Abstract

The mouse is the most extensively used mammalian model for biomedical and aging research, and an extensive catalogue of laboratory resources is available to support research using mice: classical inbred lines, genetically modified mice (knockouts, transgenics, and humanized mice), selectively bred lines, consomics, congenics, recombinant inbred panels, outbred and heterogeneous stocks, and an expanding set of wild-derived strains. However, these resources were not designed or intended to model the heterogeneous human population or for a systematic analysis of phenotypic effects due to random combinations of uniformly distributed natural variants. The Collaborative Cross (CC) is a large panel of recently established multiparental recombinant inbred mouse lines specifically designed to overcome the limitations of existing mouse genetic resources for analysis of phenotypes caused by combinatorial allele effects. The CC models the complexity of the human genome and supports analyses of common human diseases with complex etiologies originating through interactions between allele combinations and the environment. The CC is the only mammalian resource that has high and uniform genomewide genetic variation effectively randomized across a large, heterogeneous, and infinitely reproducible population. The CC supports data integration across environmental and biological perturbations and across space (different labs) and time.

Key Words: aging; Collaborative Cross (CC); mouse model; recombinant inbred mice; systems genetics; genetic reference population

Recombinant Inbred Lines and Genetic Reference Populations

Recombinant inbred (RI) lines have been generated in many organisms including mice and rats (Shisa et al. 1997). Traditional mouse RI lines are derived from two inbred strains through the brother-sister mating of F2 individuals and their descendants until each independent line reaches inbred status (Bailey 1971). A complete RI panel is composed of a defined set of inbred strains, each representing a unique mosaic of the original genomes. RI lines retain all the advantages of inbred strains (reproducibility and integration of data across space and time) while providing a simple and cost-effective approach to genetic mapping (Bailey 1971). RI resources have been widely used in genetic research and have become popular as tools for systems genetics, the integration of data across biological scales (Threadgill 2006).

The usefulness of traditional RI panels as platforms for systems genetics is limited by the presence of only two parental genomes and the number of lines in the RI panels. The first limitation of low genetic diversity is caused by the fact that existing RI panels use two classical inbred strains as founders and a large fraction of the genome of all pairs of classical inbred strains is identical by descent (IBD1; Yang et al. 2007, Yang et al. submitted). IBD in the founders leads to a nonuniform distribution of genetic variation across the genome in the RI panel (Roberts et al. 2007). More importantly, as a result of the lack of variation, IBD regions are effectively “blind spots” for genetic mapping (Roberts et al. 2007), a critical shortcoming that can be partly remedied by substituting a classical inbred strain with a wild-derived inbred strain in the founder pair. However, it is not possible to completely overcome the presence of blind spots because of the pervasive presence of intersubspecific introgression and contamination in wild-derived laboratory strains (Yang et al. submitted). Others have attempted to overcome the shortcoming of limited genetic diversity by using panels of extant inbred strains (e.g., the Mouse Phenome Project, MPP) that feature variation across most of the mouse genome (Bennett et al. 2010; Paigen and Eppig 2000; Svensson et al. 2007). The intertwined breeding history of extant strains is problematic as.

1Abbreviations used in this article: CC, Collaborative Cross; IBD, identical by descent; ORNL, Oak Ridge National Laboratory; RI, recombinant inbred; SOP, standard operating procedure
multiple variable regions of the genome have similar distribution patterns across strains, making it impossible to uniquely distinguish associations to different areas of the genome (Roberts et al. 2007; Yang et al. 2007).

The second limitation of small numbers of lines in existing RI panels has been overcome by generating larger panels such as the 81 lines of the BXD panel (C57BL/6J × DBA/2J; Peirce et al. 2004) and the 77 lines of the LXS panel (Williams et al. 2004). These panel sizes remain relatively small, but their primary weakness is that they are biparental, so a large fraction of the genomes are IBD and thus invisible for genetic analysis.

