

The application of genomics and bioinformatics to accelerate crop improvement in a changing climate

Jacqueline Batley<sup>a</sup> and David Edwards<sup>a</sup>

<sup>a</sup>School of Plant Biology and Institute of Agriculture, University of Western Australia

Corresponding author:

David Edwards

School of Plant Biology

University of Western Australia

Crawley 6009

Australia

Email:

[Dave.Edwards@uwa.edu.au](mailto:Dave.Edwards@uwa.edu.au)

## Abstract

The changing climate and growing global population will increase pressure on our ability to produce sufficient food. The breeding of novel crops and the adaptation of current crops to the new environment are required to ensure continued food production. Advances in genomics offer the potential to accelerate the genomics based breeding of crop plants. However, relating genomic data to climate related agronomic traits for use in breeding remains a huge challenge, and one which will require coordination of diverse skills and expertise. Bioinformatics, when combined with genomics has the potential to help maintain food security in the face of climate change through the accelerated production of climate ready crops.

## Introduction

There is growing acceptance that the world climate is changing and that these changes are due to the perturbation of the environment by human activity, predominantly the burning of fossil fuels and subsequent increase in atmospheric carbon dioxide concentration [1]. What continues to be debated are the exact changes that are likely to occur and their impact on agriculture. Due to the complexity of the factors influencing the climate, the accurate prediction of change is challenging [2]. However, there is general consensus that there will be warming across the globe which will lead to changes in rainfall patterns, and these changes in rainfall are likely to have a significant impact on agriculture.

Not all of the predicted changes in climate will have an immediate negative impact. Warmer temperatures may increase agricultural productivity in some temperate zones by extending the growing season. Increased carbon dioxide concentrations may increase photosynthetic efficiency, reduce water loss through transpiration and may lead to a net increase in yield. However the increase in both the regularity and severity of droughts will likely negate these positive factors in many agricultural environments, particularly at low to mid latitudes.

With the predicted changes in climate likely to disrupt global food supplies, immediate action is required to ensure global food security. In addition to the continued monitoring of climate and its impact on food crops, new crops and varieties are required which are tolerant to abiotic stress, resistant to pests and diseases and demonstrate greater nutrient use efficiency [3]. Advances will predominantly be improvements on current major crops but may include the adoption of new species which have not previously been considered for domestication.

## Climate change traits

While the impact of climate change at any particular location is difficult to predict, some general themes on crop adaptation are clear. These include improving the ability to sustain yield under drought or flooding, resilience to high and low temperature extremes, and improved resistance to pests and disease.

One of the more obvious changes in climate is the increase in atmospheric CO<sub>2</sub> concentration and this is expected to have a net positive effect on crop growth as CO<sub>2</sub> is an essential nutrient for plant growth [4]. The impact of increased CO<sub>2</sub> will be greater for C3 plants, such as wheat and rice, than C4 plants, such as maize, which have evolved mechanisms to optimise CO<sub>2</sub> efficiency. Increased CO<sub>2</sub> concentrations also improves water use efficiency as fewer stomatal openings are required for efficient gas exchange. This is important as climate change predictions for many regions include more frequent and severe droughts and floods. While it is not possible to grow crops in the absence of water, improved water use efficiency as

well as the ability to survive and sustain yield under drought stress will become increasingly important traits. Drought tolerance and water use efficiency are complex traits which have strong environmental interactions, so while there will be no 'golden bullet', selection for improved adaption to drier conditions while maintaining yield is an important aim for crop breeders. Significant advances have been made in the area of flooding tolerance in rice with the introduction of the Sub1 gene [5,6] demonstrating the impact of genomics to improve food security under climate change.

Increased plant growth and reduced transpiration associated with higher levels of atmospheric CO<sub>2</sub> are predicted to affect the nutritional quality of food, with a reduction in nitrogen content in some species [7]. This could have the greatest impact in developing countries where dietary nutrition is already often poor. The selection of quality traits frequently takes second place to overall yield, and maintaining yield and quality will become more challenging under predicted climate change scenarios.

One often overlooked impact of climate change is the predicted increase in abundance and virulence of pests and diseases [8,9], and a greater understanding of disease virulence and host resistance mechanisms is likely to reduce the impact of pests and pathogens in major crops and developed countries.

