Genet. Res., Camb. (2002), **80**, *pp.* 1–6. © 2002 Cambridge University Press
DOI: 10.1017/S0016672302005839 Printed in the United Kingdom

Review

Enabling Population and Quantitative Genomics

GREG GIBSON* AND TRUDY F. C. MACKAY
[1] *Department of Genetics, Gardner Hall, North Carolina State University, Raleigh, NC 27695-7614, USA*

(*Received 10 July 1988*)

Summary

Dissection of quantitative traits to the nucleotide level requires phenotypic and genotypic analysis of traits on a genome scale. Here we discuss the set of community-wide genetic and molecular resources, including panels of specific types of inbred lines and high density resequencing and SNP detection, that will facilitate such studies.

1. Background

The field of evolutionary quantitative genetics currently has a somewhat bipolar character. On the one hand, there is great excitement derived from advances in genotyping technologies and statistical methodology that have opened up research possibilities that could only be dreamt about a decade ago (Mackay, 2001; Doerge, 2002). It is now routine to map quantitative trait loci (QTL) to genomic intervals bounded by molecular markers and describe the broad attributes of genetic architecture of phenotypes in just about any organism, from stickleback fish to monkeyflowers, and from yeast to honeybees (Peichel *et al*, 2001; Bradshaw *et al*, 1998; Steinmetz *et al*, , 2002; Page *et al*, 2000). On the other hand, it is depressing that there are certain problems, notably the dissection of QTL to the genic and eventually nucleotide level, quantification of epistasis, and evaluation of genotype by environment interactions in natural populations, that remain refractive to dissection. The point of this minireview is to discuss the community-wide resources that will have to be developed to ensure that quantitative genetic analysis of model organisms continues both to drive advances in evolutionary theory, and to contribute to the development of approaches to the understanding of complex, multifactorial traits.

High-throughput 'genomic' approaches are having an impact on at least three major areas of evolutionary genetics, namely molecular systematics, comparative genomics, and molecular phenotyping. The US National Science Foundation has announced its intention to support the first two of these endeavors at a structural level, through workshops and program solicitations calling for 'tree of life' proposals for collaborative and interdisciplinary research to determine molecular phylogenies of the 1·7 million described species; as well as to construct up to 100 BAC libraries as the foundation for generation of physical maps across a broad sampling of taxa. Just as importantly, ecological genomics initiatives are seeing the application of these and other resources to address basic and applied biological issues from conservation genetics to mechanisms of speciation. Also, a handful of studies have begun to apply classical quantitative genetic approaches to gene expression profiling of differences among natural isolates within species and in different environments, as well as between closely-related species (Gibson, 2001). We can expect proteomic and metabolic profiling to arrive in the near future as well, dramatically increasing the tools available to fill in the events that relate genetic to phenotypic variation.

It is nevertheless essential to keep asking what other resources will be required to address the overarching evolutionary questions (Barton & Keightley, 2002). What will it take to understand the mechanisms that maintain genetic variation, to appreciate the impact of population structure on genotype and phenotype, to quantify genotype-environment interactions, and to describe the relationship between intra-specific polymorphism and inter-specific divergence? While broad sampling of organismal diversity is a welcome and

* Corresponding author: email: ggibson@unity.ncsu.edu. Tel: (919)513 2512 Fax: (919)515 3355

essential component of population and quantitative genomics, intensive phenotypic sampling of the existing model organisms will continue to be the key to the development of new perspectives on many of the classical issues in evolutionary genetics.

2. Quantitative Genomics of Model Organisms

In the past year, significant breakthroughs in the mapping of QTL to single genes have been achieved in Arabidopsis and maize, taking advantage of the well-characterized genetic resources (Remington *et al*., 2001). Several sets of recombinant inbred lines of Arabidopsis have been genotyped at high density and have proven to be a valuable resource for quantitative genetics laboratories around the world. Combining fine structure recombination mapping with transformation, it has proven possible to positionally clone loci affecting flowering time in *Arabidopsis* (El-Din El-Assal *et al*., 2001) and *Brassica* (Osterberg *et al*., 2002), while similar strategies in tomato have identified a locus responsible for fruit size (Frary *et al*., 2000). A quite different approach to QTL identification, linkage disequilibrium mapping with candidate genes, has been pursued in maize. The common accessions used by maize breeders have a complex genealogy that complicates the interpretation of genotype-phenotype associations (Matsuoka *et al*., 2002), but statistical deconvolution of the effect of population structure may be possible (Pritchard *et al*., 2000; Thornsberry *et al*., 2001) and is clearly also enabled by the existence of genetically characterized seed-stock.

