

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES

Issue: *Nutrition and Physical Activity in Aging, Obesity, and Cancer***Genomics of complex traits**James E. Womack,^{1,2} Hyun-Jun Jang,¹ and Mi Ok Lee¹¹Department of Agricultural Biotechnology, Seoul National University, Seoul, Korea. ²Department of Veterinary Pathobiology, Texas A & M University, College Station, TexasAddress for correspondence: James E. Womack, Department of Veterinary Pathobiology-Mail Stop 4467, Texas A&M University, College Station, TX 77843-4467. jwomack@cvm.tamu.edu

The analysis of complex genetic traits, including mapping and identification of causative genes, has long been an enigma of genetic biology, whether in the animal sciences or in medical sciences. Traits of agricultural interest and traits of medical interest are often under the influence of both environmental factors and multiple genes, each with modest contributions to the total variance in the trait. Although the number of known mutations underlying complex traits is still relatively small, advances in genomics have greatly enhanced traditional pathways to their analysis and gene mining. The candidate gene approach, linkage analysis, and association studies are all significantly more powerful with recent advances in genome mapping, sequencing, and analysis of individual variation. Avenues to gene discovery are discussed with emphasis on genome wide association studies (GWAS) and the use of single nucleotide polymorphisms (SNPs) as revealed by increasingly powerful commercially available microarrays.

Keywords: genomics; complex traits; animals, disease resistance; GWAS

Introduction

A phenotype with known or suspected genetic involvement that does not conform to classical Mendelian inheritance is generally described as a complex genetic trait. In truth, all genetic traits are complex because identical phenotypes among individuals carrying identical alleles at a given locus are rare or nonexistent, due to a number of causes including incomplete penetrance, genetic heterogeneity, and most often, a genetic background consisting of multiple genes modifying the phenotype with minor effects. Generally, however, we reserve the term *complex* for phenotypes that are influenced by multiple genes (polygenic) as well as environmental factors. Many of the traits important to animal agriculture fall into this class and are described quantitatively rather than qualitatively. Complex traits also include phenotypes that are discrete but are only expressed when the effects of multiple genes and/or environmental factors achieve a minimal threshold. These so-called threshold traits include many human diseases such as cancer and diabetes. Susceptibility to specific infectious diseases, a trait important

to both animal and human health, falls in this class as well.

Genetic influence has been demonstrated for thousands of traits in domestic animals and humans. The strength of the genetic component of a trait is usually expressed as a heritability estimate for quantitative traits including many economically important traits in animals, and as familial risk factors for threshold traits, including many complex diseases in humans. Although large numbers of genomic loci underlying complex traits have been mapped, very few specific mutations have been identified and proved to be causative.

Genome projects and genomic resources

In addition to providing a complete human genome sequence, along with cataloging of a wealth of naturally occurring genomic variation, the Human Genome Project spawned genome projects for a large number and wide variety of other organisms, including most of the common domestic animal species. Single nucleotide polymorphisms (SNPs) are the most common form of genetic variation in most species, and animal genomes appear to be replete with these, small insertions and deletions (indels), and copy number variants (CNVs). sbSNP, a database for short genetic variations, lists more than

Re-use of this article is permitted in accordance with the Terms and Conditions set out at http://wileyonlinelibrary.com/onlineopen#OnlineOpen_Terms

60 million entries for humans, 9.5 million for cattle, more than 3 million for dogs and chickens, 1 million for horses, and 0.5 million for pigs as of May 16, 2012 (<http://www.ncbi.nlm.nih.gov/projects/SNP/>). Among the many valuable scientific resources to come from these genome projects are microarrays which permit the screening of individuals for hundreds of thousands of genomic variants in a single assay. These so-called SNP chips and CNV arrays are commercially available for humans and the common domestic species. One chip for human variants, for example, contains markers for more than 900,000 SNPs and 900,000 CNVs. A high-density cattle SNP chip contains almost 800,000 markers.

