Itchy and scratchy: Polygenic risk scores predict self-reported mosquito attraction and bite response

James R. Ashenhurst*, Sarah B. Laskey, Michael Multhaup, Olga V. Sazonova, the 23andMe Research Team, Nicholas A. Furlotte, and Bertram Koelsch

23andMe, Inc., Mountain View, CA   *jashenhurst@23andme.com

Introduction

Mosquitos present not merely an annoyance but also a significant threat to global public health due to the transmission of disease. For example, the World Health Organization estimates that there were 207 million cases of malaria resulting in 627,000 deaths globally in 2012[1]. As such, susceptibility and response to mosquito bites are phenotypes of substantial interest. Twin studies indicate that susceptibility to mosquito bites is a heritable trait[2], and a previous genome-wide association study (GWAS) identified 15 independent genetic loci associated with self-reported attractiveness to mosquitoes, bite size, or itch intensity[3], and suggested shared phenotypic and genetic etiology.

In the present study, we assessed the predictive power of genetic risk scores (GRS) for each of these three phenotypes. Considering the shared phenotypic and genetic etiology[3], we hypothesized that combining genetic features identified across the three separate GWAS into a single superset of variants would yield better polygenic prediction than using features from a single GWAS alone.

Lastly, we explored the generalizability of European-trained polygenic risk models of the mosquito attraction phenotype to populations of non-European descent.

Methods

Participants
We selected a sample of unrelated individuals of European descent who provided informed consent and all answered three questions about self-reported attractiveness to mosquitoes relative to others, bite size response, and itch intensity. This sample was split into 80% included in GWAS and model training (N = 311,556) and 20% reserved for model testing (N = 77,890). Lastly, non-European model calibration was assessed among unrelated people of African American (N = 23,278), East Asian (N = 11,935), Latinx (N = 73,094), Middle Eastern (N = 3,266), and South Asian (N = 3,258) descent who provided consent and answered any of the three questions.

GWAS & Feature Selection
We performed three GWAS in the European sample as described previously[3], using variants assayed on any one of five custom Illumina genotyping arrays. GWAS included sex, age, the first five genomic principal components and genotyping platform as covariates. We identified variants associated with each of these three phenotypes that reached a significance threshold $p < 5e-6$, and pruned this list to select index variants within a window of 500kb.

Model Fitting & Non-European Populations
Linear additive models were fit in the training sample, and evaluated in the testing set as previously described[4]. First, models were fit using only features from a single GWAS for that specific phenotype. Next, models were evaluated using the superset of variants drawn from all three GWAS. Calibration and predictive power of the European-trained model was evaluated among the non-European groups.