The most likely animal models for quantitative genomics continue to be the mouse (*Mus musculus*) and fruitfly (*Drosophila melanogaster*), and possibly the soil nematode *Caenorhabditis elegans*. Several mouse recombinant inbred lines genotyped at high density exist and have proven utility for the mapping of many complex traits (Williams *et al*., 2001). For some traits it has been suggested that QTL can be mapped in a relatively small sample of common laboratory mouse strains, already genotyped at high density, taking advantage of the fortuitously ideal level of recombination that has been achieved since the stocks diverged from a small set of founders (Grupe *et al*., 2001). Given the completion of the mouse and human genome sequences and assumed similarity of physiological and behavioral mechanisms in these mammals, there is every reason to believe that quantitative genetic analysis of the mouse will experience a renaissance in the next decade, with only cost as the major limitation. *C. elegans* is relatively underutilized as a quantitative genetic model (Ayyadevara *et al*., 2001; Shook & Johnson, 1999) despite advantages such as the simplicity of stock maintenance, inbreeding by self-fertilization, considerable variation for a wide range of traits, and levels of polymorphism similar to those in humans (Koch *et al*., 2001), which may support linkage disequilibrium mapping. Increased attention to the ecology of this organism, and expanded sampling from around the world are urgently needed.

3. Population Genomics of *Drosophila melanogaster*

The remainder of this minireview discusses enablement of population and quantitative genomics of Drosophila, and draws on discussions that took place at a workshop on this topic that was organized by T. F. C. M. and was held at the 43rd Annual Drosophila Research Conference in San Diego, California, April 10, 2002.

An ambitious proposal to sequence the complete genomes of 50 strains of *D. melanogaster* using oligonucleotide-based resequencing was presented by Charles Langley (University of California, Davis) and Michael Zwick (Johns Hopkins University). The basis of this method is the hybridization of amplified genomic DNA to perfect match 25mer oligonucleotides arrayed at high density on silicon wafers (Wang *et al*., 1998). Oligonucleotides are synthesized *in situ* on the wafers with all four possible bases substituted at the central position of each oligonucleotide in adjacent features. The complete genome sequence of the *y*; *cn bw sp* strain of *D. melanogaster* (Adams *et al*., 2000) would be used to program the synthesis of the complete set of oligonucleotides tiling along each chromosome, one nucleotide at a time. The total cost of this endeavor would be approximately $27 million U. S. dollars, or $0·005/bp. Given the interest in developing such technology for resequencing of the human genome, it is likely that at least a portion of the fly genome will be resequenced in the near future.

Combined with the impending complete shotgun sequence of the divergent *D. pseudoobscura* genome, whole genome sequence information of this magnitude and caliber would provide an unprecedented resource for molecular evolutionary analysis. For example, several groups have begun to explore the genome-wide distribution of test-statistics for inference of selection acting on individual loci (Bustamante *et al*., 2002; Smith & Eyre-Walker, 2002; Fay *et al*., 2002), and it is likely that novel patterns of variation will be uncovered from analysis of thousands of loci and contiguous megabase stretches of DNA. These data would enable quantification of heterogeneity across the genome in rates and patterns of mutation, recombination and gene conversion.

Further, as pointed out by Andrew Clark (Cornell University), genetic dissection of Drosophila quantitative traits should serve as a model system for testing ideas to map human complex diseases. The

https://doi.org/10.1017/S0016672302005839 Published online by Cambridge University Press

current scenario for proceeding from a QTL to identifying single genes contributing to complex traits in humans is to utilize association mapping (Ardlie *et al*, 2002). Underlying this strategy are the assumptions that common diseases are caused by common allelic variants, and that single nucleotide polymorphisms (SNPs) occur in large blocks of linkage disequilibrium (LD). Thus, one should be able to narrow down the interval containing the causal disease gene(s) using only a few SNPs per haplotype block – an important practical consideration given the current cost of SNP genotyping and the large sample sizes required for such studies (see below). Molecular polymorphism data from 50 whole Drosophila homozygous genomes – i.e., with known linkage phase – will provide an opportunity to rigorously quantify the LD landscape and haplotype blocks, and test the concept of haplotype mapping by sampling haplotypes from complete polymorphism data.

