1. Abstract

The prevalence of increasingly dense genotyping chipsets prompted the need for high-throughput genetic linkage pipelines to handle the vast amounts of data they generate. Despite the new paradigm in genetics being more sequence analysis driven, the importance of linkage analysis still remains a powerful tool to pinpoint loci of interest within the tumultuous over-abundance of sequence data. Linkage analysis alone is often more than capable of identifying a single causative gene under the scope of a rare disease model for the relatively low cost of a genotyping array, and has the added advantage of reconstituting genotypes of individuals absent from the analysis via haplotype reconstruction.

Here we present our comprehensive linkage analysis pipeline consisting of a collection of well-established tools and utilities (GRR, Merlin, Alohomora) to perform analysis under all penetrance models (dominant/recessive, autosomal/X-linked) as well as large complex consanguineous pedigrees. Pre-analysis filtering selects a subset of markers indicative of informative meioses, and qualitative tests are performed to check for gender, relationship, and Mendelian inheritance consistency. A reliable lineup of linkage analysis suites (Allegro, GeneHunter, Simwalk) compute LOD scores to produce fast genome-wide and chromosome-specific linkage plots complete with sub-banding overlays and peak validation. Limitations in pedigree creation and post-analysis haplotype examination applications (HaploPainter) further prompted the development of a new visualization tool, HaploHTML5, built upon the latest advances in the HTML5 web schema. Pedigrees can be drawn and analysed in-browser, and haploblock resolution is performed using the novel approach of a best-first path-finding algorithm (A\*) implemented in pure JavaScript.

Small (<19-bit) pedigrees with an informative input set of 40,000 markers were processed in under 15 minutes (single core) and 5 minutes (multi-core). Complex (>19-bit, inbred) pedigrees required extensive code and platform-specific modifications to cater for some of the more outdated software (Allegro) which increased the runtime to hours (19 to 23-bit) and days (>23-bit). HaploHTML5 performed correct haplotype resolution for X-linked inheritance models compared to HaploPainter.