Deciphering the Role of Epistasis in Complex Trait Genetics

Francois Besnier[1] and Örjan Carlborg[1,2]

[1]Department of Animal Breeding and Genetics, SLU, Box 7023, SE-75007 Uppsala, Sweden
[2]Linnaeus Centre for Bioinformatics, Uppsala University, BMC Box 598, SE-75124 Uppsala, Sweden

Introduction

Most complex traits of importance in animal production are polygenic. A major challenge in animal genetics is therefore to unravel how multiple loci and environmental factors together determine phenotypic expression in individuals and phenotypic diversity in populations. Although most agree that interactions between the involved genes are likely, there is still debate as to how important epistatic interactions are in determining the phenotypes of interest in animal production. One of the reasons for this dispute not being resolved is a lack of sound experimental evidence for how genes independently or through interactions regulate the complex trait phenotypes of interest. It is therefore of great importance to stimulate research that will increase our knowledge in this area. In a study of an intercross population between two divergently selected chicken lines (Dunnington and Siegel (1996)) we found the first experimental support for an important contribution of gene interaction networks in long-term selection response (Carlborg et al. (2006)). The loci in the network interact through a mechanism that facilitates a continuous release of selectable genetic variation – a finding that has important implications for the evolution of complex traits under selection (Le Rouzic et al (2007), Le Rouzic and Carlborg (2008)). We now focus on dissecting the molecular mechanisms by which the involved QTL drive selection response. The first step in this work is to replicate and fine-map the epistatic QTL in an independent population. Replication of QTL is not a trivial task for several reasons. First, as the effects of the primary QTL are likely to be over-estimated (Beavis (1998)), it is difficult to select an appropriate population size to obtain sufficient power for replication. Furthermore, other stochastic experimental fluctuations in e.g. allelic frequencies or environmental influences, as well as unknown genetic interactions (Barton and Keightley (2002), Carlborg and Haley (2004)) further complicate the studies. Naturally, the challenge to simultaneously replicate multiple QTL that interact in networks is even greater. Here, we report the first results from our attempts to replicate the loci included in the epistatic QTL network reported by Carlborg et al. (2006) in a large eight-generation advanced intercross line (AIL).

Materials and Methods

Animal populations. An eight generation Advanced intercross line (AIL) was generated between two selected lines of chickens obtained by bi-directional, single trait, selection for

bodyweight at 56 days of age (referred to as the High Weight Selected "HWS" and Low Weight Selected "LWS" lines). The lines originate from a common base population, consisting of crosses of seven partially inbred lines of White Plymouth Rock chickens Dunnington and Siegel (1996). A large intercross pedigree was bred by crossing individuals from the HWS and the LWS lines from generation 40 where the sex-averaged mean body weights at the age of selection was 1522 g (SE: ± 36g) for the HWS line and 181g (SE: ± 5g) for the LWS line. The breeding of the intercross and a detailed description of the husbandry are described in Park et al. (2006). Briefly, 10 HWS males were mated with 22 LWS females and 8 LWS males were mated to 19 HWS females to produce 100 F_1 (a reciprocal intercross with both HWS and LWS progeny). About 100 individuals were produced in generations F_2, F_4, F_5, F_6 and F_7 and 300 and 400 individuals in generation F_3 and F_8 respectively.

DNA extraction, marker selection and genotyping. Nine chromosome regions with significant or suggestive QTL for body weight in the F_2 generation (Jacobsson et al. (2005), Carlborg et al (2006), Wahlberg et al. (2009)) were selected for further study. The segments are named as the original QTL in Jacobsson et al. (2005). DNA was extracted from blood by AGOWA (Berlin, Germany). 15 individuals from each parental strain were genotyped for approximately 13,000 genome-wide SNP markers as described in Wahlberg et al. (2009). A set of 304 segregating SNPs was selected in the nine QTL regions from the 13,000 SNPs. The average distance between the markers was less than 1 cM (Besnier et al, submitted). All individuals in the AIL (n=1529) were genotyped for these markers using the GoldenGate assay (Illumina, CA) at the SNP technology platform in Uppsala (Sweden).

