Summary for Discussion: Correlations Between Phenotype and Genotype

Maggie Schedl

Words: 946

When associating phenotypes to genotypes, or more specifically individual variants or single nucleotide polymorphisms (SNPs), there are multiple directions that a researcher can come at the method. These directions depend on both the question being asked, and the study system available to the researcher. For example, Nadeau *et. al*  (2016) tackled the task of looking for genes associated with wing color patterning across an entire genus of butterflies. Their reasoning behind pursuing such a broad study was that many of the species within the genus ﻿*Heliconius* have mimetic coloration patterning, and there is substantial prior information in this genus about areas of the genome that effect color patterns. Thus, they could solely focus on the *Yb* locus, and make repeated associations across the phylogenetic tree to see if there was a common region of the locus putatively effecting wing color patterning in all species of the genus. In contrast, Bosse *et. al* (2017), began with a single species, but potentially divergent populations, and no clear phenotype. Using genetic data alone, they used principle components analysis (PCA) to identify and investigate their phenotype of interest, which turned out to be bill morphology. Consequently, they had to pursue whole genome association analysis, as their phenotype is one generally unstudied in their organism and is highly polygenic. And in the “best-case” scenario, one could have an extremely well-studied organism, the ability to do molecular gene editing techniques, and a known gene that effects a phenotype of interest. Such was the case for Barret *et. al* (2019) and their investigation into the serine mutation in the very well-studied Agouti gene concerning fur color.

In terms of genome wide association analyses, the foundational method of phenotype-genotype association methods, only Bosse et al. (2017) demonstrated it in the most basic sense. If there is no previous research in your study organisms about associations between a phenotype and the genome, then a genome-wide approach is necessary. The authors did both an EigenGWAS to probe the genome for a phenotype to focus on, and then further narrowed their search for highly associated variants with a classic GWAS. If there is a great deal of previous research on genes effecting certain phenotypes, then a very detailed study can be made finding exact SNPs with large effects (Barrett et al. 2019). This is possible because fur coloring in mice is not polygenic in the sense that bill morphology in great tits is, and the Agouti gene and its surrounding genomic region is well known to effect fur color (Siracusa 1994).

However, specific individual SNPs or mutations may not be of interest to your study if you are looking across multiple species for a consensus in a gene or region controlling a phenotype (Nadeau et al. 2016). In this case their goal, and ultimately their results, was to find a gene that had different SNPs across the genus, and was consistently and repeatedly differentiated and associated with the varying phenotypes. An illuminating way to investigate this is to also consider differential expression across the locus in different species or regions in the organism. Nadeau *et. al* (2016) did this in two ways, with microarrays and with in-situ hybridization during various stages of butterfly metamorphosis. Neither Bosse, nor Barrett *et. al* considered investigating their SNPs of interest during their effect on development. Bosse *et. al* (2017) did some minimal functional exploration with preliminary gene ontology analysis of genes highlighted in their EigenGWAS peaks, but there were likely some missing GO terms, and it still remains unknown how ﻿COL4A5-C effects beak morphology on a physical level.

All three of these papers found and focused on SNPs in an exon or region of a coding gene. Perhaps more can be elucidated about polygenic traits from investigating regulatory regions of genes as well. However, there is a significant barrier to this with the large bias in general biological knowledge about coding regions/exons in comparison to non-coding and regulatory regions of genomes. While not an association study, one of the papers for last week’s journal club (Kolaczkowski et al. 2011) considered the genomic region (ie. gene, regulatory, intron etc.) as different and importantly distinct sections of their Fst distributions. This type of detailed investigation may be possible in the mouse study, in Figure 4 we can see that there are significantly associated SNPs not in the exons of Agouti, however the authors chose not to focus on those variants.

A clear goal in a phenotype-genotype association study is to tie in broader evolutionary processes. This can be done by measuring selection with field work (Barrett et al. 2019), or by measuring differences in fitness in populations of interest (Bosse et al. 2017). Barrett *et. al* compared the differences in the Serine deletion allele frequency compared to non-deletion alleles before and after the first time point, which was considered a selection event because many of the mice experience predation during the interval. Bosse *et. al* measured the number of fledged offspring across the homozygous COL4A5-C allele, ancestral allele, and heterozygous alleles to account for fitness differences. The lack of either of the previous two types of methods was obvious in Nadeau paper, even though the results of a singular gene being an effector across such a wide group of insects is remarkable. Although traits such crypsis and mimicry can certainly be argued as adaptive, under selection, or concurring different levels of fitness, it was not shown as such in this study. While clear that detections of important genes that are highly differentiated between populations or groups can be elucidated with only genomic data, when exploring the reasons *why* there is differentiation, detection of selection acting on phenotypes is a necessary component of a study.

References

Barrett, R.D.H. et al., 2019. Linking a mutation to survival in wild mice. *Science*, 504(February), pp.499–504.

Bosse, M. et al., 2017. Recent natural selection causes adaptive evolution of an avian polygenic trait. *Science*, 368(October), pp.365–368.

Kolaczkowski, B. et al., 2011. Genomic Differentiation Between Temperate and Tropical Australian Populations of Drosophila melanogaster. *Genetics*, 187, pp.245–260.

Nadeau, N.J. et al., 2016. The gene cortex controls mimicry and crypsis in butterflies and moths. *Nature*, 534(7605), pp.106–110. Available at: http://dx.doi.org/10.1038/nature17961.

Siracusa, L.D., 1994. The agouti gene: turned on to yellow. *Trends in Genetics*, 10(12), pp.423–428. Available at: https://www.sciencedirect.com/science/article/pii/0168952594901120 [Accessed February 18, 2019].