SeqSNP: A massively parallel marker screening approach – high-speed and high-throughput genomic selection for plants and livestock

Berthold Fartmann[1], Silke Winkler[1], Samuel Arvidsson[1], Sabine Osterkamp[1], Joe Don Heath[3], **Joris Parmentier[2]** and Wolfgang Zimmermann[1]

[1]LGC, Genomics division, 12459 Berlin, DE
[2]LGC, Genomics division, Hoddesdon, EN11 0WZ, UK
[3]NuGEN Technologies Inc., San Carlos, CA 94070, USA

Presenting author email: joris.parmentier@lgcgroup.com

Introduction

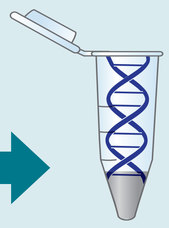
MAS (marker assisted selection) has seen massive advances in genetic gains linked to production and environmental tolerance for both plants and livestock. MAS uses markers in LD (linkage disequilibrium) with a QTL (quantitative trait locus) guided by phenotype, pedigree information and a relatively small number of SNP markers. Traits are, however, more often complex networks of genetic interactions requiring larger numbers of associated markers during selection and eventual trait fixation. Genomic selection (GS) is a method that uses high densities of markers (>~5K unique 'data points' per sample), interspersed throughout a genome, to increase association between migrating genetic material and phenotype, thus improving prediction accuracy for breeding and genotypic values. For GS analysis to be of value, input tissue mass or DNA should be minimal along with quick turnaround time, returning data with high accuracy. Increasing allele number and concomitant increased population sizes puts strains on time, throughput and resources, thus limiting the scope of this kind of analysis for many. To remove these hurdles, LGC has developed SeqSNP, which tackles all of these problems with a complete sample to data service. Here we describe the utilisation of automation and Nugen's Allegro technology to genotype 50,000 allele targets per sample in multiplex; SeqSNP also allows markers to be selected flexibly per project. We show a typical output using 96 tomato plants in a segregating population screened with 4,744 alleles across the whole genome. We directly compared a 297 marker subset tested with KASP showing high concordance of genotyping data between the two technologies.

SeqSNP process

Sample and send

Use the provided sampling kits to collect and ship samples safely or send purified DNA.

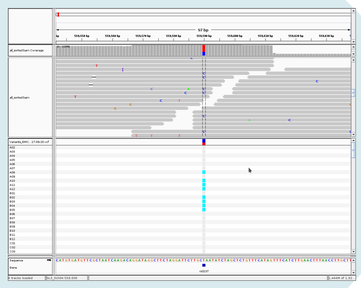
DNA purified - Multiplex libraries generated



SeqSNP

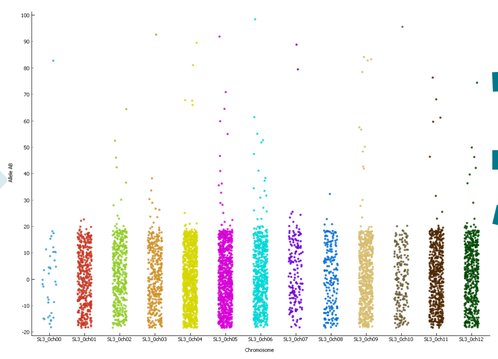
DNA extractions of 1,000s of samples at any one time followed by generation of massively multiplex SNP probe libraries.

Sequencing and variant calling



Libraries are sequenced. Data is QC'd and variant call analysis is performed and provided in an easy to understand format.

Precision analysis with high-density markers



Data delivered ready for more precise ΔZ estimates to compliment phenotypic data.

Greatly improved breeding accuracy

Use genomic selection for real-time breeding. Flexibly select probe sets depending on application/study.

Case study: 96 tomato samples and 4,744 SNPs analysed

Methods

- 96 tomato DNA samples were used from a previous genotyping study performed at our laboratory. They were stored at -20 °C for 6 months – used with the owners permission. DNA concentration ranged from 1 ng/µL to 50 ng/µL.
- To test the robustness of SeqSNP, no optimisations of probes were performed prior to testing other than homology searching for off-target mismatches using tomato reference genome ITAG SL_3.0 (Heinz-1706) (1).
- All samples had previously been genotyped at 297 positions with KASP.
- The remaining markers were selected from the Sol Genomics Network (2) and distributed at an average distance of 2 x10^5 bp or approximately 1 centiMorgan.
- We sequenced using NextSeq500 using 1 x 75 bp run mode resulting in 96 million reads and average 900,000 per sample (range 100k - 2 million reads per sample).
- Library prep to data took 3 days. All reads were processed and QC'd before mapping against the reference genome before exporting into VCF format for analysis.

Results

Average coverage 154 (min. 50 - max. 270) reads per target – we only required a minimum of 8 reads per target to successfully call variants.