Replication of epistatic QTL. In an F_2 intercross between HWS and LWS birds from generation 40 of the Virginia lines, Carlborg *et al* (2006) identified a radial network of six epistatic QTL that explained nearly half the difference in body-weight between the lines. The central locus in the network, *Growth9,* located on GGA7 was a master regulator of growth that i) turned on and off the genetic effects of other loci and ii) had its own effect reciprocally turned on and off by the radial loci in the network. Besnier et al (submitted) scanned the nine selected chromosomal segments and detected ten QTL affecting body weight at 56 days of age. The original *Growth9* QTL was shown to contain two independent QTL, designated *Growth9.1* and *Growth9.2*. Here, we evaluate whether the epistasis in radial network around *Growth9* replicate in the AIL. For this, we used the same stratification based QTL analysis as Carlborg et al. (2006). Since the original *Growth9* QTL was due to the effects of two separate loci (*Growth9.1* and *Growth9.2*), we performed two separate analyses with either *Growth9.1* or *Growth9.2* as the central locus in the network. The main features of the stratification based analysis are as follows. First, QTL genotype probabilities were estimated for the ten loci from the gametic IBD matrices of Besnier et al (submitted). Using these, the most probable genotype at each locus was determined as in Carlborg et al (2006). Second, the dataset was stratified based on the genotype (HH, HL or LL) at the central locus (*Growth9.1* or *Growth9.2* respectively). The additive genetic effects of the other eight QTL were then

estimated separately in each of the three strata. The estimates of the additive effects were obtained using linear regression of residual body weight at 56 days of age, corrected for the fixed effect of generation and sex, as in Carlborg et al (2006). Third, the dataset was stratified based on the genotype (HH, HL or LL) of each radial locus in turn and the additive genetic effect of the central locus was estimated in each strata of each radial QTL in turn. Finally, epistatic loci were defined as those that displayed either i) a significant difference between the additive effects in HH and LL genetic backgrounds and/or ii) a significant genetic effect in one homozygous genetic background but not the other.

The non-transgressive loci that were defined as epistatic in the stratification based analysis according to the rules given above were included in a joint network analysis that proceeded as follows: First, all marginal and two-way interaction effects for the epistatic loci in the network on residual body weight at 56 days of age, corrected for the fixed effects of generation and sex were estimated jointly using least squares regression as described in (Carlborg et al 2006). In the analyses, the genetic effects were modeled using the NOIA framework (Alvarez-Castro and Carlborg (2007)). Thus, in the analysis two marginal effects (additive and dominance) were estimated for each locus and four interaction effects were estimated for each pair (additive by additive, additive by dominance, dominance by additive, and dominance by dominance). The use of the 'statistical model' in NOIA provide an orthogonal model for the estimation of genetic effects, even though the population is a non-ideal F_2 regarding the allelic frequencies at each locus. From the orthogonal estimates, we constructed a multi-locus genotype-phenotype (GP) map using the 'transformation' operation in NOIA. This GP-map provides estimates of the genotype values (expected phenotypes) for all multi-locus genotype combinations in the studied network that is useful for functional studies of the epistatic interactions.

Results

Replication of epistatic QTL. Four of the eight QTL tested within alternative genetic backgrounds for *Growth9.1* had either significantly higher genetic effect in the HWS homozygote genetic background and/or a significant genetic effect in one of the genetic backgrounds and a non-significant effect in the other (Table 1). No significant epistatic effects was observed for *Growth9.2*.

Four of the five QTL (*Growth2*, *Growth4, Growth9.1* and *Growth12*) were part of the original epistatic QTL network detected by Carlborg et al (2006). The sum of their effects is 4-times higher in a HWS than in a LWS genetic background (Table 1). Together these 4 loci explain 25% of the difference between the HWS and LWS lines. The fourth locus (*Growth8*) that interacted with *Growth9.1* was not included in the original network (Carlborg et al. (2006)). Here, *Growth8* had a strong transgressive epistatic effect, i.e. the HWS alleles at *Growth8* increase body weight in a LWS genetic background at *Growth9.1*, whereas they do

not have any effect in a HWS background (Table 1). The total increase in body weight due to the network is smaller than in the original study (about 25% vs 45% of the parental line difference or 1.2 σ_P vs 2.2 σ_P).

Table 1: Genetic effects of five epistatic QTLs for body weight in the Viriginia line AIL dependent on the genetic background. The additive effects with associated standard errors (a±SE) are given for each locus (Tested QTL) in strata containing only individuals homozygous for HWS (HH) or LWS (LL) derived alleles at the Stratification QTL. The QTL where the HWS alleles increase the body-weight in the HH/LL strata respectively are given under the headings Additive/Transgressive (see text for more explanation).

