DSRG 2007: A Comparison of DNA Variant Screening Strategies – DNA Sequencing and SNaPshot

G.J. Wiebe[1], D. Adam[2], D. Bintzler[3], M. Detwiler[4], H. Escobar[5], J. Kieleczawa[6], D.S. Needleman[7], P.A. Schweitzer[8], M. Zianni[9], P.S. Adams[10]

[1]Max Planck Institute of Molecular Cell Biology & Genetics, Dresden, Germany, [2]University of British Columbia, Vancouver, BC, Canada, [3]University of Cincinnati, Cincinnati, OH, United States, [4]Roswell Park Cancer Institute, Buffalo, NY, United States, [5]University of Utah, Salt Lake City, UT, United States, [6]Wyeth Research, Cambridge, MA, United States, [7]USDA ARS Eastern Regional Research Center, Wyndmoor, PA, United States, [8]Cornell University, Ithaca, NY, United States, [9]Plant-Microbe Genomics Facility/Ohio State University, Columbus, OH, United States, [10]Trudeau Institute, Saranac Lake, NY, United States.

ABSTRACT

In recent years, an increasing number of technologies and platforms suitable for short tandem repeat (STR) and single nucleotide polymorphism (SNP) genotyping have become available. However, since information directly comparing the different techniques is limited, members of the DNA Sequencing (DSRG), Fragment Analysis (FARG), and Nucleic Acid (NARG) research groups of the ABRF, launched a joint Genotyping Pilot Study. This study involved participants from all three research groups, and the technologies surveyed in the members' laboratories included microsatellite/STR-PCR analysis, direct DNA sequencing, real-time PCR, SNaPshot, Scorpion/Amplifluor, pyrosequencing, and heteroduplex analysis.

In this poster we will focus specifically on the DNA sequencing and SNaPshot aspects of the study. Results will be evaluated for the quality and concordance of the genotyping calls reported by each participant. A detailed cost analysis and a comparison of the advantages and limitations of these two methods will also be provided. The information gained during this study will be used to assess the feasibility, and optimize the implementation, of a larger study open to the genotyping community at large.

For more information on the entire study, please attend the Joint Research Group presentation, and/or visit the accompanying poster.

METHODS

Human genomic DNA from nine individuals was used to produce twelve different samples, as shown in Figure 1. Samples 1-7 were taken directly from the respective individual. Sample 7 was further subjected to 6 rounds of GenomiPhi (Pharmacia) to produce sample 8. Samples 9-12 represent a mix of DNA from individuals H & I, ratios as indicated.

Prior to analysis by DNA sequencing and SNaPshot, samples were PCR amplified to produce fragments ranging from 350-740 bp. PCR primers were tailed with M13 universal forward and reverse priming sites that were used for the subsequent sequencing reactions. A separate set of primers was used for the SNaPshot reactions. Reactions were performed in both forward and reverse directions. SNaPshot reactions were carried out either individually or multiplexed, as per each participant's preference. All samples, primers, and a standardized reaction set-up protocol was provided for each method, however, samples were run according to each participant's usual instrument parameters.

The samples were genotyped at a known polymorphic position in 4 different genes: AR, MTHFR, TNF and TYMS. The known genotype for each sample is shown in Table 1. Participants were asked to assign genotypes for each sample according to his/her preference and experience. All submitted data files were also re-analyzed by Sequence Analysis v5.2 (SeqA) (Applied Biosystems), Mutation Surveyor (SoftGenetics) and Sequencher (Gene Codes), using the default settings. The results in each case were then compared to the participant-reported genotypes.

Concordance was calculated as the percentage of samples whose called genotype was identical to the known genotype.

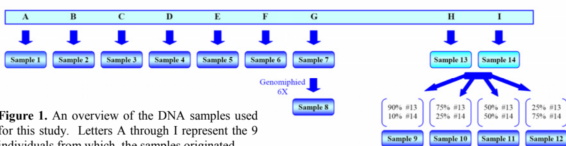

Figure 1. An overview of the DNA samples used for this study. Letters A through I represent the 9 individuals from which the samples originated.

Table 1. Known genotype for each sample.

	SNP			
	AR	MTHFR	TNF	TYMS
Sample 1	A/G	T	G	C
Sample 2	G	C/T	G	C/T
Sample 3	G	C	G	C/T
Sample 4	G	C/T	G	C
Sample 5	G	C/T	A/G	C/T
Sample 6	G	C/T	G	T
Sample 7	G	C/T	A/G	C/T
Sample 8	G	C/T	A/G	C/T
Sample 9	10A:90G	95C:5T	5A:95G	95C:5T
Sample 10	25A:75 G	87.5C:12.5T	12.5A:87.5G	87.5C:12.5T
Sample 11	50A:50G	75C:25T	25A:75G	75C:25T
Sample 12	75A:25G	62.5C:37.5T	37.5A:62.5G	62.5C:37.5T