### The role of genomics

Genomics and bioinformatics are rapidly expanding fields of research, being fuelled by the continued growth and reducing cost of DNA sequencing and genotyping [10,11]. While many of the initial genomics research areas focused on increasing our fundamental understanding of biology, there has been more recent growth of the application of genomics in area of agriculture. The translation of genomics to improve human health and nutrition will have a major impact on society both in developed and developing countries.

The increased application of genomics in agriculture is timely, coming when food production faces the dual challenge of unprecedented population growth and climate change. Draft genome assemblies are now available for many of the major crops as well as an increasing number of wild relatives [12], and international efforts through the DivSeek and African orphan crops initiatives (<http://www.divseek.org/>; <http://africanorphancrops.org/>) aim to promote and coordinate the capture of crop genomic diversity for application in crop improvement.

While one reference genome sequence provides valuable information, it does not represent the diversity within a species, and it is this diversity which is mined for selection and breeding. There are two main types of heritable genomic diversity; single nucleotide polymorphisms (SNPs) and structural variation such as copy number or presence/absence variations (CNVs/PAVs). Molecular markers developed

from SNPs have the resolution of a single nucleotide and are highly abundant across genomes [13]. They can be discovered in large numbers through genome resequencing and assayed using a wide range of approaches [14]. The importance of structural variation to both genomic and phenotypic diversity is only now being realised, with up to 20% of genes in some plant species showing absence in some individuals [15-17]. The combination of gene sequence diversity and the presence of variable genes is likely to contribute to the hybrid vigour observed when two diverse individuals are crossed [18,19].

The identification of genome diversity is of little value unless the diversity can be associated with heritable agronomic traits. Traditional trait mapping was performed on the progeny of pairwise crosses which have relatively low resolution due to limited recombination, and lack the ability to assess broader impacts of multi-gene interactions. More recent trait mapping approaches have applied multi parent populations such as nested association mapping (NAM) or magic (multiparent advanced generation intercross) populations [20-23]. These populations provide much greater resolution due to the higher frequency of recombination captured.

Genotyping populations has also undergone a revolution in the last decade, with a move from PCR based single markers such as simple sequence repeats, towards highly parallel SNP based markers using genotyping arrays from providers such as Illumina and Affymetrix, or more recently genotyping by sequencing (GBS) approaches [24,25]. With the ever reducing cost of generating DNA sequence data, GBS approaches, and particularly resequencing or skimGBS based methods [19,26] are likely to dominate the future of crop genotyping.

One of the bottlenecks of trait association has been the production of suitable phenotypic information for trait association [27]. This is being addressed through the use of high throughput automated glasshouse phenotyping as well as the rapid increase in the use of modified drones to visualise and assess crops [28]. These drones are increasingly being used for routine on-farm assessment of crop growth, and there is the potential to extend remote monitoring as routine agricultural practice to enable the assessment of cultivars under diverse geographical and environmental conditions.

### The role of bioinformatics

The flood of information from genome diversity studies, phenotypic assays as well as gene expression, proteomic, metabolomics, epigenetic and other related studies makes the application of this information for applied crop improvement a complex challenge. There are several database systems designed to manage this information for crop improvement; such as the integrated breeding platform (<https://www.integratedbreeding.net/>) and the Triticeae toolbox (<https://triticeaetoolbox.org>), however they do not capture the full range of relevant public information, and information related to historic published experiments or

genes from related species is particularly absent. Crop specific information consortia such as the wheat information system (<http://wheatis.org/>) and the rice informatics consortium (<http://iric.irri.org/>) also aim to collate genomic information and translate this for applied crop improvement. One of the limitations of managing and integrating such large and diverse information is that the standard tabular design of traditional databases does not efficiently manage the diverse interactions between data or facilitate broad queries. New developments in graph database design offer opportunities to address this limitation and systems such as ondex [29] and Neo4j [30] open up the potential to explore and mine diverse data interactions including historic published data on related species which may be relevant to the crop. The application of graph databases to genomics is still in its infancy, but as they develop and become more sophisticated they are likely to become widely adopted both for understanding fundamental biology as well as for practical applied breeding tools.

