



Maximizing the potential of multi-parental crop populations



Olufunmilayo Ladejobi^{a,b}, James Elderfield^b, Keith A. Gardner^a, R. Chris Gaynor^c, John Hickey^c, Julian M. Hibberd^b, Ian J. Mackay^a, Alison R. Bentley^{a,*}

^a The John Bingham Laboratory, NIAB, Huntingdon Road, Cambridge CB3 0LE, United Kingdom

^b Department of Plant Sciences, The University of Cambridge, Downing Street, Cambridge CB2 3EA, United Kingdom

^c The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush Research Centre, Midlothian EH25 9RG, United Kingdom

ARTICLE INFO

Article history:

Received 15 July 2016

Received in revised form 22 October 2016

Accepted 24 October 2016

Keywords:

Mapping
Trait dissection
Wheat
MAGIC
NAM

ABSTRACT

Most agriculturally significant crop traits are quantitatively inherited which limits the ease and efficiency of trait dissection. Multi-parent populations overcome the limitations of traditional trait mapping and offer new potential to accurately define the genetic basis of complex crop traits. The increasing popularity and use of nested association mapping (NAM) and multi-parent advanced generation intercross (MAGIC) populations raises questions about the optimal design and allocation of resources in their creation. In this paper we review strategies for the creation of multi-parent populations and describe two complementary *in silico* studies addressing the design and construction of NAM and MAGIC populations. The first simulates the selection of diverse founder parents and the second the influence of multi-parent crossing schemes (and number of founders) on haplotype creation and diversity. We present and apply two open software resources to simulate alternate strategies for the development of multi-parent populations.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Contents

1. Introduction	9
2. Materials and methods	10
2.1. Selecting founders	10
2.2. Simulating multi-parent populations.	12
2.2.1. Pedigrees	12
2.2.2. Simulations	12
3. Results	13
3.1. Selecting founders	13
3.2. Simulating multi-parent populations.	13
4. Discussion	15
Acknowledgements	16
References	16

1. Introduction

Expanded genetic diversity is required to address the perpetual challenges of quantitative trait dissection. In crops, mapping populations developed from two contrasting parents have been popular for creating

novel recombinants and haplotypes for key crop traits (e.g. the UK wheat reference population Avalon × Cadenza; see www.wgin.org.uk; Ma et al., 2015). Bi-parental mapping populations are simple to develop and possess high power for QTL detection (Semagn et al., 2006; Xu et al., 2016). However, combining the genomes of only two parents results in a relatively narrow genetic base and inadequately represents wider allelic diversity (Jannink, 2007). Despite this, linkage based quantitative trait locus (QTL) mapping using bi-parental populations is the most widely used method of identifying regions of genome controlling phenotypic variation (Bernardo, 2008).

* Corresponding author.

E-mail address: alison.bentley@niab.com (A.R. Bentley).

Genome-wide association (GWA) or linkage disequilibrium mapping is a complementary method exploiting linkage disequilibrium (LD) as a function of historical recombination for QTL mapping. GWA studies however are prone to detection of false positive QTLs due to unknown population structure and genetic relatedness among the lines (Lewis, 2002; Zhao et al., 2007) and statistical approaches may also over-compensate for population structure (Segura et al., 2012), thereby lowering the accuracy of QTL detection. In addition, low frequency rare variant QTLs may be undetected despite having large effects (Bresgello and Sorrells, 2006; Mackay and Powell, 2007).

Multi-parent populations (MPPs) have emerged as next-generation mapping resources combining diverse genetic founder contributions with high levels of recombination (Mackay and Powell, 2007; Cavanagh et al., 2008), overcoming some of the limitations of bi-parental and GWA populations (Huang et al., 2011). The two most commonly developed forms of MPPs in crop genetics are nested association mapping (NAM) and multi-founder advanced generation inter-cross (MAGIC) populations. Derivation from greater than two parents and structured inter-mating maximizes allelic diversity and facilitates the inclusion of novel recombinants. Creating controlled populations from crosses between multiple well-characterized parents allows the derivation of individuals which feature diverse levels and patterns of recombination and new genotype and haplotype combinations. These features are exploited for trait mapping with the contribution of multiple founders increasing the potential genetic diversity in advanced lines (Yu et al., 2008).

NAM populations were designed to increase the power and precision of QTL mapping by combining the advantages of association mapping and bi-parental populations. NAM populations can effectively capture rare alleles allowing new loci to be seen (McMullen et al., 2009). Populations are derived by crossing a single inbred parent to a successive collection of diverse inbred lines. The first NAM population was created in maize, derived from crosses between the maize reference line B73 and 25 diverse inbred lines to produce 5000 recombinant inbred lines (RILs) (Yu et al., 2008). These capture thousands of recombination events but recombination and segregation distortion varies among different families which can limit the precision of genetic dissection of quantitative traits (McMullen et al., 2009). The maize NAM has been used to study the genetic architecture of a number of morphological and disease resistance traits (Buckler et al., 2009; Tian et al., 2011; Cook et al., 2012; Bajgain et al., 2016). A NAM derived advanced backcross population has been recently developed for barley which combines wild barley landraces into the exotic background Rasmusson (Nice et al., 2016).