Tested QTL	*Stratification QTL*	*a±SE | HH[a]*	*a±SE | LL[a]*
Additive			
Growth2	Growth9.1	36.2 ± 19.8	-7.5 ± 18.0
Growth4	Growth9.1	62.0 ± 17.0	22.7 ± 17.3
Growth9.1	Growth2	64.2 ± 18.0	18.5 ± 19.9
Growth9.1	Growth4	60.1 ± 17.3	21.4 ± 16.1
Growth9.1	Growth12	41.4 ± 16.7	5.6±16.3
Growth12	Growth9.1	45.5 ± 16.0	22.0 ± 15.6
Sum A:		309.4	82.7
% population difference:		25 %	7%
Transgressive			
Growth8	Growth9.1	11.4 ± 18.9	54.1 ± 20.0
Growth9.1	Growth8	7.9 ± 21.9	50.8 ± 17.4
Sum A:		19.3	104.9

The joint effect of the network was studied using a NOIA predicted genotype-phenotype map. In the GP-map we included the four non-transgressive interacting loci that combined to increase the body weight in the AIL (*Growth2*, *Growth4*, *Growth9.1* and *Growth12*). Carlborg et al. (2006) showed that the genetic effects of *Growth9* (including both *Growth9.1* and *Growth9.2*) increased with the total number of HWS alleles in the interacting (radial) loci. In the same way, the HWS genotype at *Growth9* released the genetic effects of the three interacting loci. To evaluate whether this still holds in the AIL, we studied a number of genotypes in the complete GP-map more closely.

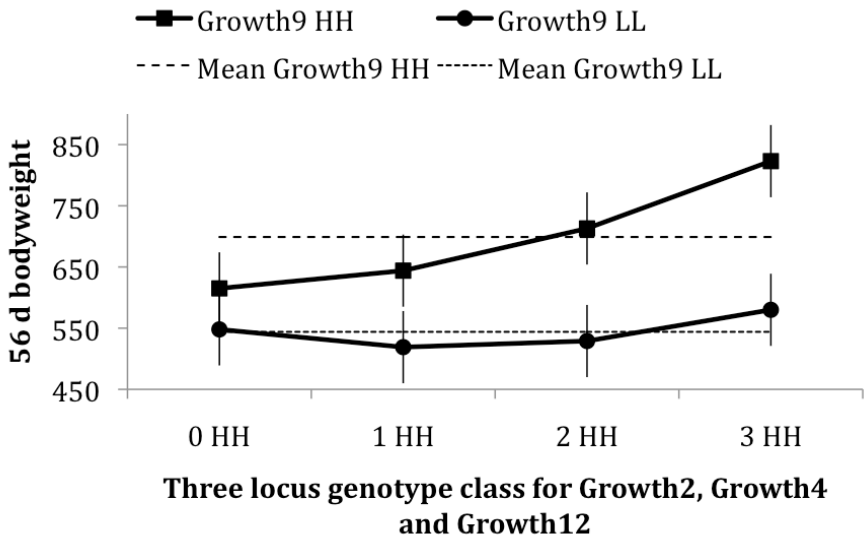

Figure 1: NOIA predicted genotype values within alternative *Growth9.1* homozygous genetic backgrounds plotted as a function of the degree of HWS homozygosity at its interacting loci *Growth2*, *Growth4* and *Growth12*. The error bars represent s.e.m., estimated using NOIA.

The original HWS and LWS genotypes, i.e. 4 locus HWS and 4 locus LWS homozygote genotypes for *Growth2*, *Growth4*, *Growth9* and *Growth12* were expected to give the highest and lowest phenotype values respectively. Figure 1 shows that this is the case also in the AIL. If the LWS homozygous genotype at *Growth9.1* suppressed the genetic effects of the other three loci, the predicted phenotype should not increase with the number of HWS alleles in a LWS homozygous *Growth9.1* genotypic background. In Figure 1 we can see that this type of interaction is present in the AIL. If the HWS homozygous genotype at *Growth9.1* releases the genetic effects of the other three loci, an increase in the genotypic values is expected when the number of HWS alleles for the radial loci is increased in the HWS homozygous *Growth9.1* genetic background. Figure 1 shows that this type of interaction is also present in the AIL. Thus, LWS homozygosity at *Growth9.1* suppress the growth-promoting effects of HWS alleles in *Growth2*, *Growth4* and *Growth12*, and HWS homozygosity at *Growth9.1* promotes it. The interaction effect is reciprocal as the HWS allele in *Growth9.1* has a very low growth promoting effect in LWS background of *Growth2*, *Growth4* and *Growth12*. Figure 1 also show that the additive effect of *Growth9.1*, which can be detected in a QTL scan for marginal effects, is thus not caused by an individual genetic effect of *Growth9.1* but rather by the combination of beneficial alleles at the four loci. The same holds for the marginal additive effects of *Growth2, Growth4* and *Growth12*.