Acknowledgments & References

1) World Health Organization, *A Global Brief on Vector-Borne Diseases*, 2014. 2) Fernandez-Grandon, G.M., *et al., PLoS One*, 10 e0122716. 3) Jones, A.V. *et al., Human Molecular Genetics*, 26(7):1391-1406, 2017. 4) Furlotte, N.A. *et al., 23andMe White Paper 23-12*, 2015. 5) Pers, T.H., *et al., Nature Communications*, 6, 5890, 2015.
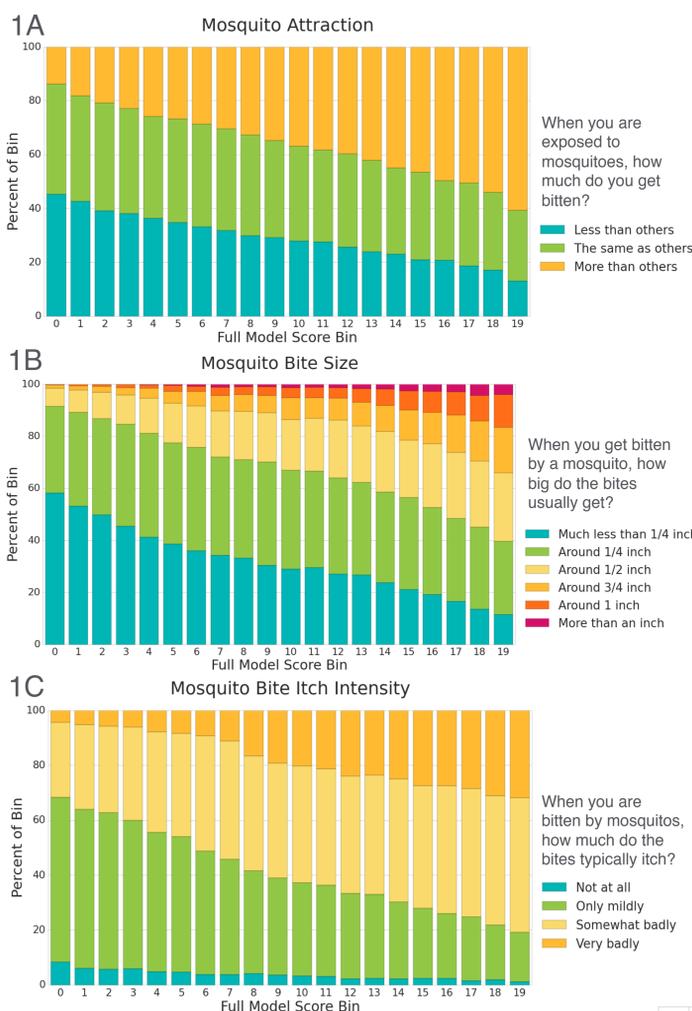
Results

GWAS & Variant Selection
The GWAS yielded 139, 105, and 128 independent markers ($p < 5e-6$) for mosquito attraction, bite size, and bite itchiness respectively, comprising a superset of 285 variants. Tissue and gene set enrichment analyses (using DEPICT[5]) implicated genes associated with these loci in blood and lymph tissues and in gene sets involved in hematopoiesis, myelopoiesis, and leukocyte activation.

Polygenic Models
Additive linear models fit in the training sample using only the phenotype-specific variants as features in a genetic risk score (GRS) explained 1-2% of the variation in the three phenotypes in the testing set (**Table 1**). Addition of age and sex features (the full model) increased the variance explained to 6-12% (**Figure 1**). GRS trained using the superset of markers (285 variants) resulted in minor improvement in model performances relative to the individual variant sets (mean increase in R^2 = 0.43%, **Table 1**) and in highly correlated full model scores (mean R^2 = 0.89).

1A Mosquito Attraction

When you are exposed to mosquitoes, how much do you get bitten?
- Less than others
- The same as others
- More than others

1B Mosquito Bite Size

When you get bitten by a mosquito, how big do the bites usually get?
- Much less than 1/4 inch
- Around 1/4 inch
- Around 1/2 inch
- Around 3/4 inch
- Around 1 inch
- More than an inch

1C Mosquito Bite Itch Intensity

When you are bitten by mosquitos, how much do the bites typically itch?
- Not at all
- Only mildly
- Somewhat badly
- Very badly

Figure 1. Phenotype prevalence across risk score bins. Plots represent the prevalence of each response choice to the survey question across the distribution of risk scores when using the superset of 285 variants and age and sex in the testing sample. Each column represents 1/20th of the population (of European ancestry) sorted by full model scores for the three phenotypes A) Mosquito Attraction, B) Bite Size, and C) Itch intensity. Full model scores meaningfully stratified risk across the sample.