A sample of 50 genomes would supply a list of most of the common molecular variants segregating in *D. melanogaster*. However, association studies based on 50 whole genomes will not have sufficient power to detect genotypic effects of the magnitude expected from QTL alleles. Simulation studies indicate that samples of at least 500 individuals are required to attain the statistical power to detect associations at SNPs in candidate genes that affect as little as 5% of the genetic component of variance for a trait (Long & Langley, 1999); samples of this magnitude are now standard in human genetics. However, this consideration does not take into account the power of Drosophila as a model system. With 50 reference genomes, it will be possible for any investigator interested in a particular candidate gene to introgress the genomic fragments including the 50 candidate gene alleles into a common inbred background and determine trait phenotypes on multiple individuals of each of these near-isoallelic lines, and/or on wild alleles as heterozygotes against mutant and wild type alleles of the candidate gene. The reduction in genetic heterogeneity and accurate assessment of genotypic values, combined with quantitative complementation testing, enables the detection of relatively small effects with only 50 alleles (Long *et al*., 1998; Lyman *et al*., 1999; Robin *et al*., 2002). Further, individual investigators studying particular candidate genes could utilize the knowledge of common variants in the design of association studies with larger samples.

4. Quantitative Genomics of *Drosophila melanogaster*

The idea of 50 reference Drosophila genomes naturally raises the issue of which strains to develop as resources for quantitative genomics. Several proposals were discussed by Sergey Nuzhdin (University of California, Davis) and Andrew Clark, all of which recognize two unique problems facing Drosophilists. These are the constraint on stock-keeping that arises from the need to maintain live cultures that must be turned every few weeks, at considerable expense, and the observation that mutation accumulation in laboratory stocks quickly compromises their quantitative genetic integrity (Wayne & Mackay, 1998). The community would thus benefit greatly from the establishment of a core set of genetic resources that could be phenotyped in parallel in a window of a three to five years, genotyping of which would support investigation of a comprehensive set of research questions.

There are three classes of genetic resource that are likely to be most useful, namely inbred lines, recombinant inbred lines, and mutation lines. These might be supplemented by, and/or derived from, samples of isofemale lines or population cages designed to represent diversity from particular habitats and localities and along clinal gradients. A practical limitation on the use of population cages for artificial selection experiments is that they should be purged of naturally occurring inversions to limit the bias inversions have on the trajectory of selection, as well as their negative impact on efforts to map QTL.

The two major classes of lines that are currently in use for QTL mapping and candidate gene association studies are inbred lines derived by at least 20 generations of full-sib mating, and chromosome substitution lines (CSLs) constructed using balancer crosses. The main advantage CSLs provide is that standardization of the genetic background for two of the three major chromosomes effectively enhances the proportion of genetic variation attributed by each QTL on the substituted chromosome, hence greatly increasing statistical power. The drawbacks are that the balancers can break down, many extracted chromosomes carry lethals so can only be studied in heterozygous condition, and three times as many stocks must be maintained than with inbred lines to capture variation throughout the genome. There are then good reasons for generating both types of resource. Before doing so, it will also be important to reach broad agreement on the population sampling strategy: is it better to sample 500 lines each from two localities, or 50 lines each from 20 localities along a cline?

Recombinant Inbred Lines (RIL) have proven invaluable to fly and plant geneticists alike for mapping of QTL, since they only need to be genotyped once and since the genetic correlation structure among traits is directly determined. One set of approximately 100 RIL constructed by E. G. Pasyukova and S. V. Nuzhdin (Nuzhdin *et al*., 1997) has been used to map QTL affecting over a dozen sets of traits in several laboratories. The combination of recombination map-

https://doi.org/10.1017/S0016672302005839 Published online by Cambridge University Press

ping with quantitative complementation to deficiencies and mutations of positional candidate genes can identify actual genes of previously unknown function that fail to complement QTL alleles (Fanara *et al.*, 2002). However, success of this approach is currently limited by the availability of deficiencies uncovering the entire genome (there are gaps in the deficiency map, such that only 70–80% of the genome is uncovered), and the availability of mutations in all genes.

