance in maize adapted to southern Africa. *Agricultural Water Management* **80**, 212–224. doi:10.1016/j.agwat.2005.07.014

Bennett SJ, Barrett-Lennard EG, Colmer TD (2009) Salinity and waterlogging as constraints to saltland pasture production: a review. *Agriculture Ecosystems & Environment* **129**, 349–360. doi:10.1016/j.agee.2008.10.013

Bonnett DG, Rebetzke GJ, Spielmeyer W (2005) Strategies for efficient implementation of molecular markers in wheat breeding programs. *Molecular Breeding* **15**, 75–85. doi:10.1007/s11032-004-2734-5

Botwright TL, Condon AG, Rebetzke GJ, Richards RA (2002) Field evaluation of early vigour for genetic improvement of grain yield in wheat. *Australian Journal of Agricultural Research* **53**, 1137–1145. doi:10.1071/AR02007

Bureau of Meteorology (2011) Climate statistics for Australian locations. (Australian Government, Bureau of Meteorology: Melbourne) Available at: http://www.bom.gov.au/climate/ [verified October 2012].

Campos H, Cooper M, Habben JE, Edmeades GO, Schussler JR (2004) Improving drought tolerance in maize: a view from industry. *Field Crops Research* **90**, 19–34. doi:10.1016/j.fcr.2004.07.003

Cattivelli L, Rizza F, Badeck FW, Mazzucotelli E, Mastrangelo AM, Francia E, Marè C, Tondelli A, Stanca AM (2008) Drought tolerance improvement in crop plants: an integrated view from breeding to genomics. *Field Crops Research* **105**, 1–14. doi:10.1016/j.fcr.2007.07.004

Ceccarelli S (1994) Specific adaptation and breeding for marginal conditions. *Euphytica* **77**, 205–219. doi:10.1007/BF02262633

Chenu K, Chapman SC, Tardieu F, McLean G, Welcker C, Hammer GL (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**, 1507–1523. doi:10.1534/genetics.109.105429

Chenu K, Cooper M, Hammer GL, Mathews KL, Dreccer MF, Chapman SC (2011) Environment characterization as an aid to wheat improvement: interpreting genotype-environment interactions by modelling water-deficit patterns in north-eastern Australia. *Journal of Experimental Botany* **62**, 1743–1755. doi:10.1093/jxb/erq459

Condon AG, Richards RA, Rebetzke GJ, Farquhar GD (2004) Breeding for high water use efficiency. *Journal of Experimental Botany* **55**, 2447–2460. doi:10.1093/jxb/erh277

Cooper M, Stucker RE, Delacy IH, Harch BD (1997) Wheat breeding nurseries, target environments, and indirect selection for grain yield. *Crop Science* **37**, 1168–1176. doi:10.2135/cropsci1997.0011183X003700040024x

Dolferus R, Xuemei JI, Richards RA (2011) Abiotic stress and control of grain number in cereals. *Plant Science* **181**, 331–341. doi:10.1016/j.plantsci.2011.05.015

Duggan BL, Richards RA, van Herwaarden AF, Fettell NA (2005) Agronomic evaluation of a tiller inhibition gene (*tin*) in wheat. I. Growth and partitioning of assimilate. *Australian Journal of Agricultural Research* **56**, 179–186. doi:10.1071/AR04153

Evans LT, Bingham J, Jackson P, Sutherland J (1972) Effect of awns and drought on the supply of photosynthate and its distribution within wheat ears. *The Annals of Applied Biology* **70**, 67–76. doi:10.1111/j.1744-7348.1972.tb04689.x

Falconer DS, Mackay TFC (1996) 'Introduction to quantitative genetics'. 4th edn. (Longman Group: Essex)

Gibson LR, Paulsen GM (1999) Yield components of wheat grown under high temperature stress during reproductive growth. *Crop Science* **39**, 1841–1846. doi:10.2135/cropsci1999.3961841x

