GET-Evidence - Feature #578

Improve reporting of variant/allele frequencies

05/19/2011 07:06 PM - Tom Clegg

Status:	Feedback	**Start date:**	05/19/2011
Priority:	Normal	**Due date:**	
Assigned To:		**% Done:**	100%
Category:		**Estimated time:**	0.00 hour
Target version:			
Billable:		**Totalhours:**	
Estimatedhours:		**Resolution:**	fixed
Hours:			

Description

Pieces:

- Current data on evidence.personalgenomes.org seems to have been imported from "all.txt" (should have used "parsed_for_getev.txt") in mball data
- Better version of parsed_for_getev.txt is available too
- Should **not** use sum(num)/sum(denom) because many of the same genomes are included in multiple populations (hapmap, 1000g, cgi)
  - Just use local genomes, and generate a report of variants whose frequencies differ wildly from 1000genomes/hapmap?

History

#1 - 05/23/2011 09:56 PM - Tom Clegg

- Status changed from In Progress to Feedback

- % Done changed from 0 to 100

- Resolution set to fixed

Done:

- Fixed update_variant_frequency.php to use population with highest denominator, instead of summing populations (985000, ec7159)
- Updated import_variant_frequency.php to use "for getev" format (ea8475c) -- also, look up "unknown" variant_ids using variant_name
- Imported parsed_for_getev_new.txt from mball on production site. E.g., COL4A1-Q1334H, ATAD3C-A91V