INTEGRATING MOLECULAR APPROACHES WITH QTL TO IDENTIFY POSITIONAL CANDIDATE GENES

H.C. Liu and H. H. Cheng

USDA, ARS, ADOL, 3606 E. Mount Hope Rd., East Lansing, MI 48823 USA

Animal products have been and will continue to be the major protein source for human nutrition. Awareness of such a role, diminishing areas to raise animals, and increasing consumer demands have fueled investigations for ways to improve animal performance and production. So far, conventional breeding schemes have been very successful in improving animal performance for economic traits by using the methods of selection and mating systems. However, the concern is that the rate of improvement by such traditional methods may become reduced with time.

Recent advances in genetic engineering and molecular biology have now made it possible for providing more precise selection and alternative routes for yielding high merit animals. In particular, molecular genetic approaches would be useful in enhancing genetic progress for traits with low heritability (e.g., reproducitvity), measured late in life (e.g., carcass yield), expressed in only one sex (e.g., egg production), or confer disease resistance by identifying individual genes having moderate to large effects on these traits. In the past decade great progress has been made to identify specific genes or regions of chromosomes that influence animal performance and disease resistance (reviewed by Andersson, 2001). These genes or chromosomal regions are called quantitative trait loci (QTL). The theory of quantitative genetics is that genes or QTL act together to produce continuous or quantitative variation in a trait between animals.

In order to identify and map QTL efficiently, genetic markers located throughout the genome are systematically examined if the segregation of alleles can explain some of the phenotypic variation observed in a pedigreed resource population that has been measured and is segregating for the trait(s) of interest. The power to identify and resolve a particular QTL is largely dependent on the size of the resource population and the magnitude of the allele substitution effect. This fact implies that the trait has been accurately measured. Unfortunately, environmental and other non-genetic factors can make this task very difficult. Furthermore for disease, which is not a continuous trait but rather a binary one where affected animals are observed only after a threshold has been exceeded, this problem is even worse. Thus, high resolution mapping of QTL and ultimately the cloning of the causative genes for disease resistance or other traits that are poorly measured will be an exceedingly difficult undertaking using a pure genetics approach.

For these reasons and others, we are motivated to look for other methods to fine map and identify position candidate genes for our previously identified QTL that confer resistance to Marek’s disease (MD). Fortunately, new methods continue to be improved or developed that can be conducted at the individual laboratory level and aid in our efforts.
In this paper, we will discuss current progress of identifying causative genes in MD resistance QTL by DNA microarrays and two-hybrid screens for protein-protein interactions.

THE STATUS OF MAREK’S DISEASE QTL
Marek’s disease (MD) is the most serious chronic concern to the poultry industry, especially for the egg producers. MD is a lymphoproliferative disease caused by a herpesvirus, the Marek’s disease herpesvirus (MDV) (Churchill and Biggs, 1967). As MDV persists and is ubiquitous in the poultry houses, all chicks are exposed at 1 day of age. Susceptible chickens develop tumors in the nerves and visceral tissues, which leads to paralysis, blindness, and eventually death. Annual losses in the U.S. by MD due to meat condemnation and reduced egg production exceed $160 million (Purchase, 1985).

Our group are developing methods to improve genetic resistance to MD as one control strategy. Ideally, selection for MD resistance would be based on genetic markers for the disease resistance genes or ones tightly linked to them, which is known as marker-assisted selection (MAS). MAS is expected to accelerate genetic progress by increasing the accuracy and timing of selection. Furthermore, for use in disease resistance, challenging breeding stock to hazardous pathogens could be avoided. Thus, the identification of QTL and causative genes in resistance to MD is the major goal in our laboratory. With this goal in mind, we have implemented various complementary approaches. First, genome-wide QTL scans have or are being conducted in various experimental and commercial resource populations (Vallejo et al., 1998; Yonash et al., 1999). Results from multiple genetic backgrounds advise us on the reliability of our declarations, the size of effect, and whether the QTL are segregating. Second, to achieve better resolution of the MD QTL, recombinant congenic strains (RCS) are being developed, which will enable us to genetically dissect the complex disease resistance into a series of single gene traits. Not only does this allow us to fine map the causative gene, the QTL can be associated with functional information since the genetic composition of each strain is known and each strain can be phenotyped for numerous traits. Third, the comparative map of human and chicken genomes (Smith et al., 1997; Smith and Cheng, 1998; Groenen et al., 1999; Burt et al., 1999; Groenen et al., 2000; Suchyta et al., 2001) is being leveraged to identify positional candidate genes. Essentially, due to a surprisingly high amount of evolutionary conservation, chromosomal segments from the chicken genome can be aligned with chromosomal segments from the human genome. As the human genome has been completely sequenced (gene rich), it is possible to predict what genes should reside within a chicken QTL region (gene-poor) even when the order of genes is not maintained between the two species.

