Introduction: Schizophrenia (SCZ) is a psychiatric disorder characterized by frequent hallucinations and delusions, with associated substantial morbidity and personal and societal costs.[1] Drug treatments are available, yet their success is relatively ineffective for many patients. Identifying the underlying genetic variants related to SCZ has been compromised by debates on coding vs. non-coding variants and uncertainty of magnitude and phenotypic effect size of these variants.[2-6] The mutations that cause Mendelian disorders typically occur at nucleotide positions that are evolutionarily conserved due to purifying selection, the strength of which is dependent upon the severity of the disorder.[3-4] However, this approach has not been applied in complex or spectrum disorders like SCZ, partly because of the number of loci and lack of knowledge on the underlying selection model. For example, several have suggested that some complex behavioral and brain-related disorders are the byproduct of historical positive selection on specific variants, which when in combination can result in diseases like SCZ.[5] We have previously used this evolutionary model with other traits to identify signatures of conservation and adaptation in resolving variants at single loci with the most clinical potential.[6] With a recent GWAS meta-analysis of thousands of SCZ cases and controls that identified 108 linkage independent loci (Fig. 1),[2] we propose to develop a new evolutionary model that explains the relationship between evolutionary conservation, selection, and etiology associated with complex traits.

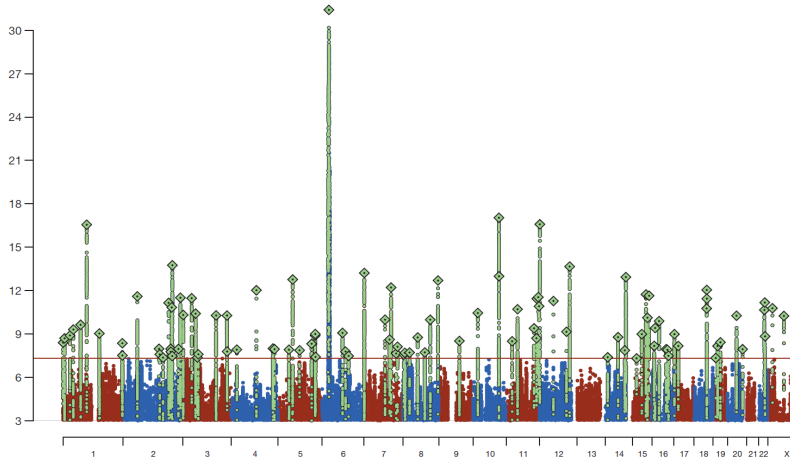

Figure 1. Chromosomal locations (x-axis) of 108 GWAS loci associated with SCZ. A total of 128 GWAS significant variants, depicted as a green "diamond" here, were identified by a genome-wide significance cut-off (horizontal red line; Y-axis $-\log_{10}$ P-value). All green points below each GWAS variant reflect variants in significant LD with that GWAS variant (adapted from PGC[2]). These loci were as long as 800 kb, and span all 23 chromosomes.

Hypothesis: If mutations underlying SCZ are associated with deleterious functional change, then we expect GWAS variants to occur at genomic sites that have high evolutionary conservation across phylogenetic time. We also predict that global human population datasets will mirror these results as evidenced by signatures of strong recent purifying selection within these loci. On the other hand, while SCZ is severe, it is not fatal; thus, an alternative hypothesis is that GWAS variants occur at sites that show a distribution of conservation strength across phylogenetic time, reflecting both positive and purifying selection, and possibly only more recently within primate lineages. In addition, we would expect that human population datasets exhibit signatures of positive selection for some genes consistent with recent human adaptation.