The main impetus behind the development of the Collaborative Cross (CC) was to overcome the limitations of existing RI panels and to create the first genetic reference population suitable for systems approaches to biomedical research (Threadgill et al. 2002). The CC project involved generation of a large panel of RI strains derived from eight founder inbred strains through randomized breeding (Churchill et al. 2004). The generation of each CC line requires three generations of intercrossing followed by at least 20 generations of brother-sister mating or other selection methods to ensure the lines are inbred (Figure 1). The overall aim of the CC project was to create hundreds of independent RI lines.

The CC: A Community Resource

Since its inception the CC was conceived as a community resource that would complement the traditional strengths of the mouse as a model of human disease and create novel avenues for scientific research (Threadgill et al. 2002). During the initial phase several scientists hosted meetings to discuss the breeding schema, the number and identity of the founder strains, and the logistics of the generation of the cross. The Complex Traits Consortium (CTC, recently renamed the Complex Traits Community) assumed leadership (Churchill et al. 2004) and the CC became its most tangible project along with the annual CTC meeting.2

A critical aspect of the early discussions to establish the CC was the identification of appropriate funding sources. Because generation of the CC was expected to take more than 5 years and its utility would become clear only after most of the lines were completed, securing funds was a challenge. Although long-term planning and support for community resources are common in some scientific fields such as astronomy, they are rare in biomedical research. Funding to initiate the CC was provided by the Department of Energy, the Ellison Medical Foundation, and the Wellcome Trust, with subsequent ongoing support from the National Institutes of Health.

The CC is now a partnership among several national and international universities and research institutions and includes three distinct breeding arms located in the United States, Israel, and Australia. All three populations are traceable to F1 mice generated by the eight founder strains at the Jackson Laboratory and sent in 2004 to (1) the Oak Ridge National Laboratory (ORNL, since relocated to the University of North Carolina [UNC] at Chapel Hill; Chesler et al. 2008), (2) the International Livestock Research Institute in Kenya (since relocated to Tel Aviv University in Israel; Iraqi et al. 2008), and (3) the Western Australia Institute for Medical Research in Perth (Morahan et al. 2008) (Figure 2). Because we are more actively involved in the generation of the US CC population, the remainder of this article focuses on the US arm, but most of the observations and conclusions also apply to the Israeli and Australian arms.

Production of the CC Population

In contrast to traditional RI lines, the CC was derived from eight genetically diverse inbred strains—A/J, C57BL/6J, 129S1/SvImJ, NOD/ShiLtJ, NZO/HILtJ, CAST/EiJ, PWK/Ph, and WSB/EiJ (Chesler et al. 2008)—to result in high levels of variation distributed uniformly throughout the genome (Roberts et al. 2007; Yang et al. 2007). This was achieved mainly by including wild-derived strains from three mouse subspecies among the eight founder strains (Figure 1C). Unlike most RI panels, no genomic region is IBD in all CC parental strains (Yang et al. 2010). The presence of more than two distinct haplotypes can be easily visualized by determining the subspecific origin of the genome of the eight founders (Figure 1D).

We recently conducted a proof-of-principle experiment using incompletely inbred CC lines that were genotyped using the high-density Mouse Diversity Genotyping Array (MDA; Yang et al. 2009). Genomewide analysis of single nucleotide polymorphism (SNP) data in the incipient strains revealed the three desired characteristics of a genetic reference population: high genetic diversity, balanced allele frequencies, and dense, evenly distributed recombination sites (Aylor et al., in press).

The eight founder inbred strains were bred using a combinatorial funnel design to yield a large number of genetically independent RI lines (Figure 1). Reciprocal matings initially involved all combinations of the eight founder strains: F1 hybrids were intercrossed, a balanced subset of nonoverlapping 4- and 8-way progeny were subsequently mated, and the final 8-way animals were sibling mated to produce incipient RI lines. The aim of this breeding scheme was to generate RI lines with many random allele combinations via recombination and chromosomal assortment.