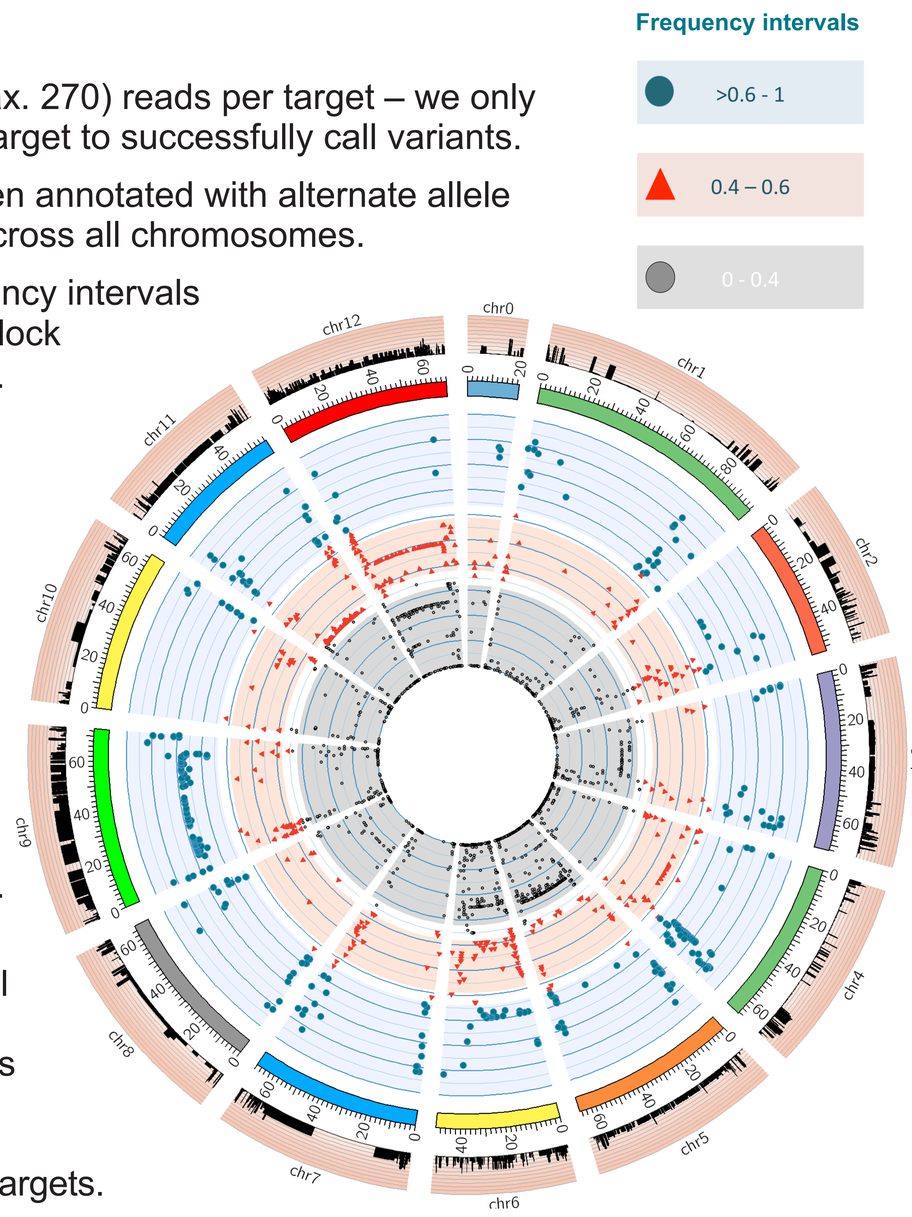
Heinz-1706 ideogram (right) has been annotated with alternate allele frequency for each position tested across all chromosomes.

Inner rings are separated into frequency intervals (see key). Outer histogram shows block clusters of variation by chromosome.

Of 4,744 targets, 6 % (284) were monomorphic but in some cases we were also able to determine multiple SNPs surrounding the target.

Of the 297 markers tested, we found that 281 (94 %) had concordance with KASP results. Of the remaining 16 markers (6 %) the majority were monomorphic indicating that surrounding sequence variations may have affected binding sites for the probes or KASP assays.

66 samples (70 %) were above 20 ng/µL reaching 98 % (4,658) of all targets. When input DNA was ≥ 30 ng/µL, 99.9 % (4739) of markers returned data. For samples with between 1-5 ng/µL SeqSNP could produce data for ~84 % (3985) of all targets.



Conclusions

Using the SeqSNP process, we were able to take a range of input sample quantities and produce high-quality genotyping data in a fraction of the time required by most NGS workflows. We were able to compare to the genotyping gold standard, KASP, with high concordance between the two sample sets. DNA concentration is very important and we recommend 30 ng/µL of high molecular weight for all projects to achieve the greatest data return. Data is analysed and provided in easy to read formats along with raw FASTAQ data. Interestingly, SeqSNP can also view variants surrounding the target up to 50 bp either side of the target SNP. Markers can be selected per project/study without fixed commitments to arrays. With short turnaround times to large quantities of data, SeqSNP will allow research and breeding to be proactively data-driven.

References
1. The tomato genome sequence provides insights into fruit evolution Nature 485, 635–641 (31 May 2012) doi:10.1038/nature11119
2. Sol Genomics Network Marker Search - solgenomics.net/search/markers