Detecting and Localizing Inversions from SNPs

KR Manke[1], SJ Emrich[2], and RJ Nowling[3]

[1]Physics and Chemistry Department, Milwaukee School of Engineering [2]Electrical Engineering and Computer Science, University of Tennessee-Knoxville [3]Electrical Engineering and Computer Science, Milwaukee School of Engineering

Abstract

Chromosomal inversions play an important role in ecological adaptation by enabling the accumulation of beneficial alleles (Love, et. al. 2016; Fuller, et al. 2017) and reproductive isolation (Noor, et a. 2001). The 2La inversion in the *Anopheles gambiae* complex has been associated with thermal tolerance of larvae (Rocca, et al. 2009), enhanced desiccation resistance in adult mosquitoes (Gray, et al. 2009), and susceptibility to malaria (Riehle, et al. 2017). Additionally, inversions must be identified and accounted for to avoid bias in population inference and association testing (Seich al Basatena, et al. 2013).

Methods

We compared three techniques for detection and localization: LD (linkage disequilibrium) averaged over sliding windows, minor allele frequencies (MAFs) averaged over sliding windows, and Manhattan plots from PC-SNP association tests. We applied these techniques to 34 *Anopheles* samples (Fontaine, et al. 2015) and 198 *Drosophila* Genetics Reference Panel v2 samples (Mackay, et al. 2012; Huang, Massouras, et al. 2014). The *Anopheles* samples have been karyotyped for 2La but not inversions on 2R, while the DGRP2 data set contains five inversions (*In(2L)t*, *In(2R)ns*, *In(3R)mo*, *In(3R)p*, and *In(3R)k*) present in five or more samples detecting and localizing inversions along a chromosome using SNPs.

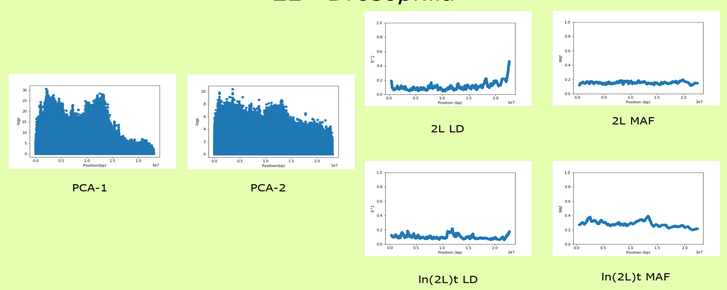
In order to calculate LD, the open-source whole genome association analysis toolset Plink was utilized to yield R^2 values that reflect SNP correlation. VCFtools is an open-source package for working with VCFs; it was utilized for filtering VCFs based on MAF as well as for producing reports on allele frequency. Asaph is a software tool for population genetic analysis from insect SNPs, with it single-SNP association tests were performed and the resulting p-values were used to identify principal components. The Python Pandas library is for data analysis and manipulation; it was used for processing the LD reports to calculate R^2 averages and for extracting the MAF for each SNP, as well as preparing the data for visualization.