A CC breeding database (CCDB) was developed to track each animal and to determine, using a random generator, which mice should become breeders (Chesler et al. 2008). Strict criteria were established to meet these randomization goals: breeders were assigned priorities and mating pairs for generation X + 1 were prioritized at weaning. Second- or third-priority mating pairs were used only if their second litter was born before a litter from the first-priority pair. This
schema proved to be overly strict and in hindsight may have caused the loss of more lines than if the first litter from any mating was used (Chesler et al. 2008). Once the preliminary studies demonstrated that all eight strains were represented in lines at intermediate stages of inbreeding, the stringency of randomization was relaxed to allow for the selection of more fecund animals. (For details on recombination, inbreeding rates, and statistical power in the CC, Broman 2005; Valdar et al. 2006.)

Figure 1 Generation and genetic diversity in the Collaborative Cross (CC). (A) Variable coat color: litter from a CC line at an early generation of inbreeding (G2:F4) in which multiple coat color alleles are still segregating. (B) Breeding scheme. The colors and letters represent the eight founder strains as follows: A/J, A/yellow; C57BL/6J, B/grey; 129S1/SvImJ, C/pink; NOD/LtJ, D/dark blue; NZO/H1LtJ, E/light blue; CAST/EiJ, F/green; PWK/PhJ, G/red; and WSB/EiJ, H/purple (these colors and letters are also used in the website describing the status and availability of CC lines; http://csbio.unc.edu/CCstatus/). (C) Phylogenetic tree of the eight CC founder strains based on single nucleotide polymorphism (SNP) data for chromosome 11. Blue denotes M. m. domesticus origin; red, M. m. musculus origin; and green, M. m. castaneus origin. Strains in solid color boxes are wild-derived inbred strains. (D) Subspecific origin of the autosomes of the eight CC founder strains. Colors follow the same conventions as for (C).

Unique Characteristics and Special Considerations

Several of the CC founder strains have particular characteristics that affect their handling, lifespan, and reproduction. Although in this section we use the founder strains as examples, a recent proof-of-principle experiment indicates that all traits studied in the CC lines span—and typically expand—the founders’ range of phenotypic variation (Aylor et al., in press).
The inclusion of three wild-derived inbred strains (CAST/EiJ, PWK/PhJ, and WSB/EiJ) ensures high levels and uniform distribution of genetic variation across the genome. In contrast with classical inbred strains (derived from pet mice bred for docility and coat color), wild-derived strains originate from wild-caught mice. Although some form of selection for docility may have occurred, the behavior of these mice is strikingly different from that of most laboratory strains and their handling requires special precautions to reduce the possibility of escape, which can be reduced by handling cages with the least amount of disturbance and by cautiously removing cage tops. Cages should be placed in hoods fitted with shields around three sides so that the only opening is occupied by the handler. Alternatively, it is useful to place the cage in a large sanitized bin while opening so that any escapees have a third level of containment.

Two of the founder lines, NOD/ShiLtJ and NZO/HILtJ, become diabetic by 4 to 6 months of age and need special attention as both strains exhibit excessive water consumption and urination due to the diabetes. The extra water consumption, combined with the large litter sizes typical of NOD/ShiLtJ mice, results in the need for much more frequent cage changes than for nondiabetic mice. In fact, changes as frequent as every 2 to 3 days are not unusual compared with the normal changing schedule of 14 days for ventilated cages. Moreover, lifespan is usually shorter in diabetic animals—NOD/ShiLtJ mice may live only 6 to 8 months under normal food and water conditions. NOD/ShiLtJ females have even
shorter lifespans and are retired from breeding at a much higher rate than any other CC founder strain. These observations may apply to CC lines susceptible to diabetes.

NZO/HILtJ mice become obese as young adults and have notoriously poor reproductive performance, so they need to be mated immediately upon weaning. Once a male has sired a litter, he is usually able to father 3 or 4 more litters; but if he does not sire his first litter before 10 weeks of age, it is extremely unlikely he will father any offspring. The NZO/HILtJ breeding colony typically requires twice as many cages as any other inbred strain.

Characteristics of the founder strains have been noted in the CC lines and will likely persist, and even increase in severity, in some lines.