Several complementary strategies were discussed to overcome these drawbacks. The limit of resolution of recombination mapping of QTL is set by the sample size, for two reasons. First, the number of recombinants in any given interval scales with the sample size, thus narrowing the intervals to which QTLs map. Second, QTLs with smaller effects are detectable as the sample size increases. Thus, with 1000 hyper-recombinant inbred lines derived from two strains, genotyped for 14,000 markers, one could potentially resolve QTLs to single genes. A major issue to be faced before embarking on the construction of further sets of RIL is the identification of appropriate parental lines that segregate for a representative set of traits. Two strategies were proposed to counter the narrow genetic base of the hyper-recombinant inbred lines. One would be to generate controlled pedigree crosses of 200 inbred lines derived from nature, and extract chromosomes from these as a permanent resource for genotyping and phenotyping. The other would be to derive from each of four inbred strains a set of 500 near-isoallelic introgression lines. Individual introgression lines would contain a small introgressed fragment from one of the inbred strains, in the common inbred background of a fifth, unrelated strain. The whole collection of lines would span the genome, enabling fine-mapping of QTL (Eshed & Zamir, 1995) and evaluation of epistasis from crosses among lines (Eshed & Zamir, 1996). Finally, availability of a deficiency kit spanning the entire genome, but in a common inbred background, would greatly improve the power of quantitative deficiency mapping. Most existing collections of deficiency lines have been generated in diverse genetic backgrounds, so much work needs to be done before this type of resource is optimized for quantitative genetics.

The above discussion focused on the problem of identifying genes impacting segregating variation for complex traits in nature. However, many Drosophila genes are predicted, with no known function, and pleiotropic effects on quantitative traits have yet to be characterized for many loci that play critical roles in development and metabolism. This is a consequence of the bias of past mutation screens to detecting mutations with large, qualitative effects and the focus on generating null mutations for which pre-adult lethality is the phenotype of interest (Mackay, 2001). To identify candidate genes for quantitative traits, we need to construct lines each containing a single mutation in an isogenic background, and measure the morphological, behavioral and life-history traits of interest in replicate, to identify mutations with quantitative effects. *P*-element insertion lines are ideal for this purpose, and have been used successfully to isolate numerous mutations in novel genes involved in olfaction (Anholt *et al.*, 1996) and sensory bristle number (Lyman *et al.*, 1996). Thousands of single *P*-element insert lines in an isogenic background would be an ideal resource for identification of candidate genes affecting complex traits.

Allied with the identification of candidate quantitative trait loci is the problem of defining the distribution and frequency of quantitative trait nucleotide effects attributable to individual SNPs. Procedures will also have to be developed that will allow rapid, cost-effective, and reliable genotyping of very large numbers of SNPs from hundreds if not thousands of strains. Resequencing of several hundred kilobases encompassing a set of agreed-upon candidate genes in 1000 different wild-type strains that would be phenotyped simultaneously for a range of traits is a further strategy that can be considered as a direct means of performing associations studies that would complement the more hierarchical QTL identification schemes discussed above.

5. Conclusion

Construction of genomic-scale resources will necessitate a collaborative effort from the whole community, as the effort and expense are well beyond the scope of an individual research award. There was considerable excitement and support for continued discussion and planning among the participants of the workshop. In addition to the afore-mentioned issues of strain selection, many thousands of lines would have to be phenotyped for a multiplicity of complex traits, including tissue-specific expression profiles, in multiple environments; and genotyped for tens of thousands of SNP markers. Given the rapid pace of technology development, we can anticipate more sensitive and comprehensive methods of quantifying whole genome transcript levels than the currently available cDNA microarray and Affymetrix oligonucleotide array platforms; as well as the development of rapid, accurate and cost-effective methods for SNP genotyping. It will also be important to establish a widely available and user-friendly database to integrate the phenotypic and genotypic data for each strain. The Drosophila population and quantitative genomics community has the unique expertise to take a leadership role in this initiative.