QTLs, QTNs, and QTM

Identification of genes responsible for monogenic traits is becoming relatively straightforward with the resources available to today's scientists. GeneCards, a database of human disease genes (<http://www.genecards.org/cgi-bin/listdiseasescards.pl?type=full>) lists 5,600 genes that are associated with human diseases, most of them monogenic (listing as of May 16, 2012). Online Mendelian Inheritance of Man (<http://omim.org/>) contains information on all known Mendelian disorders and over 12,000 genes. Even in these substantial databases, only a few genes underlying complex human disease are known.

Considerable effort in domestic animal research has gone into mapping quantitative trait loci (QTLs), the chromosomal regions (locations) that contain genomic elements contributing to variation in a trait. These studies, initially using highly polymorphic microsatellites and more recently SNPs as markers, have been highly successful. The AnimalQTLdb (<http://www.animalgenome.org/cgi-bin/QTLdb/index>) cites several thousand QTLs for pigs, cattle, chickens, sheep, and rainbow trout as summarized in Table 1.

Despite the great success in mapping QTLs, only a handful of the actual mutations responsible for trait variation have been identified. While selective animal breeding can be practiced with phase-known linked markers, identification of the genes and their specific mutations should be the scientific objective of QTL studies. Only through knowledge of genetic and cellular mechanisms accounting for variation, can phenotypes be modulated via envi-

Table 1. Summary of quantitative trait loci in animals used in agriculture

Species	No. of QTLs	No. of publications	No. of traits
Pigs	6,818	290	585
Cattle	5,920	330	407
Chickens	3,162	158	270
Sheep	684	78	189
Rainbow trout	73	8	11

From Animal QTLdb (<http://www.animalgenome.org/cgi-bin/QTLdb/index>), release 17, April 20, 2012.

ronment, diet, vaccines, therapies, or perhaps genetic modification. While the term quantitative trait nucleotide (QTN) has been valuable in describing the nucleotide variation underlying a QTL,¹ Andersson² has introduced a more generally useful term in quantitative trait mutation (QTM), which accounts for the fact that the variation may be caused by a mutation larger than a single nucleotide.

Candidate genes

The most intuitive approach to identification of specific genes and mutations responsible for phenotypes is the candidate gene approach. Prior knowledge of the function of a gene, coupled with some understanding of the physiological basis of the trait, has been a historical first approach to the discovery of gene/trait relationships. This logic has been exploited effectively for a number of monogenic human disease genes as well as some important phenotypes in domestic animals. As pointed out by Georges,³ however, the candidate gene literature is replete with nonreproducible results. Georges³ makes a rational plea for higher standards in candidate gene analysis, including (a) more rigid significance thresholds, (b) thorough characterization of the haplotype diversity and structure in the study population, (c) replication of positive associations in an independent cohort, and (d) resequencing individuals from functionally distinct haplotypes to find causative SNPs. As is usually the case, Michel Georges' advice should not go unheeded.

Linkage studies and the positional candidate approach

As discussed above and demonstrated in Table 1, a large number of QTLs have been mapped in

domestic animal species, especially those of agroeconomic significance but only a few QTMs have been discovered. The *positional candidate approach* to gene discovery employs linkage mapping to map a trait, and then searching for most likely candidates among the genes known to lie in this region. While this has been effective in both human and animal genetics for monogenic traits, it has had limited success for QTLs. One reason is that the function of most genes is still unknown. Another is that mapping resolution is usually severely limited. A typical confidence interval in a QTL mapping study might be 20 or more centimorgans, corresponding to roughly 20 million base pairs and potentially 200 or more genes. As discussed by Georges,³ several factors limit mapping resolution. These are (a) marker density (which is no longer a problem in most species with the availability of SNP technology), (b) crossover density because only recombinants provide map information, (c) QTL detectance, the accuracy of inferring the genotype/phenotype relationship at loci with small contributions to the total phenotype, and (d) the molecular architecture of the QTL, that is, linked QTMs. A classic example of success with linkage analysis is the discovery by Van Laere *et al.* of the causative mutation for muscle growth and fat deposition in pigs.⁴ Identical by Descent (IBD) mapping revealed a noncoding QTN in an intron of the IGF2 gene that abates binding of a repressor and results in higher IGF2 expression in skeletal and heart muscle. This was the first noncoding QTN reported in any species. Most successful mapping studies have enhanced crossover density, often by utilizing historical recombinants and the knowledge of linkage disequilibrium (LD) and haplotype structure in the targeted species and study populations.

