1. Abstract

The new paradigm in genetics is more sequence analysis driven, but linkage studies are frequently adopted in parallel to pinpoint loci of interest within the vast torrent of sequence data. Linkage analysis alone can identify a single causative gene under the scope of a rare disease model for the relatively low cost of using increasingly dense genotyping arrays, with 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 multi-core 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 are 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19-bit, inbred) pedigrees required extensive code and platform-specific modifications to cater for outdated software (Allegro) which increased the single-core run-time to hours (19 to 23-bit) and days (>23-bit). Correct X-linked haploblock resolution was determined via HaploHTML5, and side-by-side case/control evaluation was facilitated. A web interface is in development for the rest of the linkage pipeline for local or cloud-based deployment .