In conclusion, the molecular dissection of complex

traits at the nucleotide level will require the development of genetic resources for model organisms that are more extensive than those currently in use. There are benefits to be made from economies of scale, from collaboration that arises out of design and adoption of shared resources, and from combined analysis of a common set of lines. Genome-scale quantitative genetics will draw on high volume genotyping, phenotyping, and powerful statistical approaches, but will be facilitated by integrated strain construction and experimental design.

We thank all of the participants in the Workshop at the San Diego Fly Meeting, in particular the other presenters: Chuck Langley, Mike Zwick, Sergey Nuzhdin, and Andy Clark. Research in our laboratories is supported in part by a Program Project Grant from the NIH (P01-GM45344).

References

Adams M. D., *et al*. (2000). The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195.

Anholt, R. R., Lyman, R. F. & Mackay, T. F. C. (1996). Effects of single *P*-element insertions on olfactory behavior in *Drosophila melanogaster*. *Genetics* **143**, 293–301

Ardlie, K. G., Kruglyak, L. & Seielstad, M. (2002). Patterns of linkage disequilibrium in the human genome. *Nature Reviews Genetics* **3**, 299–309.

Ayyadevara, S., Ayyadevara, R., Hou, S., Thaden, J. & Reis, R. J. (2001). Genetic mapping of quantitative trait loci governing longevity of *Caenorhabditis elegans* in recombinant-inbred progeny of a Bergerac-BO x RC301 interstrain cross. *Genetics* **157**, 655–y66

Barton, N. H. & Keightley, P. D. (2002). Understanding quantitative genetic variation. *Nature Reviews Genetics* **3**, 11–21

Bradshaw, H. D., Jr, Otto, K., Frewen, B., McKay, J. & Schemske, D. W. (1998). Quantitative trait loci affecting differences in floral morphology between two species of monkeyflower (*Mimulus*). *Genetics* **149**, 367–382

Bustamante, C. D., Nielsen, R., Sawyer, S., Olsen, K., Purugganan, M. D & Hartl, D. L. (2002). The cost of inbreeding in *Arabidopsis*. *Nature* **416**, 531–534

Doerge, R. W. (2002). Mapping and analysis of quantitative trait loci in experimental populations. *Nature Reviews Genetics* **3**, 43–52

El-Din El-Assal, S., Alonso-Blanco, C., Peeters, A., Raz, V. & Koornneef, M. (2001). A QTL for flowering time in Arabidopsis reveals a novel allele of CRY2. *Nature Genetics* **29**, 435-440

Eshed, Y. & Zamir, D. (1995). An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine-mapping of yield-associated QTL. *Genetics* **141**, 1147–1162.

Eshed, Y. & Zamir, D. (1996). Less-than-additive interactions of quantitative trait loci in tomato. *Genetics* **143**, 1807–1817.

Fanara, J. J., Robinson, K. O., Rollmann, S., Anholt, R. R. H. Mackay, T. F. C. (2002). *Vanaso* is a quantitative trait locus for *Drosophila* olfactory behavior. *Genetics*, submitted.

Fay, J. C., Wyckoff, G. & Wu, C.-I. (2002). Testing the neutral theory of molecular evolution with genomic data from *Drosophila*.. *Nature* **415**, 1024–1026.

Frary, A., Nesbitt, T., Grandillo, S., Knaap, E., Cong, B., Liu, J., Meller, J., Elber, R., Alpert, K. & Tanksley, S. D. (2000). *fw*2·2: a quantitative trait locus key to the evolution of tomato fruit size. *Science* **289**, 85–88

Gibson, G. (2001). Microarrays in ecology and evolution: a preview. *Molecular Ecology* **11**, 17–24.

Grupe, A., Germer, S., Usuka, J., Aud, D., Belknap, J., Klein, R, Ahluwalia, M., Higuchi, R. & Peltz, G. (2001). In silico mapping of complex disease-related traits in mice. *Science* **292**, 1915–1918.

