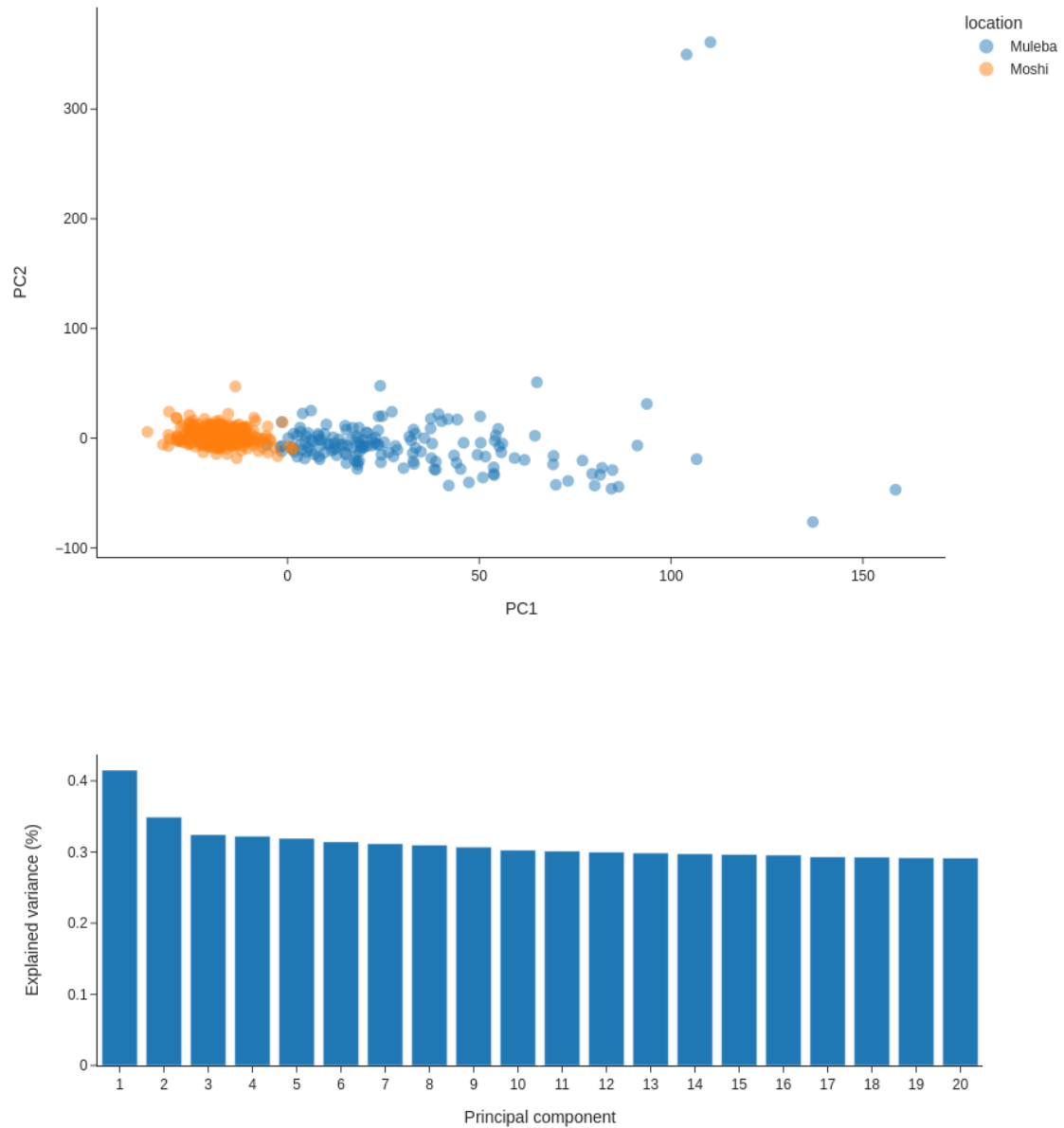


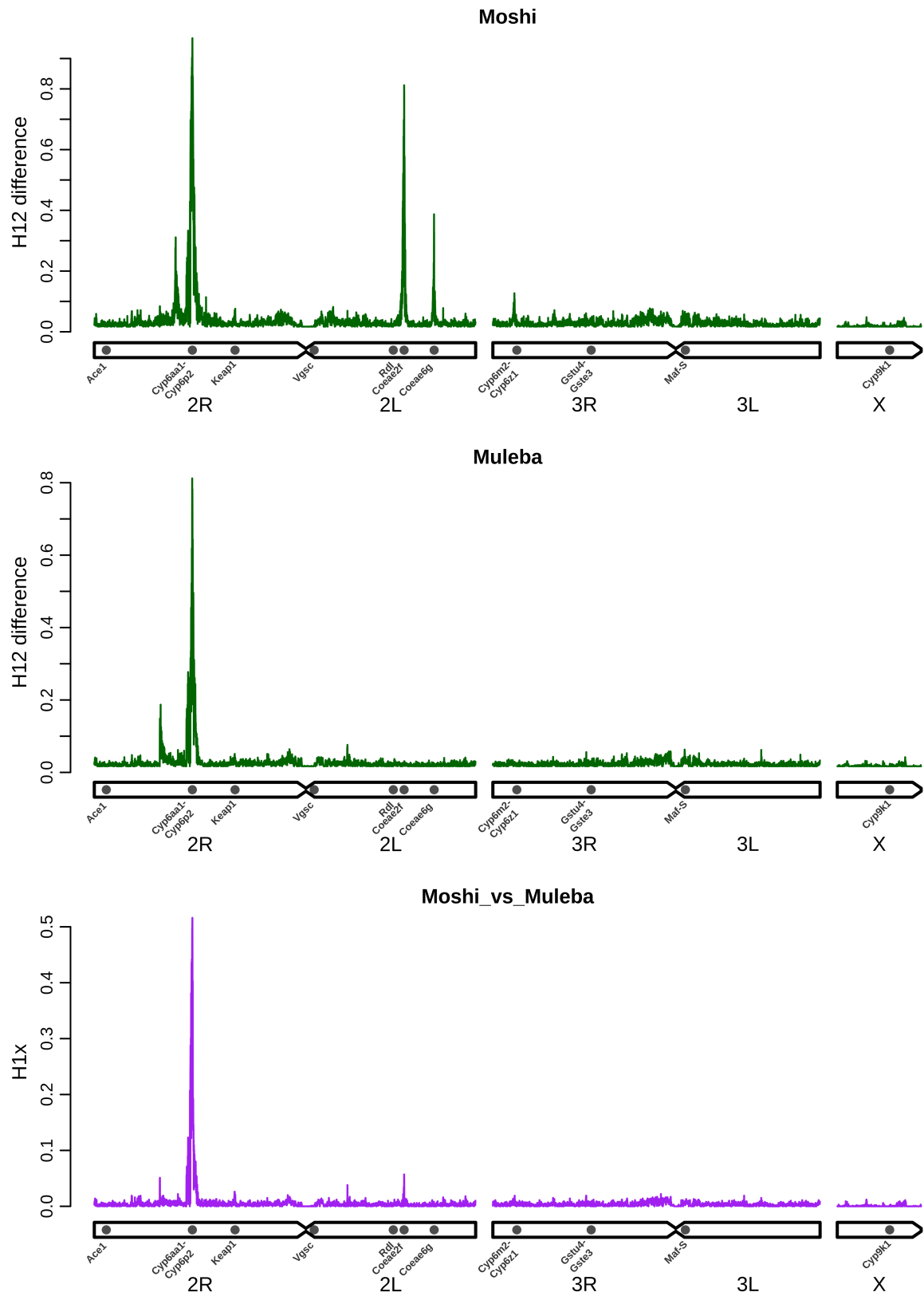
**Copy number variants underlie the major selective sweeps in insecticide resistance genes in *Anopheles arabiensis* from Tanzania.**