By using multiple approaches, we hope minimize the limitations of each individual approach and to have several lines of evidence on the location and function of each MD resistance gene. Unfortunately, with the exception of the QTL conferring MD susceptibility on our RCS, as our QTL are of small to moderate effect ($R^2$=2 to 10%), we cannot resolve our QTL sufficiently to obtain tightly linked markers. And even when we are able to align a QTL with a human chromosomal region, there is a lack of functional information for the majority of the genes, which limits our ability to test candidate genes and form hypotheses on biological pathways.
Thus, unless we implemented additional methods, our level of understanding would be restricted to 20-30 cM chromosomal regions and minimal functional information.

DNA MICROARRAYS
One recent technological advance that we have tested to complement our genome-wide QTL scans is DNA microarray. In this technique, DNA containing known and undefined genes is spotted at very high density to specific locations on a solid support to generate a microarray or DNA chip. In the "classical" or most widely used form, mRNA is reversed transcribed and hybridized to the microarray followed by quantification of the amount of material bound by spots. Thus, one is essentially performing a Northern blot only multiplied several 1000-fold. As this gives a gene expression "fingerprint," by comparing samples from two or more treatments, it is possible to associate the expression of particular genes with the phenotype. Since the original publication by Pat Brown's group at Stanford (Schena et al., 1995), this technology has been highly touted and a number of papers have successfully demonstrated the ability of this procedure to provide insights on complex biological systems.

We have begun to test DNA microarrays containing chicken cDNAs in an attempt to identify genes and pathways involved in MD resistance. In collaboration with Joan Burnside (U. of Delaware), using arrays containing selected cDNAs from a T cell-enriched library (see the U. of Delaware Chick EST web site at www.chickest.udel.edu), the expression of genes in peripheral blood lymphocytes (PBLs, which are the targets of MDV infection) from uninfected and MDV-infected line 6 (99+% inbred and MD resistant) and line 7 (99+% inbred and MD susceptible) chicks were examined in replicate experiments. An important feature to note is that we are using chickens that we already phenotypically defined, thus, we minimize problems associated with the environment and lack of penetrance. However, unlike QTL scans that examine DNA, the choice of tissues to isolate the mRNA and the time of isolation are critical. And to get sufficient mRNA (1+ µg) for labeling, it was necessary to pool the lymphocytes from 10 chicks of each treatment and group. This pooling may have had a beneficial effect as it reduced the individual variation, which is known to occur following MDV infection, even in these highly inbred lines (e.g., Bumstead et al., 1997).

In Experiment #1 where the arrays contained ~1,200 different clones, two-fold or greater gene expression difference in one or both of the comparisons (uninfected or 8 dpi) between the lines were identified in 105 genes or ESTs (Liu et al., 2001a). In Experiment #2, 256 clones displayed two-fold or greater differential gene expression between the lines that were uninfected or 8 dpi. As reported (Table 1 of Liu et al., 2001a), 25 genes or ESTs had reproducible gene expression differences between Lines 6 and 7. Variability between the experiments for the genes with differential expression was not unexpected, especially with the number of steps that are involved in process and given the differences previously observed among individuals following MDV infection. The lack of reproducibility has been reported for induction of gene expression following MDV infection of chicken embryo fibroblasts (CEF) where only 12 of 29 genes identified were replicated (Morgan et al., 2001).

Using all the time points (i.e., uninfected and 4, 6, 8, and 11 dpi) in Experiment #2, genes or ESTs that were differentially expressed between the lines during the course of MDV infection
based on vector analysis were revealed (see Table 2 of Liu et al., 2001a). Fifty of the 55 genes or ESTs were previously screened in Experiment #1. Of these 50 genes, 11 also exhibited two-fold or greater expression difference when measured in Experiment #1. Our second and more sophisticated analytical method of vector analysis appears to be more biologically relevant based on the genes identified; however, the experiment needs to be repeated as the results may be a consequence of more time points. Interestingly, four of the 12 genes or ESTs (homology to MIP, pat.pk0067.c9; stem cell antigen 2, pat.pk0051.a3; MHC class I, pat.pk0052.h1; quiescence specific protein, pk0035.g9) that were repeatedly induced by MDV infection of CEF were also identified in our cluster analysis to be differentially expressed between our lines. With the exception of quiescence specific protein, the remaining three genes had elevated levels of expression in Line 7 compared to Line 6.