### Global impact of climate change

It is clear that the impact of climate change on agriculture will vary across the globe, both in the effect on crops and also the ability of societies to adapt and respond to the changing climate. Many developing countries have not experienced the intensification of agriculture undertaken in developed countries and many regions rely on subsistence farming of a limited number of crops species. These regions are particularly susceptible to the impact of climate change and are unlikely to benefit greatly from the advances in genomics seen in more mainstream crops without funding for specific crops, combined with support to increase agricultural intensification and associated food storage and distribution networks. Once these broader improvements in agricultural practice are in place however, there is the opportunity to rapidly advance crops that currently have little investment by shortening breeding cycles and improving genetic gain. In the short term, countries with intensive agricultural production are likely to benefit most from advances in genomics as established industries invest and compete for their share of the seed market leading to continued improvement in varieties for yield, quality and climatic resilience (Figure 1).

### Conclusions

There is a danger that without targeted action to promote food security, climate change may lead to major food shortages and increase the disparity of wealth and wellbeing across the globe. Advances in genomics and bioinformatics have the potential to accelerate the production of improved crops, improving global food security in the face of climate change.

### Acknowledgements

The authors would like to acknowledge funding from the Australian Research Council (LE140100061, LP140100537, LP130100061, LP130100925, LP110100200, LP0989200, DP0985953, LP0883462, LP0882095).

Figure 1. Pathways for genomics assisted breeding of new crops.

The application of advanced breeding strategies for the production of new crops will increasingly be based on knowledge of genome variation and its relationship with agronomic traits.

## References

\*\*1. IPCC: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Field, C.B., V.R. Barros, D.J. Dokken, K.J. Mach, M.D. Mastrandrea, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L. White (eds.)]. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press; 2014.

Report of the Intergovernmental Panel on Climate Change is the most current authoritative account of predicted climate change scenarios and its impact on global agriculture.

\*2. Asseng S, Ewert F, Rosenzweig C, Jones JW, Hatfield JL, Ruane AC, Boote KJ, Thorburn PJ, Rotter RP, Cammarano D, et al.: Uncertainty in simulating wheat yields under climate change. *Nature Clim. Change* 2013, 3:827-832.

Highlights the challenge of predicting changes in yield under climate change.

3. Abberton M, Batley J, Bentley A, Bryant J, Cai H, Cockram J, Costa de Oliveira A, Cseke LJ, Dempewolf H, De Pace C, et al.: Global agricultural intensification during climate change: a role for genomics. *Plant Biotechnology Journal* 2015:n/a-n/a.

4. Lawlor DW, Mitchell RAC: The effects of increasing CO<sub>2</sub> on crop photosynthesis and productivity - a review of field studies. *Plant Cell and Environment* 1991, 14:807-818.

\*5. Mackill DJ, Ismail AM, Singh US, Labios RV, Paris TR: Development and rapid adoption of submergence-tolerant (Sub1) rice varieties. *Advances in Agronomy* 2012, 115.

An excellent example of how a single gene can improve climate resilience in a major crop.

6. Xu KN, Xu X, Fukao T, Canlas P, Maghirang-Rodriguez R, Heuer S, Ismail AM, Bailey-Serres J, Ronald PC, Mackill DJ: Sub1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice. *Nature (London)* 2006, 442.

7. Cotrufo MF, Ineson P, Scott A: Elevated CO<sub>2</sub> reduces the nitrogen concentration of plant tissues. *Global Change Biology* 1998, 4:43-54.

\*8. Garrett KA, Dendy SP, Frank EE, Rouse MN, Travers SE: Climate change effects on plant disease: Genomes to ecosystems. In *Annual Review of Phytopathology*. Edited by; 2006:489-509. *Annual Review of Phytopathology*, vol 44.]

This paper highlights the indirect impact of climate change on crops through changes in plant disease.