Graybosch RA, Peterson CJ (2010) Genetic improvement in winter wheat yields in the Great Plains of North America, 1959–2008. *Crop Science* **50**, 1882–1890. doi:10.2135/cropsci2009.11.0685

Intergovernmental Panel on Climate Change (2007) Synthesis report. Contribution of Working Groups I, II and III to the fourth assessment report of the Intergovernmental Panel on Climate Change. (IPCC: Geneva)

Isbell R (2002) 'The Australian soil classification' (CSIRO Publishing: Canberra)

Jeffrey SJ, Carter JO, Moodie KB, Beswick AR (2001) Using spatial interpolation to construct a comprehensive archive of Australian climate data. *Environmental Modelling & Software* **16**, 309–330. doi:10.1016/S1364-8152(01)00008-1

Keating BA, Carberry PS, Hammer GL, Probert ME, Robertson MJ, Holzworth D, Huth NI, Hargreaves JNG, Meinke H, Hochman Z, *et al*. (2003) An overview of APSIM, a model designed for farming systems simulation. *European Journal of Agronomy* **18**, 267–288. doi:10.1016/S1161-0301(02)00108-9

Kirigwi FM, van Ginkel M, Trethowan R, Sears RG, Rajaram S, Paulsen GM (2004) Evaluation of selection strategies for wheat adaptation across water regimes. *Euphytica* **135**, 361–371. doi:10.1023/B:EUPH.0000013375.66104.04

Kirkegaard JA, Simpfendorfer S, Bambach R, Holland J, Moore K, Rebetzke GJ (2004) Effect of previous crops on crown-rot infection and yield of tolerant bread wheat and susceptible durum wheat in northern NSW. *Australian Journal of Agricultural Research* **55**, 321–334. doi:10.1071/AR03178

Liu H, Mei H, Yu X, Zhou G, Liu G, Luo L (2006) Towards improving the drought tolerance of rice in China. *Plant Genetic Resources* **4**, 47–53. doi:10.1079/PGR2006111

Lopes MS, Reynolds MP (2010) Partitioning of assimilates to deeper roots is associated with cooler canopies and increased yield under drought in wheat. *Functional Plant Biology* **37**, 147–156. doi:10.1071/FP09121

Mathews KL, Chapman SC, Trethowan R, Pfeiffer W, Ginkel M, Crossa J, Payne T, DeLacy I, Fox PN, Cooper M (2007) Global adaptation patterns of Australian and CIMMYT spring bread wheat. *Theoretical and Applied Genetics* **115**, 819–835. doi:10.1007/s00122-007-0611-4

Michelmore RW, Para I, Kesselli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis – a rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 9828–9832. doi:10.1073/pnas.88.21.9828

Midmore DJ, Cartwright PM, Fischer RA (1984) Wheat in tropical environments 2 – crop growth and grain yield. *Field Crops Research* **8**, 207–227. doi:10.1016/0378-4290(84)90064-9

Moeller C, Asseng S, Berger J, Milroy SP (2009) Plant available soil water at sowing in Mediterranean environments – is it a useful criterion to aid nitrogen fertiliser and sowing decisions? *Field Crops Research* **114**, 127–136. doi:10.1016/j.fcr.2009.07.012

Nicholls N, Drosdowsky W, Lavery B (1997) Australian rainfall variability and change. *Weather* **52**, 66–72. doi:10.1002/j.1477-8696.1997.tb06274.x

O'Shaughnessy SA, Hebel MA, Evett SR, Colaizzi PD (2011) Evaluation of a wireless infrared thermometer with a narrow field of view. *Computers and Electronics in Agriculture* **76**, 59–68. doi:10.1016/j.compag.2010.12.017

Passioura J (2007) The drought environment: physical, biological and agricultural perspectives. *Journal of Experimental Botany* **58**, 113–117. doi:10.1093/jxb/erl212

Passioura J (2010) Scaling up: the essence of effective agricultural research. *Functional Plant Biology* **37**, 585–591. doi:10.1071/FP10106