Discussion

Advances in methods for large scale molecular characterization of individuals has provided the means to explore the contribution of epistasis to phenotypic expression for many complex traits of agricultural importance. One of the difficulties with QTL mapping in general is that it has proven difficult to replicate QTL and identify the genetic mechanisms underlying the original observations (Barton and Keightley (2002)). Whether these difficulties are due to use of too small populations for replication, for the inheritance of the QTL being more complex than pure additivity or that the initial discoveries are false positives is often not known. It is, however, worth to note that knowledge of non-additive modes of inheritance at loci has proven to be highly useful for the success in molecular dissection of QTL (van Laere et al. (2003)). One challenge when aiming for robust and powerful detection of epistasis is the requirement of a more powerful experimental dataset than for detection of single QTL with marginal (i.e. additive and dominance) effects. As the number of multi-locus genotypes increases exponentially with the number of included loci, the requirement for larger datasets is obvious. In our experience, experimental F_2 populations used for explorations of first order (two-locus) epistasis should preferably include in the order of 1000 individuals and if possible even more. As there are still rather few datasets available of this size, a comprehensive picture of the general importance of epistasis in the regulation of different types of complex traits is still lacking. Most of the initial reports of epistasis have therefore been for high heritability traits and the most likely explanation for this is that the power of QTL mapping is greater for these traits. The general expectation is, however, that epistasis should be of greater importance in the regulation of low-heritability traits, including disease resistance and fertility, but more studies are needed to explore this type of traits further. Replication and genetic dissection of the underlying molecular mechanisms of epistatic interactions will require even more resources than the initial QTL mapping. As a result of this, there are very few studies in the literature that have replicated, fine-mapped or cloned epistatic QTL. Here we report the, to our knowledge, first study that replicates a network of interacting loci in vertebrates. The success in this study provides an important precedence in that it experimentally shows that genetic interactions are a real contributor to complex trait phenotypic expression and that the epistatic mechanisms that can be observed are stable over generations. Further work is needed to identify the molecular mechanisms underlying the detected interactions, but our results should stimulate others to include explorations of epistasis in primary as well as replication studies.

The primary focus in most animal breeding programs is to achieve maximum genetic gain in every generation. Theory, as well as experience, has shown that the additive genetic variance in a population is useful for predicting the genetic gain from one generation of selection. Thus, in most animal breeding applications, the importance of an allele substitution is evaluated based its average substitution effect in the population of interest (i.e. its contribution to the additive genetic variance). When epistasis is present, however, the

average substitution effect might not be the same in all populations or even in the same population in different generations. Eitan and Soller (2004) introduced the term SIGV (Selection Induced Genetic Variation) to describe a situation where epistasis affects selection response by increasing the selectable variation in response to selection. Let us use an example to illustrate how this works. Consider a locus (A) whose additive genetic effect (a) is independent of other loci and equal to a. Its contribution to the additive genetic variance is $V_A(A)=2pqa^2$ regardless of the genotype at other loci. Now consider the alternative scenario where there is epistasis that makes the additive genetic effect of locus A dependent of the genotype at locus B. Here, the additive effect of locus A is $2a$ in genotype BB, a in genotype Bb and 0 in genotype bb. When f(B)=0.5, the contribution of locus A to V_A is similar to when there is independence, i.e. $V_{A|f(B)=0.5}=2*p*q(0.25*2a+0.5*a)^2=2pqa^2$. However, when the allele frequency at locus B changes, so does the contribution of locus A to the additive genetic variance in the population. E.g. when f(B)=0.8, its contribution is $V_{A|f(B)=0.8}=2*p*q(0.64*2a+0.16*a)^2=2.88pqa^2$ and when f(B)=0.2, its contribution is $V_{A|f(B)=0.2}=2*p*q(0.04*2a+0.16*a)^2=0.48pqa^2$. Thus, depending on the allele-frequency at the epistatic locus, the average allele substitution effect, and thus its contribution to the selectable additive genetic variance, changes. When selecting on this system, the response will be an increase in the allele frequencies of the A and B alleles in the two loci. As a result, the total selectable additive genetic variance will increase and there will be a release of cryptic (or standing) genetic variation as Selection Induced Genetic Variation (SIGV). The existence of SIGV was experimentally validated by Carlborg et al. (2006) that showed its contribution to selection response in the Virginia chicken lines. When studying the long-term effects of selection, it is therefore important to be aware that epistasis can dramatically change the contribution of allele substitutions to the additive genetic variance over time. Allele substitutions that might appear to contribute marginally to the additive genetic variance in one population might make a large contribution in another, although the potential for ultimately shifting the phenotypic mean in the population in a longer perspective is the same in both cases (Le Rouzic and Carlborg (2008)). The same is true within populations, where alleles in loci might appear to make a small contributions in one generation, but be a main contributor in a later generation when allele frequencies have changed in response to selection. It has been argued that interactions at the level of genes is unlikely to generate much interaction at the level of variance in a population and that the implication of this is that interactions are of minor importance in evolutionary biology, medicine and agriculture (Hill et al. (2008)). The same data can, however, be interpreted in a completely different way. The finding that much of the additive variance in a population can controlled by gene-interactions is interesting as it strongly emphasizes that it is of great importance to identify the molecular interaction mechanisms that allow long-term selection response in populations to phenotypes beyond what can be explained by standard additive genetic models. Our results is a clear example of where identification of multi-locus genetic interactions is the key to understand how the Virginia lines have been able to respond to selection to reach levels far outside the original range of phenotypes in the base population without depleting the additive variance.