	Response Phenotypes			Variant Superset (285 SNPs)			Phenotype-specific SNP sets		
	1	2	3	4	5	6	7	8	9
1. Mosquito Attraction	-	0.236	0.347	0.026	0.023	0.025	0.022	0.015	0.019
2. Bite Size	0.236	-	0.334	0.017	0.017	0.017	0.014	0.012	0.014
3. Bite Itch Intensity	0.347	0.334	-	0.024	0.021	0.025	0.018	0.014	0.021
4. Mosquito Attraction PRS (285 SNPs)	0.026	0.017	0.024	-	0.877	0.924	0.841	0.58	0.696
5. Bite Size PRS (285 SNPs)	0.023	0.017	0.021	0.877	-	0.873	0.665	0.778	0.66
6. Itch Intensity PRS (285 SNPs)	0.025	0.017	0.025	0.924	0.873	-	0.695	0.582	0.841
7. Mosquito Attraction PRS (139 SNPs)	0.022	0.014	0.018	0.841	0.665	0.695	-	0.448	0.49
8. Bite Size PRS (105 SNPs)	0.015	0.012	0.014	0.58	0.778	0.582	0.448	-	0.488
9. Itch Intensity PRS (128 SNPs)	0.019	0.014	0.021	0.696	0.66	0.841	0.49	0.488	-

Table 1. Pairwise R^2 between phenotypes and model scores. Coefficient of determination between response phenotypes and GRSs derived from the superset of SNPs or from the phenotype-specific SNP sets in the testing sample.

Table 2. Mosquito attraction full model R^2 in each group.

Population	N	R^2
European	77,890	0.06
African American	23,278	0.03
East Asian	11,935	0.05
Latinx	73,094	0.05
Middle Eastern	3,266	0.03
South Asian	2,258	0.02
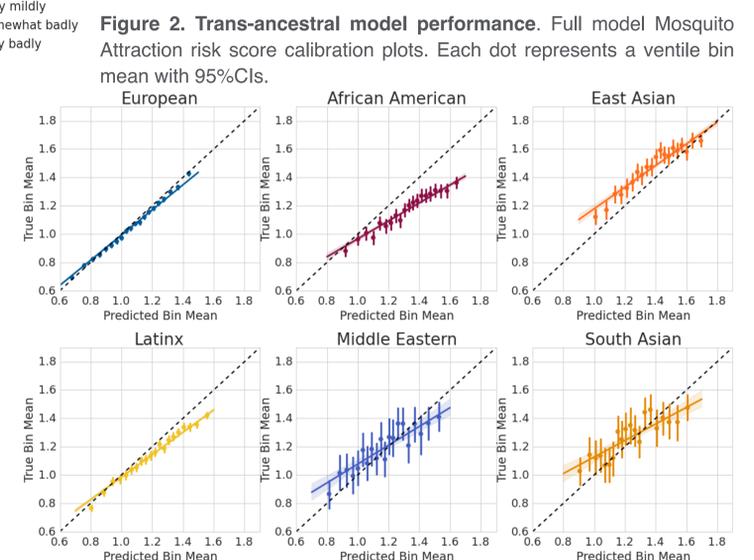
Figure 2. Trans-ancestral model performance. Full model Mosquito Attraction risk score calibration plots. Each dot represents a ventile bin mean with 95%CIs.

Discussion

Despite their modest predictive power, the full models meaningfully stratify participants by phenotype (**Figure 1**). For example, there was an over four-fold increase (60.6% vs 13.8%) in the fraction of participants rating themselves as "more attractive" to mosquitos than their peers among those in the highest 20th versus the lowest 20th of the score distribution. Scores derived from these models could be used to identify individuals who might most benefit from protective strategies to avoid mosquito bites.

Our hypothesis was confirmed that prediction using the superset of variants associated with these three phenotypes yielded stronger models of each, suggesting that polygenic modeling may benefit from information derived across multiple highly correlated phenotypes. Predictive power among people of non-European descent was not equal across groups (**Table 2**), with calibration of the European-trained mosquito attraction model being worse among people of African American, Middle Eastern, and South Asian descent (**Figure 2**). This may be due to smaller sample sizes, but differences in minor allele frequencies and linkage disequilibrium structure may also result in reduced generalizability of European-trained models to other ancestral groups even if the true causal loci are the same.