**Extinction**

The extinction rate in the US CC population is greater than 50% (Chesler et al. 2008), possibly because of the presence of incompatible combinations of alleles originating in different subspecies (Figure 1C).

The US population was started in two phases separated by 2 years. The extinction rate in the first phase was significantly higher than in the second; for example, the cumulative extinction rates at the G2:F2, G2:F3, and G2:F4 generations of inbreeding (when it is possible to make fair comparisons between phases 1 and 2) are statistically different ($p < 0.01$). The improvement in extinction rates reflects breeding changes implemented in 2007 to increase the likelihood of survival of phase 1 lines and to minimize the loss of phase 2 lines (Chesler et al. 2008). The changes in breeding involved avoiding combinations of nonproductive F1s, increasing the number of breeding pairs per generation in later generations, and decreasing the time before using lower-priority mating pairs when the first-priority pair was nonproductive. These changes had minimal impact on the randomization inherent to the CC breeding scheme. However, the extinction rate and the relatively small average litter size of the most advanced CC lines highlight the importance of maintaining large colony sizes and cryopreservation of stock to avoid reduction in the CC population.

**Housing**

Breeding to generate G2 mice started in 2004 at the ORNL William L. and Liane B. Russell Vivarium, a state-of-the-art barrier facility completed just before the initiation of the CC project and ideally suited to support the large-scale breeding program required. Mice were housed in Thoren ventilated racks with duplex cages, which allowed for higher density of breeding pairs without compromising the space needed per animal (each side of a duplex cage can accommodate three adults or a breeding pair and a litter). Each CC line was allotted four breeding pens in the early generations: one from generation X and three from generation $X + 1$. Keeping one box from generation X as a known fertile pen and backup helped to reduce the loss of lines. The ORNL vivarium standard operating procedure (SOP) called for the use of autoclaved cages, irradiated food, ALPHA-dri bedding, and automated hyperchlorinated water. Each cage also had a nestlet for nesting material. The number of animal care staff was limited to reduce disturbance and staff entered the facility through an air shower after donning full-body personal protection equipment.

In late 2008, the Department of Energy determined that the recently opened ORNL vivarium should be closed and in March 2009 the CC was transferred to UNC. Investigators at UNC had been involved in the development of the CC resource since its inception and had secured NIH support to complete the breeding of the US CC population, making UNC a logical place for relocation. The relocation schedule was a complex undertaking. Animals were transported once a week for 8 weeks by ORNL staff in a dedicated mouse transport van that carried 250 to 350 cages per trip (breeding boxes with pups less than 7 days old were not transported).

As a precaution, at least one breeding cage from each CC line remained at ORNL during colony transportation. Only about 10 animals out of the thousands moved died due to the move.

The CC population was housed in a temporary location in Research Triangle Park (RTP) until completion of a new vivarium in the UNC Genetics Medicine building. The RTP SOP called for the mice to be housed in ventilated, hot-washed Techniplast cages (the equipment available). Food delivery followed manufacturer recommendations and tap water was used to fill water bottles. The bedding used was bed-o’cobs® and a nestlet was placed in each breeding cage. We originally included a PVC pipe as well, but quickly discontinued its use after the deaths of some pups when the pipe rolled onto them during routine checks. The animals otherwise adapted quickly to the UNC conditions and exhibited no adverse effects on breeding or health status in the year they resided at the RTP location.

A second move, from RTP to the UNC Genetic Medicine vivarium, took place in February 2010 in five large shipments, two on each of 2 days in one week and the final shipment a week later to move cages housing pups that were less than 7 days old at the time of the first shipment. Care was taken while packing and there was no evidence of loss of animals during or after transport. However, after both moves it took about 4 weeks before litter production resumed normal levels.

The UNC housing conditions are the same as at RTP and, as per SOPs at all three facilities, the CC mice are housed in ventilated racks and the cages changed every 2 weeks. Based on our discussions with the UNC Institutional Animal Care and Use Committee (IACUC) and Division of Laboratory Animal Medicine, a revision to the UNC SOP has increased litter survival by calling for a 24-hour delay in cage changes after the birth of a litter.

