Phenotype to Genotype Discussion 2-19-19 Summary

Matías Gómez and Maggie Schedl

### **Papers discussed:**

**Barrett et al. (2019)**. Linking a mutation to survival in wild mice. *Science*, 363(6426), 499–504.

**Bosse et al (2017)**. Recent natural selection causes adaptive evolution of an avian polygenic trait. *Science*, 358(6361), 365–368.

**Nadeau et al. (2016)**. The gene cortex controls mimicry and crypsis in butterflies and moths. *Nature*, 534(7605), 106–110.

**Figure and Description Matching Activity**

An activity at the beginning of the discussion was formatted to have everyone receive a figure from one of the papers and a small description of another figure. The goal was to match the figure with the description, and then categorize it with the correct paper. This activity aimed at making associations on general ideas or methods of what the figures portrayed just by the plots/graphs/images themselves. This way, one could test their familiarity with the methods utilized, and their own familiarity with the papers.

**Bosse et. al 2017**

From the matching activity, Figure 3 (a,b & c) from Bosse et. al (2017) was highlighted as a confusing figure. B and C show a haplotype bifurcation diagram for the UK population of *P. major* surrounding the identified SNP locus COL4A5-C and the ancestral allele COL4A5-T. The main point of the figure is to show the extended length of the haplotype in the individuals with the homozygous COL4A5-C SNP, this could be indicative of a hard sweep for this allele if it is supported by relatively high values of the *Fst* statistics. Theoretically, if a hard sweep has taken place on any given locus its neighboring bases should exhibit low genetic diversity, as the selective pressure acting upon it would tend to favour the maintenance of that genomic region within the same linkage block. Bifurcation diagrams have the potential for there to be a node split at every base left or right from the core SNP. If there is a haplotype difference, a split is created. Thus, the extension of a single haplotype by a continuous thick block with few bifurcations. Overall, identifying genomic regions with unexpectedly high local haplotype homozygosity relatively to neutral expectation is a powerful strategy to ascertain candidate genes responding to natural or artificial selection (see [here](https://watermark.silverchair.com/bts115.pdf?token=AQECAHi208BE49Ooan9kkhW_Ercy7Dm3ZL_9Cf3qfKAc485ysgAAAn8wggJ7BgkqhkiG9w0BBwagggJsMIICaAIBADCCAmEGCSqGSIb3DQEHATAeBglghkgBZQMEAS4wEQQMnp9jcTqfn-c3y-_mAgEQgIICMt0NfRZ2cI-dg0hX4wGHSEv-dPq2pdQRbXNbDgD0UkOQiv0kgIp9S4s_hsLrct8n3Sr7l1qxvjGn_bQrJ_qUz4f9sc9doVYPy6djvOBAuvW0_hRynsCdr7ERAF4OLxLZ9j7xNoxjnluouc6fdsCCg3a2kwUYfERjyUd8PgQBQZpUOnoR69yiInE8dqxZLBlAdRYH29lEJSBBIjFopqKt1usSSM7vr8RprrVHOKRatY-JJ_vU32TBZPiuUXUEcp1UReXvuWLhuRDlHcE1cqrNUBSrvHTSJHJj2leUiz_4QXb3lJvKg7yUGH92gWSnmzhE746dOdSJHZdFKAZWyEeyQNbrtl9o66LPgth19cug--xwLPyhQYvODyMfWbc739smQsKzaEmmCcCwZ2KQq6HrB2m3SawMRWxAl7vONLKGgPKSmh_HgRSnaudSN1FgCSIwfBPyeERLDxAE4hxTzkqi7MvvPifgggxG-LZfvUY5UpTil0aJfJT7ahFn1xWO0hM8kVtDfwc2jkN4KzmFTh3CvhaJsVDsclHayZiLYXZ1ko4nTKbBebcifbV628mQFI81o-ghPtw1g_bcxoKcsdUmaRS2mJg37GgpRRA4wbNMqBy_vr5x8jbOjzymoGJ7s-ijUtXpev8r-4Zf2ey3SBeheJn0b4OPsz14CfCzqpYpdkvw_HqUX1QUg3hfRLoejbQb9G7cs7nGYWzIPIaLavUCo5EW--wW72DVcsbBaeH-0NP6hxs)).