Eric R. Lucas, Sanjay C. Nagi, Bilali Kabula, Arjen E. Van't Hof, Emily J. Rippon, Dimitra Pipini, Nicholas J. Harding, Naomi A. Dyer, Chris S. Clarkson, Alistair Miles, Martin J. Donnelly, David Weetman

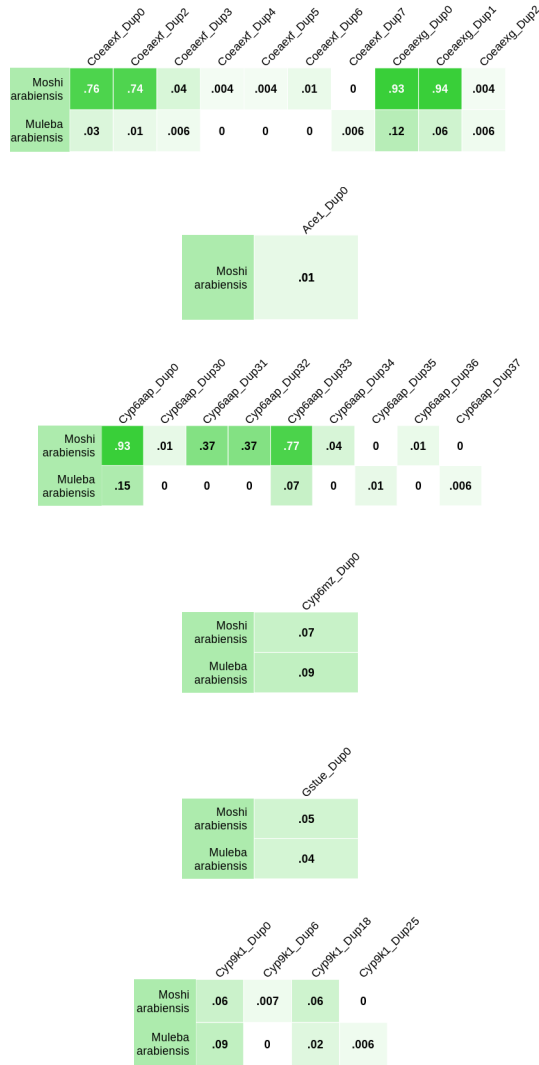
**Electronic Supplementary Material**  
**Supplementary figures and tables**



**Fig. S1:** PCA (using quality-filtered biallelic SNPs from genomic region 3L:15,000,000-41,000,000, eu-chromatic and free of chromosomal inversions). Top panel shows clustering of samples by region. Bottom panel show variance explained by the first 10 PCs, indicating that PCs 3 onwards explain similar levels of variance and are thus likely only capturing noise.



**Fig. S2:** Selection scans showing genome-wide H12 signal in Moshi (top) and Muleba (middle), as well as shared signals of selection (H1x) between the two sites (bottom).



**Fig. S3:** Frequency (proportion of samples carrying at least one copy) of known CNV alleles detected using diagnostic reads around the *Coeaezf* cluster, the *Coeaezf* cluster (combined into a single subfigure), *Ace1*, the *Cyp6aa* / *Cyp6p* cluster, the *Cyp6m* / *Cyp6z* cluster, the *Gste* cluster and *Cyp9k1*. Only CNV alleles with frequency > 0% are shown. Cell darkness provided as a visual aid for the magnitude of the value in each cell. The genomic coordinates of each CNV allele can be found in Supplementary Data S@. In each cluster, the “Dup0” column indicates the presence of increased copy number in any of the genes in the cluster. Where this is larger than the sum of known alleles, it suggests the presence of uncharacterised CNV alleles. “Del” alleles in *Ace1* represent secondary deletions within the *Ace1*-Dup1 CNV.