Genome-wide association studies

Following the success in humans,⁵ animal geneticists can now bypass the mapping steps to a large extent in the search for QTLs and eventually the underlying QTMs. As mentioned above, high-density SNP chips covering genomes with tens to hundreds of thousands of SNPs are now available for the most common domestic animal species.² Genome-wide association studies (GWAS) have been prolific in humans, where a catalog of published GWAS⁶ (www.genome.gov/gwastudies) includes 1,260 publications and 6,408 SNPs as of May 16, 2012. As

pointed out in several reviews,^{2,3,7} the power to mine mutations underlying QTLs for complex traits is much better in domestic animals than in human populations. While domestic animals have experienced the evolution of diverse phenotypes, their young history of approximately 10,000 years permits powerful genetic dissection of phenotypic diversity.² It is now well established that the LD in most domestic species is more extensive than in humans, thus permitting GWAS studies with fewer markers. Moreover, a definition of haplotype structure in different breeds or breeding populations can be exploited to reveal associations with far fewer individuals than is required in a human study. In dogs, with exceptionally high LD, power calculations have suggested that as few as 100 cases with an equal number of controls will detect loci that give a fivefold risk to a complex trait.⁸ In fact, a monogenic trait, hair ridge in Rhodesian and Thai Ridgeback dogs was mapped and the causative mutation identified from an association study with 10 cases and 10 controls.⁹

Other domestic animals have benefitted from SNP technology and GWAS as well. A number of genes for monogenic traits have been mined in horses, along with some for QTLs. GWAS in four different breeds identified a common haplotype associated with *in vitro* CD3⁺ T cell susceptibility to equine arteritis virus.¹⁰ Another study using GWAS with SNPs confirmed a variant in the equine myostatin gene as a predictor of racing distance in thoroughbreds.¹¹ While GWAS has been used mostly for milk production traits, meat quality traits, and reproduction traits in cattle, several studies have tackled the complex traits of resistance/susceptibility to infectious diseases. These include tuberculosis susceptibility in Holstein-Friesians¹² and susceptibility to paratuberculosis infection and Johne's disease^{13–17} in several breeds of cattle and sheep. Neiberger *et al.* have found loci on cattle chromosomes 2 and 26 linked with bovine respiratory disease and associated with persistent infection of bovine viral diarrhoea virus.¹⁸ In pigs, several studies have targeted meat quality, fatness, and reproductive traits, and one has identified candidate genes for *E. coli* susceptibility.¹⁹ GWAS in chickens have focused on both meat production^{20,21} and egg production.^{22,23}

Conclusions

Complex traits remain an enigma in both human and animal genetics. While large numbers of QTLs

underlying complex traits are known, the causative mutations remain elusive. However, the recent advances in genome sequencing projects, particularly the discovery of vast numbers of SNPs and their availability in affordable arrays, have opened the door to a new era in gene discovery. These tools, especially when used in GWAS, offer more than a faint hope of discovery of large numbers of genes contributing to complex traits in the very near future. The promise is especially important to studies in host resistance to infectious diseases, an important new frontier in animal and human health.

Acknowledgment

This study was supported by the WCU Biomodulation Major program at Seoul National University.

Conflicts of interest

The authors declare no conflicts of interest.