Passioura J (2012) Phenotyping for drought tolerance in grain crops: when is it useful to breeders? *Functional Plant Biology* **39**, 851–859. doi:10.1071/FP12079

Potgieter AB, Hammer GL, Butler D (2002) Spatial and temporal patterns in Australian wheat yield and their relationship with ENSO. *Australian Journal of Agricultural Research* **53**, 77–89. doi:10.1071/AR01002

Rattey A, Shorter R, Chapman S, Dreccer F, van Herwaarden A (2009) Variation for and relationships among biomass and grain yield component traits conferring improved yield and grain weight in an elite wheat population grown in variable yield environments. *Crop and Pasture Science* **60**, 717–729. doi:10.1071/CP08460

Rattey A, Shorter R, Chapman S (2011) Evaluation of CIMMYT conventional and synthetic spring wheat germplasm in rainfed sub-tropical environments. II. Correlated response for grain yield components and physiological traits due to selection for grain yield. *Field Crops Research* **124**, 195–204. doi:10.1016/j.fcr.2011.02.006

Rebetzke GJ, Condon AG, Richards RA, Farquhar GD (2002) Selection for reduced carbon-isotope discrimination increases aerial biomass and grain yield of rainfed bread wheat. *Crop Science* **42**, 739–745. doi:10.2135/cropsci2002.0739

Rebetzke GJ, Botwright TL, Moore CS, Richards RA, Condon AG (2004) Genotypic variation in specific leaf area for genetic improvement of early vigour in wheat. *Field Crops Research* **88**, 179–189. doi:10.1016/j.fcr.2004.01.007

Rebetzke GJ, van Herwaarden A, Jenkins C, Ruuska S, Tabe L, Lewis D, Weiss M, Fettell N, Richards RA (2008*a*) Quantitative trait loci for water soluble carbohydrates and associations with agronomic traits in wheat. *Australian Journal of Agricultural Research* **59**, 891–905. doi:10.1071/AR08067

Rebetzke GJ, Condon AG, Richards RA, Appels R, Farquhar GD (2008*b*) Quantitative trait loci for carbon isotope discrimination are repeatable across environments and wheat mapping populations. *Theoretical and Applied Genetics* **118**, 123–137. doi:10.1007/s00122-008-0882-4

Rebetzke GJ, van Herwaarden A, Biddulph B, Chenu K, Moeller C, Richards RA, Rattey A (2012) Protocols for experimental plot sampling, handling and processing of cereal experiments – standardised methods for use in field studies. PrometheusWiki. (CSIRO Publishing: Collingwood) Available at: http://prometheuswiki.publish.csiro.au/tiki-index.php [verified October 2012].

Richards RA, Rawson HM, Johnson DA (1986) Glaucousness in wheat: its development and effect on water-use efficiency, gas exchange and photosynthetic tissue temperatures. *Australian Journal of Plant Physiology* **13**, 465–473.

Richards RA, Rebetzke GJ, Watt M, Condon AG, Spielmeyer W, 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. doi:10.1071/FP09219

Ristic Z, Bukovnik U, Prasad PVV (2007) Correlation between heat stability of thylakoid membranes and loss of chlorophyll in winter wheat under heat stress. *Crop Science* **47**, 2067–2073. doi:10.2135/cropsci2006.10.0674

Shpiler L, Blum A (1986) Differential reaction of wheat cultivars to hot environments. *Euphytica* **35**, 483–492. doi:10.1007/BF00021856

Sirault XRR, Fettell N, Condon AG, Rebetzke GJ (2004) Does leaf rolling slow water use to maintain leaf area in a terminal drought? In 'Cereals – Proceedings of the 54th Australian Cereal Chemistry Conference and 11th Wheat Breeders Assembly, Canberra' (Eds CK Black, JF Panozza, GJ Rebetzke) pp. 52–55.