Study Design: We will generate nucleotide sequence datasets including each of the 108 loci based on multiz46way alignments and corresponding phylogenetic trees (UCSC GenomeBrowser), relative to different depths of vertebrate (n=46), mammal (n=36), and primate (n=10) clades. Nucleotide-specific conservation scores, and their statistical significance at each depth will be generated using the phyloP likelihood ratio test (LRT) .[7] Statistical phyloP outliers within each depth are identified using nonparametric box-and-whisker plot analyses. To compare

locus-wide phyloP LRT score distributions across phylogenetic depths to look for consistencies, we designed a novel approach that takes the sites within each of the 108 loci and z-transforms them at each of the 3 depths. Variants within each locus, including the GWAS hit, will be evaluated for deviation in relative conservation by comparison to the z-score value corresponding to the 95th percentile of sites at each locus, and at each depth. Finally, we will use data from genome-wide scans of positive selection across these 108 loci; for example, a recent hierarchical boosting (HB) score of 11 different positive selection tests was published to identify possible regions of positive selection.[5] Patterns of long-term evolutionary conservation from phyloP across depths will be correlated with HB patterns within human populations to test hypotheses of purifying and positive selection noted above. Populations to be sampled will include, but not limited to, 1000 Genomes data samples (Phase I release, 2012): CEU (Utah residents with European ancestry), CHB (Han Chinese in Beijing), and YRI (Yoruba, Nigeria).

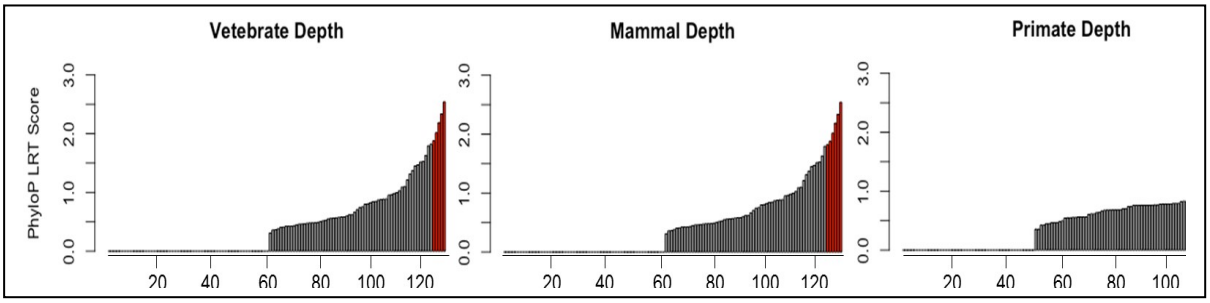

Figure 2. PhyloP LRT score histograms. Conservation estimates, ranked (x-axis) from low to high phyloP LRT, are shown for the 128 GWAS variant sites at each of three phylogenetic depths. Statistical outliers are in red.

Preliminary Results: We have conducted an analysis of only the 128 GWAS variants at this point, which shows phyloP LRT distributions of the three phylogenetic depths (Fig. 2) are not significantly different (P-values > 0.16). Almost 50% of the variants occur at sites with little evolutionary conservation (low phyloP), nonetheless, our non-parametric analysis detected several variant sites that were outliers (Fig. 2). Interestingly, our z-transformation of the phyloP LRT scores, revealed that many of the GWAS variants >95th percentile at that locus, suggesting that these sites are unusually conserved within their genomic regions. On the other hand, some sites at which GWAS hits occurred showed significantly different conservation across depths, i.e., highly conserved in vertebrate depth only. These results appear to be consistent with our alternate hypothesis in finding a distribution of functional impact; nevertheless, it remains to be seen what this approach reveals for the thousands of variants in significant LD within these loci, as well as patterns in recent evolutionary time depth in human populations that we have planned.
Intellectual Merits and Broader Impacts: As human genome and GWAS projects increase, studies like this one that use novel evolutionary analyses of these data will be needed to identify functional and adaptive variants. This project, while specific to SCZ, can be adapted to any complex disorder or trait in any organism. The broader impact of this work would directly fund an African-American female doctoral student in a field of evolutionary genetics and bioinformatics where these individuals are underrepresented.

References

1. Knapp et al. 2004. Schizophr. Bull. 30, 279–293 2. Psychiatric Genomics Consortium. 2014. Nature 511:241; 3. Subramanian and Kumar. 2006. BMC Genomics 7:306; 4. Kumar et al. 2011. TIG 27:377; 5. Polimanti and Gelernter. 2017. PLoS Genetics; 6. Stover & Verrelli. 2011. Mol Biol Evol 28:533; 7. Siepel et al. 2006. in Proc 10th Int'l Conf Res Comput Mol Biol.