References

- Mackay, T.F. 2001. Quantitative trait loci in *Drosophila*. *Nat. Rev. Genet.* **2**: 11–20.
- Andersson, L. 2009. Genome-wide association analysis in domestic animals: a powerful approach for genetic dissection of trait loci. *Genetica*. **136**: 341–349.
- Georges, M. 2007. Mapping, fine mapping, and molecular dissection of quantitative trait loci in domestic animals. *Annu. Rev. Genomics. Hum. Genet.* **8**: 131–162.
- Van Laere, A.-S. *et al.* 2003. A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature* **425**: 832–836.
- Hirschhorn, J.N. & M.J. Daly 2005. Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.* **6**: 95–108.
- Hindorf, L.A. *et al.* 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* **106**: 9362–9367.
- Andersson, L. & M. Georges 2004. Domestic-animal genomics: deciphering the genetics of complex traits. *Nat. Rev. Genet.* **5**: 202–212.
- Lindblad-Toh, K. *et al.* 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**: 803–819.
- Salmon Hillbertz, N.H. *et al.* 2007. Duplication of FGF3, FGF4, FGF19 and ORAOV1 causes hair ridge and predisposition to dermoid sinus in Ridgeback dogs. *Nat. Genet.* **39**: 1318–1320.
- Go, Y.Y. *et al.* 2011. Genome-wide association study among four horse breeds identifies a common haplotype associated with in vitro CD3+ T cell susceptibility/resistance to equine arteritis virus infection. *J. Virol.* **85**: 13174–13184.
- Hill, E.W., B.A. McGivney, J. Gu, R. Whiston & D.E. Machugh. 2010. A genome-wide SNP-association study confirms a sequence variant (g.66493737C > T) in the equine myostatin (MSTN) gene as the most powerful predictor of optimum racing distance for thoroughbred racehorses. *BMC Genomics* **11**: 552.
- Finlay, E.K., D.P. Berry, B. Wickham, E.P. Gormley & D.G. Bradley. 2012. A genome wide association scan of bovine tuberculosis susceptibility in Holstein-Friesian dairy cattle. *PLoS One* **7**: e30545.
- Gonda, M.G., B.W. Kirkpatrick, G.E. Shook & M.T. Collins. 2007. Identification of a QTL on BTA20 affecting susceptibility to *Mycobacterium avium* ssp. *paratuberculosis* infection in US Holsteins. *Anim. Genet.* **38**: 389–396.
- Neibergs, H.L., M.L. Settles, R.H. Whitlock & J.F. Taylor. 2010. GSEA-SNP identifies genes associated with Johne's disease in cattle. *Mamm. Genome*. **21**: 419–425.
- Zanella, R. *et al.* 2011. Identification of loci associated with tolerance to Johne's disease in Holstein cattle. *Anim. Genet.* **42**: 28–38.
- Purdie, A.C., K.M. Plain, D.J. Begg, K. de Silva & R.J. Whittington. 2011. Candidate gene and genome-wide association studies of *Mycobacterium avium* subsp. *paratuberculosis* infection in cattle and sheep: a review. *Comp. Immunol. Microbiol. Infect. Dis.* **34**: 197–208.
- Settles, M. *et al.* 2009. A whole genome association analysis identifies loci associated with *Mycobacterium avium* subsp. *paratuberculosis* infection status in US holstein cattle. *Anim. Genet.* **40**: 655–662.
- Neibergs, H. *et al.* 2011. Loci on *Bos taurus* chromosome 2 and *Bos taurus* chromosome 26 are linked with bovine respiratory disease and associated with persistent infection of bovine viral diarrhoea virus. *J. Anim. Sci.* **89**: 907–915.
- Fu, W.X. *et al.* 2012. A genome-wide association study identifies two novel promising candidate genes affecting *Escherichia coli* F4ab/F4ac susceptibility in swine. *PLoS One* **7**: e32127.
- Xie, L. *et al.* 2012. Genome-wide association study identified a narrow chromosome 1 region associated with chicken growth traits. *PLoS One* **7**: e30910.
- Gu, X. *et al.* 2011. Genome-wide association study of body weight in chicken F2 resource population. *PLoS One* **6**: e21872.
- Liu, W. *et al.* 2011. A genome-wide SNP scan reveals novel loci for egg production and quality traits in white leghorn and brown-egg dwarf layers. *PLoS One* **6**: e28600.
- Shen, X. *et al.* 2012. The GTPase activating Rap/RanGAP domain-like 1 gene is associated with chicken reproductive traits. *PLoS One* **7**: e33851.