Slavich PG, Read BJ, Cullis BR (1990) Yield response of barley germplasm to field variation in salinity quantified using the EM-38. *Australian Journal of Experimental Agriculture* **30**, 551–556. doi:10.1071/EA9900551

Smith AB, Cullis BR, Thompson R (2001) Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trends. *Biometrics* **57**, 1138–1147. doi:10.1111/j.0006-341X.2001.01138.x

Smith AB, Lim P, Cullis BR (2006) The design and analysis of multi-phase plant breeding experiments. *Journal of Agricultural Science (Camb.)* **14**, 393–409.

Stone PJ, Nicolas ME (1995) A survey of the effects of high temperature during grain filling on yield and quality of 75 wheat cultivars. *Australian Journal of Agricultural Research* **46**, 475–492. doi:10.1071/AR9950475

Trethowan RM, Reynolds MW, Sayre K, Ortiz-Monasterio I (2005) Adapting wheat cultivars to resource conserving farming practices and human nutritional needs. *The Annals of Applied Biology* **146**, 405–413. doi:10.1111/j.1744-7348.2005.040137.x

Trevaskis B, Bagnall DJ, Ellis MH, Peacock WJ, Dennis ES (2003) MADS box genes control vernalization-induced flowering in cereals. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 13 099–13 104. doi:10.1073/pnas.1635053100

Turner N, Asseng S (2005) Productivity, sustainability, and rainfall-use efficiency in Australian rainfed Mediterranean agricultural systems. *Australian Journal of Agricultural Research* **56**, 1123–1136. doi:10.1071/AR05076

van Herwaarden AF, Richards RA (1999) Source–sink relationships in wheat: plant density matters. In 'Abstracts of the international conference on assimilate transport and partitioning, Newcastle, Australia, 15–20 August 1999'.

Williams JH, Hook RA, Hamblin A (2002) 'Agro-ecological regions of Australia. Methodologies for their derivation and key issues in resource management.' (CSIRO Land and Water: Canberra)

Zadoks JC, Chang TT, Konzak CF (1974) A decimal code for the growth stages of cereals. *Weed Research* **14**, 415–421. doi:10.1111/j.1365-3180.1974.tb01084.x

Zheng B, Chenu K, Dreccer MF, Chapman SC (2012) Breeding for the future: what are the potential impacts of future frost and heat events on sowing and flowering time requirements for Australian bread wheat (*Triticum aestivum*) varieties? *Global Change Biology* **18**, 2899–2914. doi:10.1111/j.1365-2486.2012.02724.x

Appendix 1. Characterisation of environment types

To characterise the drought environment types for each site, two sets of simulations were conducted using the modelling platform APSIM (Agricultural Production Systems Simulator; Keating *et al*. 2003). All simulations were performed for *Triticum aestivum* L. cv. Hartog grown at three locations across the Australian wheatbelt (Table S1, available as Supplementary Material to this paper) using 123 years of historical climate data (1889–2011; Jeffrey *et al*. 2001) and available soil information.

The initial set of simulations was used to estimate the moisture available in the soil at sowing. For each year, a fallow was simulated from 1 November (with an initial soil water content of 20% of the plant-available water (PAW) capacity). Soil water content was estimated for sowing opportunities between 1 May and 15 June. A sowing opportunity was defined to occur when (i) cumulative rainfall was at least 10 mm in the previous 10 days in Narrabri, at least 10 mm in the previous five days in Yanco or with a decreasing constraint over time for Merredin (sowing opportunity from 1 May if rainfall was at least 20 mm over 3 days, decreasing linearly up to 1 July if rainfall as at least 5 mm over 3 days); and (ii) the PAW in the entire profile was higher than 80 mm in Narrabri. At the other locations (Yanco, Merredin), sowing was not conditional upon the amounts of PAW. Results revealed that the initial soil water conditions were similar for all the sowing opportunities.

The second set of simulations was run to characterise the seasonal water deficit patterns occurring over the 123 years. Four sowing dates and five different initial soil water conditions were used for each site (each representing the median of 20% of the conditions encountered during the sowing window for the considered sites over 123 years).

