ABSTRACT: The Collaborative Cross (CC) is a mouse genetic resource for high resolution mapping of disease susceptibilities and other complex traits. The CC will consist of 600 recombinant inbred lines of which 350 are at advanced breeding generations and available for study. These lines were generated by reciprocal crosses between 8 founder lines: 5 laboratory strains, and 3 derived from wild accessions (two Mus subspecies, and one Mus domesticus). The CC resource presents extremely wide genetic diversity relative to existing mouse resources, and has shown unprecedented high resolution QTL mapping. Genotypes and phenotypes are stored centrally and are available freely, making the CC a ‘Genotype once, phenotype once’ resource. Since many diseases and complex traits (e.g., meat quality, carcass composition, dietary requirements), are common to mice and livestock, the CC resource will be a superb mouse model for exploring the genomics of agricultural animals.

Keywords: QTL mapping; complex traits; disease susceptibility

Introduction

Livestock populations, even when reared under the same age-group, herd or flock management conditions, display wide phenotypic variation in performance and functional traits; susceptibility to disease; and response to environmental modifications and veterinary treatments. Heritability estimates assign 5% to 50% of this phenotypic variation to genetic sources. QTL mapping has anchored much of this genetic variation to chromosomal regions; in some instances the mapped QTL account for a major fraction of total variation in breeding values. However, for most traits, biometrical analysis and QTL mapping leave a major fraction of trait variation unexplained. Breeding values and QTL mapping assess the average effect of a gene substitution in an average environment and genetic background. Thus, deviation of an individual animal from breeding value expectation, can be due to three factors: Micro-environmental variation, and epistatic and gene x environment interactions. Precision agriculture attempts to identify predictive phenotypic or genetic biomarkers which explain part of this residual variation, and use this information to provide individuated management that will increase animal performance, health and welfare.

Fulfilling the promise of precision agriculture depends to a large extent on identifying the genes determining differential susceptibility to lifestyle and pathogen challenge, or differential response to management treatment. This poses a problem, as genetic variation in most susceptibility and response traits is “complex”, i.e., due to the interplay of numerous genetic factors interacting with one another and with the environment. Identifying the individual genetic factors underlying this variation in animal populations, has had limited success because of the difficulty of controlling and standardizing challenge, and because of the limitations of animal population and pedigree structure for genomic analysis.

Mice and livestock are susceptible to many of the same pathogens and lifestyle challenges, and almost all mouse genes have homologues in the livestock genome. Furthermore, numerous inbred mouse lines are readily available, enabling the genetics of disease susceptibility and response to treatments to be studies in crosses between lines, or as association studies across lines. Nevertheless, for the most part, mouse models based on existing mouse inbred-lines, have not succeeded in identifying genes that are relevant to animal agriculture. The reasons for this are twofold. Most existing mouse strains derive from the same narrow genetic base. This limits the genetic diversity that can be uncovered in crosses among strains; while the complex historical pedigree of the current mouse strains limits the precision of association studies across strains. Thus, the existing collection of mouse strains is inadequate to meet the demands of precision agriculture.

These developments in agriculture closely parallel corresponding developments in human medicine. Here too, there is a shift to precision medicine, which takes into account varied responses of different individuals to lifestyle and pathogen challenges, and to pharmaceutical or lifestyle treatments. The goal is to identify predictive biomarkers that will enable personalized medicine leading to more successful outcomes. To meet the demands of personalized medicine, the mouse genetic community developed the Collaborative Cross (CC), a “next generation” mouse genetic reference population (Threadgill et al. (2002)). It is the purpose of this paper to introduce this resource to the animal genomics community. We believe that it has equally as much to offer to livestock as to human genomics.

Material and Methods

The CC inbred lines are derived by reciprocal crosses (see Figure 1) among 8 divergent founder lines: 5
standard laboratory lines, and three recent inbred isolates of wild mouse subspecies \((M.\ m.\ castaneum,\ M.\ m.\ musculus\) and \(M.\ m.\ domesticus\)). Currently, over 350 (of 600 planned) lines are at advanced stages of inbreeding at Tel Aviv University, Israel; University of North Carolina, USA; and Geniad Ltd, Western Australia; also active initially were Jackson Lab and Oxford University. A Material Transfer Agreement (MTA) can be obtained from any of these five labs (Welsh et al. (2012)).

Figure 1: Development of one line of CC mice. A, B, C, D, E, F, G, H, are the eight founder strains of the CC lines. Di-, quatro- and octo-allelic lines are developed by successive crosses. The octo-allelic lines are the starting point of inbreeding process leading to an RIL.