MAGIC populations are developed by inter-crossing multiple (typically four, eight or sixteen) parental lines in a balanced funnel crossing scheme. The resulting RILs are highly recombined mosaics of the founder genomes. Multi-cross populations were first proposed for mouse known as heterogeneous stock and collaborative cross populations (Mott et al., 2000; Valdar et al., 2006b; Threadgill and Churchill, 2012) and for plants by Mackay and Powell (2007). They are also similar to the *Arabidopsis* multi-parent recombinant inbred line (AMPRIL) population described by Huang et al. (2011) which was developed from diallel crossing of eight *Arabidopsis* accessions from diverse geographical origins. In MAGIC, high levels of recombination result in low LD and give high mapping resolution. A high density MAGIC linkage map has recently been developed in wheat (Gardner et al., 2016). MAGIC populations have been developed in many plant species including *Arabidopsis* (Kover et al., 2009), tomato (Pascual et al., 2015), barley (Sannemann et al., 2015), maize (Dell'Acqua et al., 2015), sorghum (Higgins et al., 2014), wheat (Huang et al., 2012; Mackay et al., 2014) and rice (Bandillo et al., 2013).

Trait mapping in structured MPPs involves the use of statistical models developed based on their theoretical properties. Many

models for genetic data analysis have been generated by computer simulation to determine the properties and outcomes of an experimental design. For example, simulation studies in MPPs can be applied to determine the optimal number of founder lines, crosses and the size of the population needed to effectively track the genetic architecture of quantitative traits (Myles et al., 2009). Kover et al. (2009) simulated the effects of MPP size on mapping resolution and power for QTL detection determining that QTL detection error rates decreased when population size increased and QTL could be mapped to smaller intervals. Simulation studies typically generate *in silico* data describing population specific genetic polymorphism which are then used to describe, solve or predict. Because *in silico* data sets are not subject to the same inconsistencies as real datasets, they predict outcomes for specified scenarios (Yu et al., 2006, 2008; Hoban et al., 2011). Verbyla et al. (2014) simulated the effect of a joint analyses of multiple environmental and multiple trait datasets on QTL detection accuracy and to infer QTL-by-environment interactions in MAGIC.

MPPs are increasingly used in crop genetics and schemes for their creation vary in design. In this paper we present simulations using two open source software applications that analyse the selection of founders and the properties of both NAM and MAGIC population types. We compare schemes in which the number of crosses and the number of parents vary. The function of MPPs can be viewed as the creation of haplotype diversity for fine mapping and selection and the different schemes were therefore quantified as the number of haplotypes created for a range of MPP configurations.

2. Materials and methods

2.1. Selecting founders

Two methods of selecting subsets of individuals from populations to maximize genetic diversity have previously been implemented using PowerMarker analysis software (Liu and Muse, 2005) and can be used to select founding individuals for MPPs. These methods are (i) selection using total number of segregating alleles and (ii) selection using average gene diversity (Nei, 1973). The PowerMarker analysis software used a simulated annealing algorithm that allowed for efficient selection of individuals from within a large set of germplasm for which performing an exhaustive search would be infeasible. However, PowerMarker is no longer actively supported and a functional version of the software is no longer publicly available. To fill this void, we implemented a complementary method using genetic algorithms. These genetic algorithms were developed using the R package 'GA' (Scrucca, 2013) which provides a flexible, general-purpose package for this purpose. This flexibility was used to define custom objective functions and genetic operators for implementing each method. The scripts used to implement these methods are available (<http://www.niab.com/pages/id/326/Resources>) and are also available as Supplementary information.

The performance of these methods was examined using the 376 wheat varieties in the TriticeaeGenome association mapping panel (Bentley et al., 2014; dataset available as above). Each line was genotyped with 2535 polymorphic DArT markers (Jaccoud et al., 2001). Each method was used to select two, four, eight, sixteen and twenty six line subsets that could be used to generate MPPs. Average performance of each method was measured across ten replicates and compared to selection of random individuals on the basis of percentage of polymorphic loci and average gene diversity. Selection of the two line subset was compared against the best possible subset for percentage of polymorphic loci and average gene diversity using an exhaustive search of all possible combinations.

Download English Version:

<https://daneshyari.com/en/article/5584267>

Download Persian Version:

<https://daneshyari.com/article/5584267>

[Daneshyari.com](https://daneshyari.com)