Results

2L - *Anopheles*

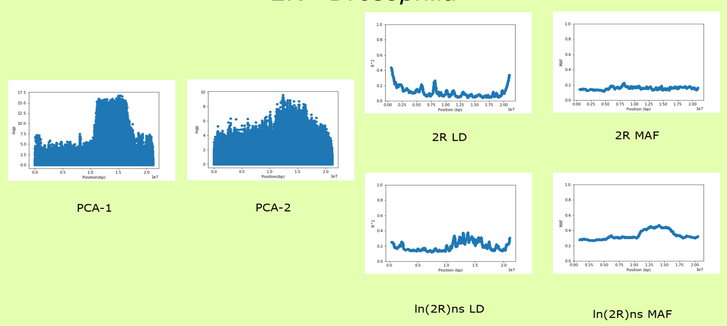
Results

2L - *Drosophila*

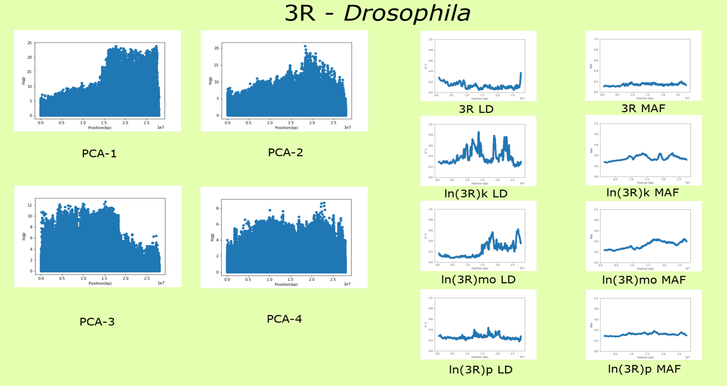


2R - *Anopheles*

2R - *Drosophila*



3R - *Drosophila*



Discussion & Conclusion

When applied to all of the samples, the LD and MAF analyzes were unable to detect any of the inversions except for 2La in *Anopheles*. The detection power for LD and MAF were increased by isolating samples by population (*Anopheles*) or homozygous inverted and heterozygous karyotypes (DGRP2). LD was able to identify 2La among Burkina Faso samples and one 2R inversion among Mali samples. MAF analyses further identified 2La and two separate 2R inversions among both Burkina Faso and Mali samples. Both of the LD and MAF analyzes detected *In(2R)ns*, *In(3R)k*, and *In(3R)mo* from analysis of the inverted DGPR2 samples but gave ambiguous results for *In(2L)t* and *In(3R)p*.

Unlike the LD and MAF analyses, the PC-SNP association tests detected inversions from whole data sets. The 2La and two 2R inversions in *Anopheles* and *In(2L)t* and *In(2R)ns* in the DGRP2 samples were unambiguously detected. An inversion was detected on 3R but *In(3Rk)*, *In(3R)mo*, and *In(3R)p* could not be separated.

In conclusion, PC-SNP association tests are more accurate than both LD and MAF analyses and do not require knowledge of the samples' karyotypes. PC-SNP association tests have the potential to significantly improve inversion detection and localization.

References

1. Love, R. R., et al(2016), Chromosomal inversions and ecotypic differentiation in Anopheles gambiae: the perspective from whole-genome sequencing. Mol Ecol 25: 5889-5906. doi:10.1111/mec.13888
2. Fuller ZL, Leonard CJ, Young RE, Schaeffer SW, Phadnis N (2018) Ancestral polymorphisms explain the role of chromosomal inversions in speciation. PLOS Genetics 14(7): e1007526. doi:10.1371/journal.pgen.1007526
3. Mohamed A. F. Noor, Katherine L. Grams, Lisa A. Bertucci, Jane Reiland (2001) Chromosomal inversions and the reproductive isolation of species. Proceedings of the National Academy of Sciences 98 (21) 12084-12088; doi: 10.1073/pnas.221274498
4. Emilie M Gray, Kyle AC Rocca, Carlo Costantini and Nora J Besansky (2009) Inversion 2La is associated with enhanced desiccation resistance in Anopheles gambiae. Malaria Journal 8:215. doi:10.1186/1475-2875-8-215
5. Kyle AC Rocca, Emilie M Gray, Carlo Costantini and Nora J Besansky (2009) 2La chromosomal inversion enhances thermal tolerance of Anopheles gambiae larvae. Malaria Journal 8:147. doi:10.1186/1475-2875-8-147
6. Alejandro Cáceres, Juan R. González (2015) Following the footprints of polymorphic inversions on SNP data: from detection to association tests. Nucleic Acids Research 43(8):30. Page e53. doi:10.1093/nar/gkv073
7. Ma J, Amos CI (2012) Investigation of Inversion Polymorphisms in the Human Genome Using Principal Components Analysis. PLOS ONE 7(7): e40224. doi:10.1371/journal.pone.0040224
8. Ronald J. Nowling and Scott J. Emrich. 2018. Detecting Chromosomal Inversions from Dense SNPs by Combining PCA and Association Tests. In Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics (BCB '18). ACM, New York, NY, USA, 270-276. doi:10.1145/3233547.3233571
9. Michael C. Fontaine et al (2015) Extensive introgression in a malaria vector species complex revealed by phylogenomics. Science 347(6217):1258524. doi:10.1126/science.1258524
10. Trudy F. C. Mackay et al (2012) The Drosophila melanogaster Genetic Reference Panel. Nature 482:173.
11. Wen Huang et al (2014) Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines. Genome Research 24:1193. doi: 10.1101/gr.171546.113