Each CC line contains an equal proportion of genome from each of the founder lines, but as a result of recombination, each line is a unique combination of segments from the eight founders (Figure 2) (Roberts et al., (2007)). Once the founder-haplotype composition of a line is known, the entire DNA sequence is inferred from the sequence of the founder lines. Further studies can then use this CC line without further genotyping, the "genotype once" paradigm. Additionally, all phenotypes obtained on the CC lines will be stored in a unified database and made freely available. Thus, in principle, the lines need only be phenotyped once for any trait, "Genotype once, phenotype once!"

Results and Discussion

Genetic diversity in the CC resource. The wild mouse strains greatly increased genetic variation in the CC. \(M.\ m.\ musculus\) and \(M.\ m.\ castaneum\): each differed at about 17 M SNPs, from the reference C57BL/6J strain, while wild \(M.\ m.\ domesticus\) differed at 6 M SNPs; standard laboratory strains differ at only 4 M SNPs. Across all eight founders, there were a total of 36 M SNPs (Keane et al. (2011)). This enormous sequence variation, is mirrored at least to some extent, by much increased genetic variation at the trait level, as measured by the Genetic Coefficient of Variation \((CV_G)\) also termed the evolvability parameter (the ratio of the genetic standard deviation of a trait to its mean). For a typical animal population, \(CV_G\) are in the range 0.04 to 0.08, depending on heritability. For a wide spectrum of traits in the CC lines, \(CV_G\) was much higher, in the range 0.15 to 0.60 (Iraqi et al, 2014). Consequently, QTL mapping using the CC tends to uncover novel QTLs originating from the wild-derived strains. Figure 3, shows variation among the CC lines in body fat %; \(CV_G\) was 0.22.

Figure 2: Reconstructions of the genome of a representative CC line. The X-axis shows the 19 autosomes. The Y-axis shows the 8 CC founders, Regions attributed to a single founder appear as dark horizontal bands in the corresponding lane. Regions where two or more putative founders cannot be distinguished are gray (Durrant et al, (2011)).

Figure 3: Percentage of total % body fat in 20-week old mice using DEXA scan. The x-axis represents the 15 tested CC lines while y-axis represents % body fat with SE (unpublished data).

Figure 4. Genome scan of susceptibility to \(Aspergillus\ fumigatus\) in the CC resource. The x-axis is genome location, y-axis is the logP of the test of association between locus and survival time. 50%, 90% and 95% confidence levels are indicated by the horizontal grey lines (Durrant et al. (2011))

QTL mapping in the CC resource. QTL mapping in the CC resource is implemented as for an F2 population; each CC line represents an individual of the population. However, effective mapping power is much greater than an F2. Due to crossing over during the formation of the lines, the CC population exhibits two-fold map expansion, increasing accuracy of QTL map location accordingly. Because the CC lines are inbred, all genes are present in homozygous state, so that all lines are informative for mapping. Multiple individuals can be phenotyped in each line reducing environmental variation. Taking these aspects into account, the mapping power of the CC is increased many-fold relative to standard F2 populations (Valdar et al. (2006)). In practice, mapping power turned out to be even greater than expected. This was due to a number of factors: (i) the afore-mentioned QTL with very
powerful effects were introduced by the wild Mus founders (ii) the powerful mapping analysis employed, based on identifying each marker allele according to the founder population from which it is derived. This converted the marker analysis to an 8-allele analysis based on founder origin, rather than a two-allele analysis based on marker genotype (Mott et al, (2000)). (iii) It was possible to go back from the QTL mapping to the founder strain and identify regions of similar effect across different strains. This limits the location of the QTL to the chromosomal region common to the different founders presenting the same QTL allele. In sum, by phenotyping a relatively small number (<100) of CC lines with modest replication (about ~ 4 individuals per line), it was possible in many instances to achieve mapping resolution below 1Mb (Iraqi et al. (2014)).

**Conclusions**

It is clear from the above that the CC resource will be equally effective in uncovering the genomic basis for susceptibility to livestock diseases, and even more effective for genetic analysis of many performance and functional traits. Carrying out a mapping analysis in the CC resource will require defining the mouse equivalent to the livestock trait, often a simple adaptation of phenotyping methodologies, e. g., for meat quality. As data on a wide variety of phenotypes accumulates for the CC lines, it will be possible to select CC lines for novel phenotyping that will be most informative. A tremendous effort is now underway to develop user-friendly computational methodologies for the CC resource and these will be available to analysis of traits of agricultural interest as well. Leads to candidate genes, gene by gene and gene by environmental interactions will be generated by the CC resource for further confirmation and exploration in livestock populations. We can foresee a time when a mouse model will be an indispensable component of livestock genomics.

**Literature Cited**