Koch, R., van Luenen, H., van der Horst, M., Thijssen, K. & Plasterk, R. H. (2001). Single nucleotide polymorphisms in wild isolates of *Caenorhabditis elegans*. *Genome Research* **10**, 1690–1696

Long, A. D. & Langley, C. H. (1999). The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. *Genome Research* **9**, 720–731

Long, A. D., Lyman, R. F., Langley, C. H. & Mackay, T. F. C. (1998). Two sites in the *Delta* gene region contribute to naturally occurring variation in bristle number in *Drosophila melanogaster*. *Genetics* **149**, 999–1017

Lyman, R. F., Lai, C & Mackay, T. F. C. (1999). Linkage disequilibrium mapping of molecular polymorphisms at the *scabrous* locus associated with naturally occurring variation in bristle number in *Drosophila melanogaster*. *Genetical Research* **74**, 303–311

Lyman, R. F., Lawrence, F., Nuzhdin, S. & Mackay, T. F. C. (1996). Effects of single P-element insertions on bristle number and viability in *Drosophila melanogaster*. *Genetics* **143**, 277–292

Mackay, T. F. C. (2001). The genetic architecture of quantitative traits. *Annual Reviews of Genetics* **35**, 303–339

Mackay, T. F. C. & Lyman, R. F. (1998). Polygenic mutation in *Drosophila melanogaster*: genotype × environment interaction for spontaneous mutations affecting bristle number. *Genetica* **103**, 199–215

Matsuoka, Y., Vigouroux, Y., Goodman, M., Sanchez, G., Buckler, E. & Doebley, J. (2002). A single domestication for maize shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Sciences of the USA* **99**, 6080–6084

Nuzhdin, S. V., Pasyukova, E. G., Dilda, C., Zeng, Z.-B. & Mackay, T. F. C. (1997). Sex-specific quantitative trait loci affecting longevity in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the USA* **94**, 9734–9739

Osterberg, M. K., Shavorskaya, O., Lascoux, M. & Lagercrantz, U. (2002). Naturally occurring indel variation in the *Brassica nigra COL*1 gene is associated with variation in flowering time. *Genetics* **161**, 299–306

Page, R. E., Jr, Fondrk, M., Hunt, G., Guzman-Novoa, E., Humphries, M., Nguyen, K. & Greene, A. S. (2000). Genetic dissection of honeybee (*Apis mellifera L.*) foraging behavior. *Journal of Heredity* **91**, 474–479

Peichel, C. L., Nereng, K., Ohgi, K., Cole, B., Colosimo, P., Buerkle, C., Schluter, D. & Kingsley, D. M. (2001). The genetic architecture of divergence between threespine stickleback species. *Nature* **414**, 901–905

Pritchard, J. K., Stephens, M., Rosenberg, N. & Donnelly, P. (2000). Association mapping in structured populations. *American Journal of Human Genetics* **67**, 170–181

Remington, D. L., Ungerer, M. & Purugganan, M. D. (2001). Map-based cloning of quantitative trait loci: progress and prospects. *Genetical Research* **78**, 213–218

Robin, C., Lyman, R. F., Long, A. D. & Langley, C. H. & Mackay, T. F. C. (2002). *hairy*: A quantitative trait locus for Drosophila bristle number. *Genetics*, in press

Shook, D. R. & Johnson, T. E. (1999). Quantitative trait loci affecting survival and fertility-related traits in *Caenorhabditis elegans* show genotype-environment interactions, pleiotropy and epistasis. *Genetics* **153**, 1233–1243

Smith, N. G. C. & Eyre-Walker, A. (2002). Adaptive protein evolution in *Drosophila*. *Nature* **415**, 1022–1024

Steinmetz, L. M., Sinha, H., Richards, D., Spiegelman, J., Oefner, P., McCusker, J. & Davis, R. W. (2002). Dissecting the architecture of a quantitative trait locus in yeast. *Nature* **416**, 326–330

Thornsberry, J., Goodman, M., Doebley, J., Kresovich, S., Nielsen, D. & Buckler, E. S., III (2001). *Dwarf*8 polymorphisms associate with variation in flowering time. *Nature Genetics* **28**, 286–289

Wang, D. G., *et al*. (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280**, 1077–1082.

Wayne, M. L. & Mackay, T. F. C. (1998). Quantitative genetics of ovariole number in *Drosophila melanogaster*. II. Mutational variation and genotype-environment interaction. *Genetics* **148**, 101–210

Williams, R. W., Gu, J., Qi, S. & Lu, L. (2001). The genetic structure of recombinant inbred mice: high-resolution consensus maps for complex trait analysis. *Genome Biology* **2, 46**, 1–18