PHENOTYPE-GENOTYPE ASSOCIATION - DISCUSSION REVIEW

Ian Bishop

BIO-594

2-18-19

Word Count:?

These three research articles are good examples of how one might use population genomics and a host of other molecular biology methods to examine divergent phenotypic traits and their association to particular genomic features, be they regulatory intronic SNPs, amino acid deletions, or widely dispersed SNPs. While the study systems vary both in genomic and ecological resources, each group systematically uncovers phenotype-genotype associations. Each study sought to validate the association via experimental means with methods as varied as fluorescence *in situ* hybridization (FISH), correlating reproductive success and foraging behavior to alternative genotypes, and induction of negatively selected phenotypes by transgenic mutation.

First off, Nadeau et al. (2016) (NA) examined the genomic locus *Yb*, known to influence Lepidopteran wing color. They sequenced the locus region in three *Heliconius* species of varying wing phenotypes (mostly yellow bars on fore- or hindwing, absent or present) and uncovered more than 100 SNPs in and around the gene *cortex* that were fixed for different phenotypes. All of the strongly associated segregating sites were located in *cortex* introns or its 5' UTR. Because the SNPs were all in non-coding regions, the authors concluded that the variants were regulatory, not functional. They also thought variation in phenotype could be associated with splice variants, a hypothesis they tested by running RT-PCR. (They found several splice variants related to non-constitutive exons, but didn’t pursue them further). The most powerful analysis conducted was the test for differential expression of *cortex* between species and between wing segments with different phenotypes. They ran a microarray of probes for all predicted genes in the *H. melpomene* genome, as well as tiled probes from the region containing *cortex*, in two species, one with and one without yellow markings. They found *cortex* to be differentially expressed, and specifically at sites within *cortex* introns. This pretty much settled things, but NA followed up with one more test: they performed *in situ* hybridization in larval wing tissue and found that the gene was upregulated in specific wing areas that correspond to specific color patterns in adult specimens. The *cortex* gene is a member of the fzy gene family, which is insect-specific and is most commonly associated with cell cycle regulation (anaphase promotion).

Second, Bosse et al. (2017) (BE) studied on-going divergent selection in UK populations of the great tit (*Parus major*). Because BE did not have an a priori target gene or genomic region of interest, they conducted an EigenGWAS analysis, which treats an explanatory principal component as the phenotype of interest, on ~500,000 SNPs in 2300 birds across three populations, two in the Netherlands and one in the UK. This analysis identified particular outlier regions under putative divergent selection, and the most commonly GO term associated with genes in these regions was "palate development". Thus we have a lead. Next, a second GWAS was conducted to look for SNPs associated with palate development in UK birds. This revealed 3 genes, *SOX6, PTHrP* and *COL4A5*, the last of which was most associated with beak length and hence was the focus going forward. While these genes were associated with beak length, they explained minimal variation in that phenotype, from which the authors concluded that the phenotype was under polygenic selection. From here, several interesting features were examined, including high LD values for the *COL4A6* SNP most associated with beak length (genomic architectural signatures of divergent selection), correlation of reproductive success and COL4A5 allele (conferred fitness advantage), and differences in bird feeder visits among genotyped, radio-tagged individuals that differentiated at the segregating site of interest in *COL4A6* (support for hypothesis that bird feeders foraging is the selective mechanism responsible for beak elongation).

Finally, Barrett et al. (2018) (BA) documented a rich story of natural selection. They began by identifying a phenotype (dorsal brightness) under selection for improving crypsis, to experimentally determining that *agouti*-related SNPs changed in frequency when individuals were relocated to different selective environments. Then they identified a specific candidate deletion as increasing dorsal brightness, and also determined the functional impact of that deletion (change in attractin binding affinity via surface plasmon resonance). Finally, as in NA, they put a nail in this association by comparing transgenic C57BL/6 lab mice with and without the candidate serine deletion in *agouti*, which demonstrated that the deletion led to cryptic lightening of mouse hair.

Overall, these studies were very dense with respect to how many analytical methods were employed, but it allowed each to tell an interesting story of phenotype-genotype association. One particularly salient distinction between the study design of BO and that of NA and BA is that the latter two studies begin with a particular, well-studied locus in mind. The foundational knowledge that model organisms provide can be cost effective and can open up different methods that require a more targeted genomic approach. For example, it allowed NA to constrain the region for which they purchased tiled microarray probes, and it allowed BA to focus their high-coverage sequencing effort on the *Agouti* gene. In a sense, knowing a priori which gene or genomic region to target based on community-wide study of particular model organisms and specific phenotypic traits depends on another salient feature: whether or not the trait of interest is polygenic. In BO's case, there was no particular region known ahead of time because beak length, the actual phenotype under selection, is governed by a widely dispersed set of alleles. It may be the case that the massive genomic resources available for *Parus major* enabled the study of such a polygenic feature in the first place. So far I’ve found enjoyed the pairing of model and non-model systems for each theme – it gives us a perspective of what is more or less possible with the creatures we study. This last part is excellent, just spend a lot less time summarizing each study and more time contrasting them.

***References:***

Barrett, R. D. H., Laurent, S., Mallarino, R., Pfeifer, S. P., Xu, C. C. Y., Foll, M., … Hoekstra, H. E. (2018). The fitness consequences of genetic variation in wild populations of mice. *BioRxiv*. <https://doi.org/10.1101/383240>

Bosse, M., Spurgin, L. G., Laine, V. N., Cole, E. F., Firth, J. A., Gienapp, P., … Slate, J. (2017). Recent natural selection causes adaptive evolution of an avian polygenic trait. *Science*, *358*(6361), 365–368. <https://doi.org/10.1126/science.aal3298>

Nadeau, N. J., Pardo-Diaz, C., Whibley, A., Supple, M. A., Saenko, S. V., Wallbank, R. W. R., … Jiggins, C. D. (2016). The gene cortex controls mimicry and crypsis in butterflies and moths. *Nature*, *534*(7605), 106–110. <https://doi.org/10.1038/nature17961>