To validate the array measurements, real-time quantitative RT-PCR was performed to measure mRNA abundance, which is more sensitive and accurate than Northern blots. In all, five genes were assayed using the same RNA that was employed in the DNA microarrays in Experiment #1. In all cases, a single product was amplified and the relative ranking of the line and treatment with regard to mRNA copy number matched the results from the DNA microarrays.

Beyond the statistical analyses, as it is unclear what gene expression differences must be observed to show biological relevance, it was important for us to get some supporting functional information. Fortunately, the arrays results are support by what is already known, i.e., line 7 is readily stimulated by mitogens while line 6 shows minimal stimulation (Lee and Bacon, 1983). For example, MHC class I, genes with high homology to human HLA-DRα promoter binding protein and tapasin are all induced upon MDV infection in line 7, however, the same genes exhibit little or no change in line 6 following MDV infection. Similarly for T cell receptors, T cell receptor (TCR) ζ and γ chain, and CD3ε are induced by MDV infection in line 7 but not in line 6.

To integrate the array results with our genome-wide QTL scans, genes of interest were selected for linkage mapping. Fifteen genes or ESTs with two-fold differential gene expression or differential response to MDV challenge from Experiment #1 were chosen; the mapping was initiated before the completion of Experiment #2. Seven of these genes were previously identified and sequenced while the remaining eight genes had varying levels of sequence homology to other genes in GenBank. Seven of the genes mapped to locations that were predicted based on the chicken-human comparative map (Schmid et al., 2000), which contained 338 orthologs, while two other genes (LAMR1 and ADL0379) formed a new conserved syntenic group between human chromosome (HSA) 3 and chicken chromosome (GGA) 3 (Table 1). Thus, we confirmed the finding that it is reasonable to locate chicken genes based purely on the chicken-human comparative map. More importantly, one of the differentially-expressed genes, GH1, is known to be a MD resistance gene (see below) and another gene, SCYC1, is a prime candidate for a MD QTL on GGA1.

While the small number of genes examined limits the power of this preliminary experiment, it demonstrates that DNA microarrays yield a wealth of information as every gene provides some information (expression level, potential function), which is not the case for QTL scans.
Mapping of the differentially expressed genes may be an attractive method for identifying positional candidate genes for the MD QTL. Even if a particular gene does not lie in an MD QTL region, each gene yields clues as to what relevant biological pathway is involved and, thus, other genes in the pathway can be tested. With several groups identifying more unique cDNAs, this technology will improve rapidly.

**PROTEOMICS**

Besides variation in DNA or RNA, it is becoming more feasible to systematically examine protein variation and interactions associated with MD resistance. For example, the yeast two-hybrid system developed by Fields and Song (1989) screens a cDNA library that has been fused to the activation domain (AD) of a transcriptional activator to identify proteins that interact with bait (protein of interest) that is fused with the DNA-binding domain (BD). As the AD and BD do not need to be physically connected to promote transcription, if two proteins interact, a reporter gene is expressed if AD and BD are brought into close proximity of each other.

Our thought was if we hypothesize that some chicken proteins that interact with MDV proteins are involved in the immune response and genetic resistance to MD, then we could utilize the two-hybrid system, which when combined with genetic mapping, would quickly identify positional candidate genes for MD resistance. We recently demonstrated that this strategy does work. In this initial study, we choose the MDV SORF2 gene as bait since SORF2 overexpression in the RM1 strain may account for the change in phenotype from the parental JM102 strain (Jones et al., 1996). Two different proteins were found to specifically interact with SORF2, which were later confirmed by other biochemical assays (Liu et al., 2001b). One protein could not be identified with any homology based on its DNA sequence; however, the other protein was growth hormone (GH).