The class then briefly summarized this study. The authors compared genotypic and phenotypic variation across three populations of the great tit (*Parus major*) bird, two from the Netherlands and one from the UK, a species with long-term observations and abundant genomic resources. The main result of the study was that the UK population has differentiated from the Netherlands populations in mean bill size, which the authors attributed to the prevalent use of supplemental bird feeders in the the UK. The authors initially adopted a bottom-up approach to search for loci associated with the differences between the UK and Netherlands population by inferring a phenotype from the first eigenvector of the genome-wide PC analysis. Most outlier regions were related to skeletal development and morphogenesis functions, which were represented in a gene ontology (GO) network. This kind of networks reflects the relationships between the GO terms based on similarity of their associated genes (see [here](http://apps.cytoscape.org/apps/cluego)), and this is how they chose the phenotype of bill morphology. They utilized museum skins to obtain bill phenotypic measures and field monitoring of bird feeders use by great tits along with their morphological measurements.

**Barrett et al. (2019)**

The discussion then moved on to the study done on deer mice (*Peromyscus maniculatus*) and its differential survival rates associated with mutations on a locus that controls coat color. On the Sand Hills in Nebraska, a recently geological (8,000-10,000 years ago) shift in ground color is thought to have exerted a strong selective pressure on fur color for crypsis. To test this hypothesis the authors set up field enclosures where light-colored and dark-colored mice were translocated to dark-soil and light-soil environments, respectively. Since the enclosures excluded terrestrial predators but were open to birds of prey, it was expected to observe different survival rates of mice in each kind of enclosure caused by avian predation only. All mice had been previously genotyped for the locus that codes for coat coloration (*Agouti*) so the original allele frequency was known and could be estimated as the population in each enclosured decreased. Results showed that the survival rate was twice as high in dark enclosures relative to light ones. Volcano plots were used to look for evidence of adaptive selection on this locus vs neutral simulated expectations. This kind ofscatter-plot is used to quickly identify changes in large data sets composed of replicate data (see [here](http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/650/index.php?manual=Volcano_plots_inspecting_result_statistical_analysis.html)). A significant change in allele frequency was indeed found between the beginning of the experiment and three months later in the dark soil plot, confirming the hypothesis of adaptation to matching ground color. They used a genome wide association study to look for SNPs associated with dorsal coat brightness. A heatmap of linkage disequilibrium, used to measure the non-random association of alleles at two or more loci in a general population (see [here](https://link.springer.com/referenceworkentry/10.1007%2F978-3-642-16483-5_3368) and [here](https://cran.r-project.org/web/packages/LDheatmap/vignettes/LDheatmap.pdf)), showed that seven candidate variants for selection belonged to three linkage blocks that are responding to selection for coat color in this species. They selected the 𐊅Ser allele for further study because it was the most associated with the phenotype, and they used the molecular biological technique to show that the 𐊅Ser caused a decrease in binding of the agouti protein and less pheomelanin production, showing a clear link from the mutation to the lighter coat color phenotype.

**Nadeau et al. (2016)**

Briefly we touched on the Nadeau paper in the time remaining. We focused on the broad question that the authors based their study on. Phenotype and genotype association analysis does not have to be completely narrowly focused, the authors in this study looked for signs of repeated loci that could be implicated in adaptive evolution of the same trait across different species. Previous research from this group showed that color patterning is adaptive in *Heliconius* butterflies. They looked broadly across the genus and found that SNPs in the gene *cortex* consistently were associated with wing color patterning, even across different species where homologous genes are not necessarily in the same position or have the same genetic structure (Figure 2). We also briefly pointed out how this paper explored gene expression in a way that the other two did not. They used [microarray analysis](https://www.nature.com/scitable/definition/microarray-202) on the known [Yb locus](https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.0040303) to further confirm that *cortex* was differentially expressed during times of development that wing color scales are being formed. And they backed that up with in-situ hybridization showing direct expression of cortex in specific patterns in the developing wings that exactly match adult color pattern. This type of analysis is a very good example of going beyond a statistical association to a functional association that really provides insights on how adaptation has occurred in this genus.