9. Gregory PJ, Johnson SN, Newton AC, Ingram JSI: Integrating pests and pathogens into the climate change/food security debate. *Journal of Experimental Botany* 2009, 60:2827-2838.
10. Edwards D, Batley J: Plant genome sequencing: applications for crop improvement. *Plant biotechnology journal* 2010, 8:2-9.
11. Edwards D, Batley J, Snowden RJ: Accessing complex crop genomes with next-generation sequencing. *Theoretical and Applied Genetics* 2013, 126:1-11.
12. Brozynska M, Furtado A, Henry RJ: Genomics of crop wild relatives: expanding the gene pool for crop improvement. *Plant Biotechnology Journal* 2015:n/a-n/a.
13. Edwards D, Forster JW, Chagné D, Batley J: What are SNPs? In *Association Mapping in Plants* Edited by Oraguzie NC, Rikkerink EHA, Gardiner SE, De Silva HN: Springer 2007:41-52.
14. Imelfort M, Duran C, Batley J, Edwards D: Discovering genetic polymorphisms in next-generation sequencing data. *Plant Biotechnology Journal* 2009, 7:312-317.
15. Li YH, Zhou GY, Ma JX, Jiang WK, Jin LG, Zhang ZH, Guo Y, Zhang JB, Sui Y, Zheng LT, et al.: De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nature Biotechnology* 2014, 32:1045-+.
16. Lin K, Zhang N, Severing E, Nijveen H, Cheng F, Visser R, Wang X, de Ridder D, Bonnema G: Beyond genomic variation - comparison and functional annotation of three *Brassica rapa* genomes: a turnip, a rapid cycling and a Chinese cabbage. *BMC Genomics* 2014, 15:250.
17. Springer NM, Ying K, Fu Y, Ji TM, Yeh CT, Jia Y, Wu W, Richmond T, Kitzman J, Rosenbaum H, et al.: Maize Inbreds Exhibit High Levels of Copy Number Variation (CNV) and Presence/Absence Variation (PAV) in Genome Content. *Plos Genetics* 2009, 5.
18. Miller M, Song QX, Shi XL, Juenger TE, Chen ZJ: Natural variation in timing of stress-responsive gene expression predicts heterosis in intraspecific hybrids of *Arabidopsis*. *Nature Communications* 2015, 6.
19. Voss-Fels K, Snowden RJ: Understanding and utilizing crop genome diversity via high-resolution genotyping. *Plant Biotechnology Journal* 2015:n/a-n/a.
20. Schnaithmann F, Kopahnke D, Pillen K: A first step toward the development of a barley NAM population and its utilization to detect QTLs conferring leaf rust seedling resistance. *Theoretical and Applied Genetics* 2014, 127:1513-1525.

21. Zhang NY, Gibon Y, Wallace JG, Lepak N, Li PH, Dedow L, Chen C, So YS, Kremling K, Bradbury PJ, et al.: Genome-Wide Association of Carbon and Nitrogen Metabolism in the Maize Nested Association Mapping Population. *Plant Physiology* 2015, 168:575-583.
22. Pascual L, Desplat N, Huang BE, Desgroux A, Bruguier L, Bouchet JP, Le QH, Chauchard B, Verschave P, Causse M: Potential of a tomato MAGIC population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era. *Plant Biotechnology Journal* 2015, 13:565-577.
23. Sallam A, Martsch R: Association mapping for frost tolerance using multi-parent advanced generation inter-cross (MAGIC) population in faba bean (*Vicia faba* L.). *Genetica* 2015, 143:501-514.
24. Poland JA, Brown PJ, Sorrells ME, Jannink J-L: Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. *Plos One* 2012, 7:e32253.
25. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE: A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *Plos One* 2011, 6.
26. Bayer PE, Ruperao P, Mason AS, Stiller J, Chan CKK, Hayashi S, Long Y, Meng J, Sutton T, Visendi P, et al.: High resolution skim genotyping by sequencing reveals the distribution of crossovers and gene conversions in *Cicer arietinum* and *Brassica napus*. *Theoretical and Applied Genetics* 2015, (Accepted January 2015).
27. Fahlgren N, Gehan MA, Baxter I: Lights, camera, action: high-throughput plant phenotyping is ready for a close-up. *Current Opinion in Plant Biology* 2015, 24:93-99.
28. Liebisch F, Kirchgessner N, Schneider D, Walter A, Hund A: Remote, aerial phenotyping of maize traits with a mobile multi-sensor approach. *Plant Methods* 2015, 11.
29. Köhler J, Baumbach J, Taubert J, Specht M, Skusa A, Rüegg A, Rawlings C, Verrier P, Philippi S: Graph-based analysis and visualization of experimental results with ONDEX. *Bioinformatics* 2006, 22:1383-1390.
30. Miller JJ: Graph Database Applications and Concepts with Neo4j. In *Proceedings of the Southern Association for Information Systems Conference*, Atlanta, GA, USA March 23rd-24th: 2013.