Jun 28, 2017

Dear Dr Hutchins,

Thank you for considering our paper for publication, and for your and the reviewers comments, which will improve the final form of our publication.

We addressed all the concerns, and indicated them clearly (in red) in the following letter.

Kind regards,

Reviewer: 1

Thank you very much for your thorough review, observations and helpful suggestions.

Comments to the Author

Tekman et al. in their manuscript entitled “HaploForge: A comprehensive pedigree drawing and haplotype visualization web application” present a web application to visualize haplotypes and draw pedigrees, that allows for both autosomal and X-linked haplotypes from outbred and consanguineous pedigrees. The application colors IBD segments and processes output from commonly used linkage analysis packages.

1. The authors compare their newly developed tool to a visualization program developed back in 2005 (HaploPainter); however, newer tools have been developed such as Family Genome Browser (Bioinformatics, 2015, 31:2262-2268), which also provides “comprehensive” tools for analysis and visualization of pedigree and variant data, including parent of origin, potential recombination events, IBD, LD, annotations for the variants, etc . The Family Genome Browser can be accessed via the web or installed locally. What features does HaploForge possess that are not possessed by newer tools such as the Family Genome Browser?

The merits of the **Family Genome Browser** (FGB) are discussed extensively in section 4.5 (Comparison with Additional Haplotyping Software), and improvements to HaploForge based upon it are mentioned in section 4.7 (Future Work).

The author notes that the desktop rendition of FGB was based upon the evaluation of the source provided in their github repository, since access to desktop binaries was not possible due to broken links.

2. The authors state that their application can “navigate high numbers of markers” and “non-trivial families” (e.g., 27 members). Are there limits to the number of markers or pedigree members that can be explored? For example, can whole genome sequence data be explored? Is it possible to analyze pedigrees with hundreds of members?

We have added an additional simulation study to show how well HaploForge scales (see Section 4.4). We demonstrate that for common scenarios (up to 400 individuals and 1000 markers) HaploForge loads in less than a second. We show good performance with up to a million markers in nuclear families. We tested pedigrees with up to 10,000 individuals, which worked with fewer markers dependent on the amount of RAM available.

3. HaploForge colors IBD information on the pedigree. It is not clear how HaploForge uses the linkage segregation information to determine haplotypes/IBD versus the A\* search algorithm.

HaploForge uses the phased genotypes output by programs that perform haplotype reconstruction such as Allegro and GeneHunter. The A\* search algorithm is necessary because of ambiguous cases where parents are homozygous. The A\* search procedure very simply finds the flow of alleles that minimizes the number of recombinations. IBD colouring is performed by identifying contiguous blocks of markers in-between recombinations.

Reviewer: 2

Thank you very much for your thorough review and appreciation of our work.

Comments to the Author

Tekman et al present a novel program to analyze and visualize the flow of genetic information through a family. The two main aims of this new application are 1) greater accuracy to determine flow of alleles in the family compared with HaploPainter and 2) improved clarity of the representation of these haplotypes, eg by allowing side-by-side comparison of multiple family members across a pedigree. Although I lack the expertise to critically appraise the extent to which the first aim is achieved, the example using X chromosomal haplotypes suggests that HaploForge does represent a significant improvement compared with Haplopainter. The second aim is substantially achieved by this application and I conclude that HaploForge will be valuable to researchers and geneticists analyzing family trees, especially for the purposes of conducting linkage studies.

Minor points:

1. The extent to which this application is useful in pedigrees being studied using whole genome or whole exome sequencing is less apparent, and discussion of this issue would be helpful, given that, especially in a research context linkage analysis is arguably less critical than it was in the pre-Next Generation Sequencing era.

Indeed, please see Section 4.3 that highlights this issue.

2. The authors state that that "there are many programs available for visualization" but they only compare HaploForge with HaploPainter. This is justified by the fact that HaploPainter is the most widely cited example, but brief descriptions of other available programs would help to provide more context for HapoForge.

A number of other haplotype programs were initially considered but were not mentioned because their features did not match HaploForge, i.e. draws pedigrees and renders haplotypes.

Section 4.5 compares the features of the two additional applications that can draw pedigrees and render haplotypes, and new supplemental material provides a table that outlines the features of 8 others.

3. Can HaploForge facilitate identification of extended haplotypes that are identical-by-state across different families (suggesting evidence of a founder effect among families not known to be related to each other)?

This is the purpose of our scoring function which has now been expanded on at the end of section 3.3.