Animals are identified by ear notches and/or cage card notations (eventually all weanlings will be identified by ear notching). Numbering is usually consecutive, in order of weaning starting with #1 in a line. The entire litter (including
weanlings that are not needed for breeding or saved for experiments) is recorded on the mother’s breeding card. Weanlings are recorded both in a separate ledger for each CC line and in the CCDB.

**Breeding Considerations**

**Weaning**

Litters are typically weaned at 3 weeks of age, which is a frequency sufficient to prevent the birth of multiple litters in breeding pens. But in some CC lines the pups are not sufficiently mature to be weaned by that age. At UNC the SOP is to wean at 10 grams. But NZO/HILtJ mice reach 10 g as early as 12 to 14 days of age, whereas the wild-derived strains (CAST/EiJ, PWK/PhJ, and WSB/EiJ) do not weigh 10 g until 24 to 28 days of age. Thus weaning by weight is not feasible because NZO/HILtJ pups would be weaned as early as 12 days of age, when few pups would survive, and wild-derived strains might not be weaned until 28 days of age despite their ability to survive independently at 21 days. Technicians working with the CC are trained to determine when animals can survive and the IACUC has approved exceptions for weaning by age and robustness rather than weight.

A second IACUC-approved exception allows for two litters from one dam to be in the same pen concurrently. Cages are checked daily by the animal care staff for overcrowding. A newborn litter in a mating pair cage with a weaning-age litter is not unusual in highly productive colonies and is promptly corrected when noted. With the use of a log of cages with higher than acceptable densities for the CC breeding program at UNC, the rate of deaths due to trampling or neglect by the dam is determined to be less than 5%.

**Mating**

The practice since the start of the CC project has been to set up mating pairs at weaning in order to increase the speed of inbreeding and to accurately track time to fertility in the different lines. Animals are usually pair mated, but, depending on availability, some are harem mated (one male mated to one to four females in a single pen). Pregnant harem-mated females are removed to separate pens to litter, then returned to the mating pen after the pups are weaned.

A third IACUC exception allows the sire to stay in the cage with the dam even in the presence of large litters (>11 pups), enabling the pair to continue mating. This exception applies mainly to early generations of the CC in which hybrid vigor prevails. As inbreeding progresses the litter sizes drop and usually by the G2:F4 generation the average litter sizes are four to eight. There are three main justifications for this exception:

- As discussed, three of the CC founder strains are wild-derived inbred lines and the behaviors of many CC lines resemble those of these wild stocks. Each time a cage is opened there is a high risk that the dam, sire, and/or pups will jump from the cage and be lost. Separating the male from the female before the litter is born would increase the number of times the cage needs to be opened, thus increasing the chance of losing the sire during a cage opening, and would decrease the likelihood of postpartum matings. These outcomes would be in direct conflict with two of the goals of the project: rapid turnover of generations and production of a high number of animals.
- Because breeding is a major goal at this stage, there are about 2,500 breeding cages. If the males are removed from each of the approximately 1,000 breeding cages at any given time, the cage census would dramatically increase. Many of the CC lines show signs of fertility loss, so every breeder could mean the difference of survival of the line.
- The average litter size in the founder inbred lines and in the CC lines themselves ranges from 2 to 12 and litters as large as 14 to 16 are not unusual. The UNC SOP calls for a maximum cage density of a breeding pair plus up to six pups or one adult plus 11 pups after the pups reach 14 days of age. Larger litters are allowed with a breeding pair until 14 days of age. In other projects, these goals are achievable by genotyping pups early or culling one sex if specific classes of animals are not needed. For the CC every animal is important and needs to be kept alive.