Details of the second set of simulations are presented in the Table S1. Each environment (location × year × sowing date × initial soil water combination) was characterised based on a daily crop water stress index, corresponding to the ratio between water supply and demand (Chenu *et al*. 2011). For each simulation, the temporal pattern of this water stress index was centred at flowering and averaged every 100°Cd, between emergence and 450°Cd after flowering. A cluster analysis was applied to the water deficit patterns of each location using the partitioning clustering function *clara* in the R statistical package (R Foundation for Statistical Computing, Vienna, Austria) in order to identify four major environment types (ET1 to ET4) at each site. An average pattern of water deficit was calculated to describe each environment type for each location, and the frequency of occurrence of these environment types was interpreted with respect to the different sites (Fig. 2). The simulated yield for the environment types of each site is presented in Fig. 3. Analyses and map graphing were performed using the R package.

Appendix 2. Managed environment facility site characteristics

For the phenotyping of water productivity traits to be relevant, the experimental sites must be representative of those in the major production zones. These range from the subhumid, subtropical slopes and plains in the mideastern regions of the continent to the temperate, seasonally dry slopes and plains in the south-west and south-east (Williams *et al*. 2002). The growing conditions at Narrabri (30.34°S, 149.76°E; 212 m above sea level (a.s.l)) in north-western New South Wales (NSW) are characteristic for the Australian subtropics. The climate is characterised by hot summers and mild winters with a summer-dominant rainfall pattern (Fig. 2). Rainfall variability is one of the highest worldwide compared with similar climatic regions elsewhere (Nicholls *et al*. 1997). The average annual rainfall at Narrabri is 660 mm, and the average annual maximum and minimum temperatures are 26.5°C and 11.7°C (Bureau of Meteorology 2011). The experimental soil is a cracking montmorillonitic clay soil classified as grey vertosol (Isbell 2002) with a high plant-available water capacity (>200 mm). Success in the production of winter wheat depends largely on the water storage capacity of soils and the amounts of summer rainfall before sowing.

The Merredin (31.50°S, 118.22°E; 318 m a.s.l) site in south-western Western Australia (WA) and Yanco (34.60°S 146.40°E 164 m a.s.l) in southern NSW represent the wide range of environmental conditions experienced in the temperate environments of the Australian wheatbelt (Williams *et al*. 2002). The climate at both sites is characterised by hot summers and cool winters. Winter rainfall dominates in the temperate western regions of the continent, where the climate can be also described as 'typical Mediterranean' (Turner and Asseng 2005). Rainfall is more evenly distributed throughout the year in the eastern temperate environments (Fig. 2). Rainfall variability is smaller than in the eastern Australian subtropics (Nicholls *et al*. 1997) but still high enough to potentially result in crop failures (Turner and Asseng 2005).

Merredin has an average annual rainfall of 313 mm, and average annual maximum and minimum temperatures of 24.7°C and 10.7°C (Bureau of Meteorology 2011) (Fig. 2). The experimental soil belongs to the chromosol group, and has a sandy clay loam texture in the top and a light medium clay texture at maximum rooting depth (~70 cm) (Isbell 2002). The plant-available water capacity is ~80 mm. The combination of soils with a sandy texture, low to moderate plant-available water capacity and winter-dominant rainfall means that winter wheat grows largely on in-season rainfall in the western temperate region (Moeller *et al*. 2009).

Yanco experiences growing conditions that are representative of those of the eastern temperate environments. The average annual rainfall is 392 mm, and the average annual maximum and minimum temperatures are 24°C and 10.4°C, respectively (Bureau of Meteorology 2011) (Fig. 2). The experimental soil at Yanco has been classified as chromosol and has a clay-loam texture (Isbell 2002). Although rainfalls are somewhat evenly distributed throughout the year on average, high rates of evaporation during summer prevent much of the summer rainfall being available for a following winter crop (Turner and Asseng 2005).