This result was significant because Kuhnlein and co-workers (1997) demonstrated an alteration in GH1 allele frequency in chicken strains selected for MD resistance. Furthermore, there is evidence that GH modulates the immune system in many species (reviews by Gala, 1991; Auernhammer and Strasburger, 1995). Genetic mapping placed the GH gene (GH1) on the small linkage group E59, which was unfortunately not previously surveyed in our MD QTL scans. Thus, an association study was conducted in a commercial White Leghorn resource population, which demonstrated that GH1 is significantly associated with several MD resistance traits. Interestingly, the GH1 association was only seen in a particular MHC background indicating an epistatic relationship between genes in the MHC and GH.

Given that the complete sequence has been determined for MDV (Lee et al., 2000), it is not difficult to imagine a systematic approach to identify all chicken-MDV interactions similar to the one employed in yeast where all interactions between all open reading frames (ORFs) were examined (Uetz et al., 2000), followed by mapping to reveal positional candidate genes or test for new trait associations. With the recent introduction of commercial two-hybrid systems based in *E. coli* (e.g., BacterioMatch by Stratagene), it is reasonable for a single investigator to screen up to 100 different baits in a year with no need for costly equipment or a large consumables budget! And as entire cDNA libraries are screened, unlike DNA microarrays,
there are no obvious limitations on the number of members being surveyed. Finally, even if an identified gene does not have a genetic basis, it does provide potential functional information and bait for additional two-hybrid screens.

Table 1. Comparison of genomic methods

<table>
<thead>
<tr>
<th></th>
<th>QTL Scans</th>
<th>DNA Microarrays</th>
<th>2-Hybrid Screens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maturity of technology</td>
<td>Proven</td>
<td>Still evolving</td>
<td>Relatively proven</td>
</tr>
<tr>
<td>Coverage</td>
<td>Extent of informative markers genotyped</td>
<td>All arrayed genes</td>
<td>Extent of cDNA library</td>
</tr>
<tr>
<td>Material requirement</td>
<td>DNA from multi-generational and measured resource population</td>
<td>RNA from tissues of defined animals and cDNAs for arrays</td>
<td>Cloned genes and cDNA library of known tissue</td>
</tr>
<tr>
<td>Equipment requirement</td>
<td>Moderate to high</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Consumable costs</td>
<td>High</td>
<td>High</td>
<td>Low to moderate</td>
</tr>
<tr>
<td>Analysis requirement</td>
<td>Moderate (single marker) to high (multiple markers, QTL, traits)</td>
<td>Low (fold differences) to moderate (clustering)</td>
<td>None</td>
</tr>
<tr>
<td>Outcome</td>
<td>QTL</td>
<td>Many differentially expressed genes</td>
<td>Up to several interacting proteins</td>
</tr>
<tr>
<td>Functional information</td>
<td>Limited</td>
<td>Yes but depends on having genes of known function</td>
<td>Maybe</td>
</tr>
<tr>
<td>Main advantages</td>
<td>Identifies regions of moderate to large effect that have a genetic influence on the trait</td>
<td>Relatively simple to perform, directly tests each gene, clues on biological pathways</td>
<td>Very simple, a limited number of candidate genes are identified, interacting proteins can be tested further</td>
</tr>
<tr>
<td>Main disadvantages</td>
<td>Resource populations take time and can difficult to generate and measure, QTL can be difficult to fine map</td>
<td>Small differential gene expression differences and structural changes will be missed, still require association studies</td>
<td>Interacting proteins may not have a genetic basis, bias to detect only strong interactions, still require association studies</td>
</tr>
</tbody>
</table>
CONCLUSIONS

More so than other scientific endeavors, genomic research is multi-disciplinary and heavily influenced by new or improved technological advancements. Each approach and advancement offers an opportunity to complement existing methods, especially genome-wide QTL scans. As discussed and shown in Table 1, DNA microarrays and two-hybrid screens can provide a viable number of candidate genes that can be mapped and tested for association to reveal positional candidate genes for QTL. It is our opinion that multiple approaches that examine multiple levels (DNA, RNA, or protein) are needed to efficiently identify positional candidate genes.

Additional opportunities will be presented when the complete genome sequence is determined for your species of interest. For example, in human methods now exist to identify unique proteins in 2-D gels by automated digestion with trypsin, mass spectrometry, and protein database searches (e.g., Jungblut et al., 1998). Given the rapid pace of the field, an investigator will have difficult decisions on what technology to incorporate and how to assimilate all the data. Fortunately, this is a relatively pleasant problem.

REFERENCES