RESULTS

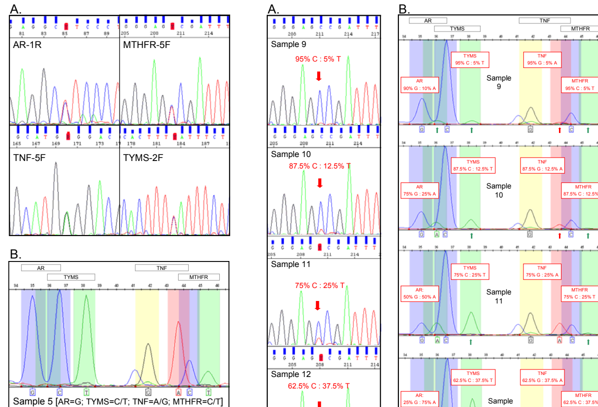

Figure 2. An example of typical results obtained with pure samples, utilizing (A) direct sequencing or (B) SNaPshot.

Figure 3. An example of typical results obtained from (A) direct sequencing on the MTHFR SNP, for mixed samples 9-12, and (B) SNaPshot analysis of all SNPs for mixed samples 9-12.

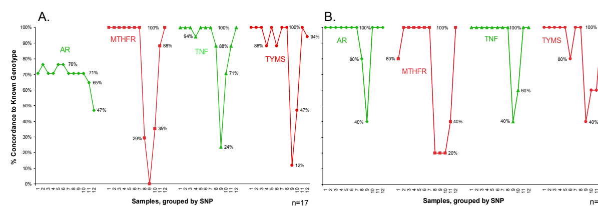

Figure 4. Accuracy of reported genotype calls by (A) direct sequencing and (B) SNaPshot. The percentage of results that are in concordance with the expected genotype from all participants for each SNP is given.

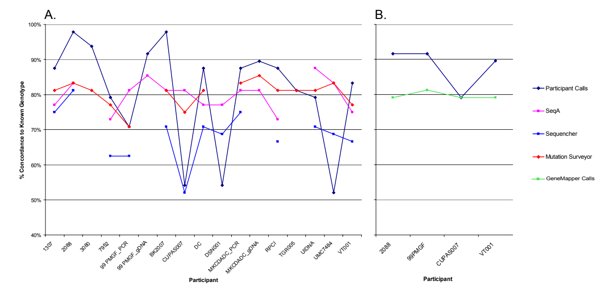

Figure 5. A comparison of different analysis methods, obtained for (A) direct sequencing and (B) SNaPshot. The percentage of results that are in concordance with the expected genotype is reported across all samples from a single participant for each analysis method.

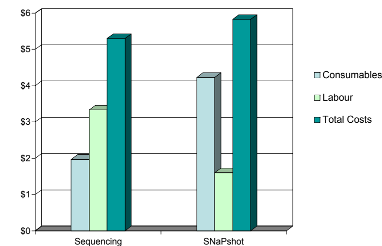

Figure 6. Cost comparison between DNA sequencing and SNaPshot. Note that the amounts are for individually-processed samples, and cost-reduction is possible with SNaPshot for multiplexed samples.

DISCUSSION

For the majority of the samples, participants were able to accurately assign the correct genotype using either DNA sequencing or SNaPshot with ≥80% accuracy. There are, however, some noteworthy exceptions. The AR locus was particularly problematic using DNA sequencing, due to the repetitive region occurring just upstream of the region of interest in the forward direction. This made it very difficult to assign a correct base-call in the forward direction, and some participants were unwilling or unable to assign a genotype to these samples based only on the reverse direction sequence. As only a single-base extension is performed in the SNaPshot reaction, this region had no bearing on the results.

As expected, sample mixtures that deviated from the normal 50/50 mix of bases found in a heterozygous individual were more problematic across all loci (samples 9-12) (refer to Figures 2 & 3). Those samples in which the 2nd base was less than ≤10% of the first (sample 9) were particularly difficult for both techniques (Figure 4).

Also of note is the "GenomiPhi effect". Samples 7 and 8 are identical for all loci, except that sample 8 was first subjected to 6 rounds of amplification by GenomiPhi prior to genotyping. It seems that there is some selection in these samples for T over C and G over A (data not shown).

Overall, SNaPshot performed better, with 35/48 samples genotyped correctly by all participants, compared to only 22/48 for DNA sequencing (Figure 4). For both methods, the participants' results were correct more often than those obtained with the various analysis software using default settings, suggesting that user experience may play a role in the outcome (Figure 5).

DNA sequencing costs per individual SNP are lower ($5.30) than those for SNaPshot ($5.83), however, the prices reported here do not reflect the ability to multiplex with SNaPshot (Figure 6). For example, in this study, it was possible to run reactions for all 4 loci simultaneously, which would bring the per sample cost down from $5.83 to only $1.46. Manufacturer claims of 10-fold multiplex would reduce costs even further.

ACKNOWLEDGEMENTS

Thank you to all of our collaborators in the FARG and NARG research groups, and to our co-workers who helped to support this endeavor.

We would also like to acknowledge Mike Klein, University of Utah, for preparation and verification of all PCR reactions used in this study.