**Marker-Assisted Inbreeding**

To accelerate the completion of the CC project and to minimize the loss of lines, the number of breeding cages has been expanded for later generations of inbreeding. In addition, genotype-based approaches are being used to characterize the genetic makeup and ensure the genetic integrity of each line, and to accelerate inbreeding. To accomplish this, the breeding pairs that are obligate ancestors of all extant mice in a CC line are genotyped using the MDA (Yang et al. 2009). Typically these breeding pairs then have to be euthanized to enable preparation of the high-molecular-weight DNA from their spleens that is necessary for the high-density SNP array. Areas of residual heterozygosity are determined and a second, much smaller panel of SNPs is used to track these regions. For the lower-density array DNA is prepared using tail or ear clips from pups just prior to or at weaning. From animals genotyped using the low-density arrays breeding pairs are selected to be most conducive to rapid inbreeding rather than randomized mating.

Simulations indicate that it is possible to significantly reduce the number of generations to full inbred status with minimal impact on the genetic architecture of the CC population (Catherine Welsh, Fernando Pardo-Manuel de Villena, and Leonard McMillan, unpublished results). High-density genotypes and haplotype reconstructions for the lines undergoing marker-assisted inbreeding (MAI) are available (http://csbio.unc.edu/CCstatus/).

A set of CC lines at the most advanced stages of inbreeding was recently imported from the Israeli CC population and most were producing litters within 6 weeks of their arrival at
UNC. These lines will also undergo MAI similar to the UNC population.

Once a line is determined to be inbred, it will be re-derived into a barrier facility at UNC’s Mutant Mouse Regional Resource Center so that it is available to any facility in the world. As of November 2010, we have one inbred line and expect to complete 50 lines by mid-2011 and then 50 more every 6 months to a total of at least 200. Each confirmed inbred line will undergo rederivation, which should make the lines available 6 to 12 months after they are confirmed inbred.

The CC and Other Mouse Resources

Notwithstanding its unique characteristics, the CC is intimately connected with other major mouse resources (Figure 2). It is derived from extant inbred strains that are widely used by the mouse genetics community and represented in other genetic reference populations such as traditional RI panels, the MPP, the Hybrid Mouse Diversity Panel (an updated and improved phenome panel; Bennett et al. 2010), and strains commonly used to produce genetically engineered models (GEM; Guan et al. 2010). These relationships can be exploited to integrate and leverage existing data to inform new experiments and to interpret previous experiments.

In addition, CC lines were the starting material for the Diversity Outcross, a heterogeneous stock designed to achieve very high mapping resolution that was seeded with mice from intermediate generations of the US CC population and is being bred at the Jackson Laboratory (GAC, unpublished results) (Figure 2). Finally, CC lines can be intercrossed among themselves or backcrossed to any other line or GEM to generate recombinant intercross (RIX) and recombinant backcross (RIB) panels. A RIX population with a unique loop design will be generated as part of an NIH-funded Center of Excellence in Genome Sciences (Figure 2).

Conclusion

The Collaborative Cross is a community resource that will eventually integrate data from a wide variety of biomedical disciplines. Its benefits to both individual users and the whole biomedical research community will be proportional to its integration of the many disparate datasets generated by individual researchers into a single public database. It is thus essential to develop and implement bioinformatic tools that will allow both geneticists and nongeneticists to take full advantage of this unique resource. Ideally these tools will be complemented with training programs that maximize access and enhance the efficient use and maintenance of the CC population.

Acknowledgments

This work was supported by grants from the National Institutes of Health (U01CA105417, U01CA134240, and P50MH090338), Ellison Medical Foundation (AG-IA-0202-05), and US Department of Energy (DE-AC05-00OR22725 with UT-Battelle, LLC). Essential support was also provided by the Dean of the UNC School of Medicine, the Lineberger Comprehensive Cancer Center at UNC, and the University Cancer Research Fund from the state of North Carolina. We thank Leonard McMillan for unpublished results and Figure 1B, and Jennifer Shockley and Stephanie Hansen for Figure 1A.

References


Threadgill DW. 2006. Meeting report for the 4th annual Complex Trait Consortium meeting: From QTLs to systems genetics. Mamm Genome 17:2-4.