Conclusion

Here we present the first replication of a multi-locus QTL interaction network that underlies long-term selection response for increased body weight in chicken. The replicated interaction mechanism facilitates SIGV (Selection Induced Genetic Variation), i.e. a release of selectable additive genetic variation in response to selection due to the change of allele frequencies at epistatic loci that in a reciprocal way releases the genetic effects of other loci. This finding provides strong evidence for the importance of genetic interactions in determining selection response for agriculturally important traits and calls for further studies of epistasis in the future.

References

Alvarez-Castro, J. and Carlborg, Ö. (2007). *Genetics*. 176 1151-1167.

Barton, N.H. and Keightley, P.D. (2002). *Nat. Rev. Genet*. 3, 11-21.

Beavis, W. D. (1998). QTL analyses: power, precision, and accuracy, pp. 145–162 in Molecular Dissection of Complex Traits, edited by A. H. Paterson. CRC Press, New York.

Besnier, F. and Carlborg, Ö. (2009) *BMC Genetics*. 10:57

Carlborg, Ö. and Haley, C. (2004). *Nat. Rev. Genet*. 5: 618-625.

Carlborg, Ö., Jacobsson, L., Åhgren, P., Siegel, P., Andersson, L. (2006). *Nat. Genet*. 38:418-20.

Dunnington, E. A. & Siegel, P. B. (1996) Poult. Sci. 75, 1168-79.

Eitan, Y. & Soller, M. in Evolutionary Theory and Processes: Modern Horizons. Papers in Honour of Eviatar Nevo (ed. Wasser, S.P.) 153–176 (Kluwer Academic, Dordrecht, The Netherlands, 2004).

Hill, W.G., Goddard, M.E. and Visscher, P.M. (2008). *PLOS Genet*. 4:e1000008.

Jacobsson, L., Park, H.B., Wahlberg, P., Fredriksson, R., Perez-Enciso, M., Siegel, P.B., Andersson, L. (2005). *Genet. Res*. 86 115-125.

Le Rouzic, A. and Carlborg, Ö. (2008). *Trends in Ecology and Evolution* 23:33-37.

Le Rouzic, A., Siegel, P. and Carlborg, Ö. (2007). *BMC Biology* 5:50.

Park, H.B., Jacobsson, L., Wahlberg, P., Siegel, P.B., Andersson, L. (2006). *Physiol Genomics*. 25 216-223.

Pong-Wong, R., George, A.W., Woolliams, J.A. and Haley, C.S. (2001). *Genet. Sel. Evol*. 33 453–471.

Rönnegård, L., Besnier, F. and Carlborg, Ö. (2008). *Genetics*. 178 2315-2326.

Van Laere, A.S., Nguyen, M., Braunschweig, M., Nezer, C., Collette, C., Moreau, L., Archibald, A.L., Haley, C.S., Buys, N., Tally, M., Andersson, G., Georges, M., Andersson, L. (2003). *Nature*. 425 832-836.

Wahlberg, P., Carlborg, Ö., Foglio M., Tordoir X., Syvanen A.C., Lathrop M., Gut I.G., Siegel P.B., and Andersson, L. (2009). *BMC Genomics* 10:24