illumina®

# Sequence-Based Genotyping Brings Agrigenomics to a Crossroads

For some applications, sequence-based genotyping provides a lower-cost alternative to microarrays in performing genetic variation studies.

## Introduction

Today, agrigenomics researchers have a wide variety of technologies at their disposal for collecting genetic information. Array-based approaches to SNP screening have been the method of choice in analyzing and associating traits with regions of the genome for many plants and animals. As sequencing costs continue to decline, new approaches that leverage next-generation sequencing (NGS) technology are being developed to perform genotyping studies. The term next-generation sequencing–based genotyping (NGG), or sequence-based genotyping, encompasses genotyping methods that leverage NGS technology. NGG includes targeted, reduced representation, and hybridization-based approaches to discover and genotype SNPs, often simultaneously in many individuals or specimens. This application spotlight provides insight into different NGG methods, their benefits, and the role that conventional array technology will play in the future.

### Arrays Pave the Way in Agrigenomic Genotyping

In the late 1980s, researchers began identifying specific regions of DNA that influenced phenotypic traits in certain species. Efforts soon turned to the development of accurate and cost-effective genetic tests that could characterize the genotype of these regions in a sample. User-friendly PCR-based markers such as short tandem repeats (STRs or SSRs) were ultimately replaced with single nucleotide polymorphisms (SNPs) as the chosen marker for genotyping studies. Not only are SNPs present in high abundance within genomes, but when screened in high densities for a given species, they enable the efficient tracking of the transfer of genetic regions from parent to offspring. SNP-based assays are now routinely used to identify trait–marker associations and perform genomic selection, parentage testing, and marker-assisted selection.[1]

Optimizing marker density to detect trait associations is one of the main challenges when developing genotyping tools. Trait associations rely on detecting recombination units (haplotype blocks), making it essential to optimize marker density for the targeted “diversity population” for genotyping to be performed at an affordable cost per sample.

Many critical steps are involved in building robust genotyping arrays, including initial SNP discovery, diversity assessment, and SNP selection.[2–3] After these steps, a filtered, high-quality subset of SNPs is deployed onto a high-density genotyping platform, such as the Infinium® Assay. The cost per sample often limits the use of SNP microarrays to research applications where the screening populations are small.

Yet, many agricultural applications could benefit profoundly from genotyping, including the screening of breeding populations.[4] By leveraging genetic screening, farmers and livestock breeders could gain immediate feedback, supporting better informed breeding decisions and accelerating their return on investment (ROI). Genotyping tools with a lower cost per sample could enable genetic screening to be performed routinely on large populations, with an attractive ROI offsetting the implementation cost of the technology.

## Sequencing Advances Can Deliver More Cost-Effective Genotyping

The rapid evolution of sequencing technology has resulted in higher throughput and a lower cost per sample, often positioning NGG as a cost-effective and efficient agrigenomics tool for genotype screening, genetic mapping, purity testing, screening backcross lines, constructing haplotype maps, and performing association mapping and genomic selection.[5–7] The number of NGG methods continues to grow, with each offering the fundamental benefits sequencing provides, including reduced ascertainment bias, identification of variants other than SNPs (small insertions, deletions, and microsatellites), and an ability to perform comparative analysis across samples in the absence of a reference genome (Table 1).

## Methods of Sequence-Based Genotyping

For small genomes (eg, *Drosophila)* or high-profile research species (eg, *Arabidopsis)*, genotyping and variant screening can be completed using standard whole-genome sequencing/resequencing (WGS) methods relative to a reference. For larger genomes where funding is limited, sequence-based genotyping (or NGG) methods have been developed.

NGG advances are greatest for methods that can be performed at a lower cost than WGS. Crop researchers supporting applications in genomics-assisted breeding and genomic selection have been the primary drivers of developing lower-cost protocols.[8–9]

**Table 1: Genotyping by Sequencing Benefits and Considerations.**

| Benefits Over Array Genotyping | Explanation | Considerations |
|---|---|---|
| Low-cost genotyping now | NGG methods often use homebrew library preparation with multiplexing. Per-sample costs < $20 USD. | • Populations with low diversity (eg, cotton) will exhibit fewer polymorphisms than those with higher diversity (eg, maize). Therefore, cost per data point will be higher for low-diversity species.[10]<br>• Targeted methods of enrichment and restriction enzyme methods both require fine-tuning of coverage across highly multiplexed samples for optimal cost benefit. |
| Lower-cost genotyping in the future | NGG methods are well-positioned to leverage future sequencing cost improvements. | • As sequencing protocols and analyses are defined, published, and shared, consistency in data management and sample and gene bank tracking will be key to optimizing resources.[11]<br>• As protocols use higher levels of sample multiplexing and lower coverage per individual (eg, skim sequencing), tolerance for ambiguity in heterozygote detection must be considered.<br>• Sequencing data analysis methods, while constantly improving, are still less streamlined than array methods for data analysis. This can be a barrier for new users with species that have little genetic information (ie, no reference genome). |
| Low ascertainment bias | Ascertainment bias, especially in high-diversity species, presents challenges with array genotyping for lines that have a parental background different from the reference or SNP discovery population. Sequencing methods have a lower burden of *a priori* knowledge. | • Pulldown or amplicon methods have potential to have some bias if they depend on hybridization. Restriction site–associated methods will be bias free if restriction sites are conserved among targeted lines of interest.[5,9] |
| Increased dynamic range detection offered by sequencing in polyploid species | Higher allele dosage detection levels of sequencing over array methods enable increased allele detection sensitivity of multiple genomes in polyploid species. | • Filtering criteria for sequencing data might require adjustments for each species protocol defined.<br>• Illumina GenomeStudio® software supports automated polyploidy calling. |
| Insight into non-model genomes where no *a priori* genomic information is available | Some sequencing protocols, like those relying on restriction enzyme cut sites, can be completed in the absence of a reference genome.[5] | • Transcriptome assembly or contigs (eg, > 10 kb) can act as a putative reference from some sequencing applications.<br>• When using a reference that is distant from the targeted species (ie, use of the bovine as a reference for whale), there is some risk that a high mismatch rate for rare variants might bias toward high MAF SNPs. |

## Skim Sequencing

Low coverage or scalable/tunable skim sequencing has been demonstrated in wheat chromosomal lines as effective for SNP discovery and is useful for detailed diversity analysis, marker-assisted selection, and sequence-based genotyping.[12–13] It offers numerous advantages, including an established library preparation protocol, an established informatics application pipeline that enables SNP calling within reads rather than relative to a reference, and redundancy checks that minimize false positives.[12] The amount of data generated using skim sequencing can also be modulated by rerunning the samples to increase coverage, avoiding library preparation optimization and challenges in sample tracking.

## Enrichment

By using PCR or hybridization probes, a suite of methods can be used to isolate a specific genomic fraction by either removing unwanted components (target enrichment, Figure 1A) or selecting desired targets (targeted pulldown, Figure 1B) for subsequent sequencing.[14–15] Sequencing is focused on regions of interest, offering sufficient overlap in sequencing coverage to call SNPs reliably. Particularly in plants, these methods avoid lost sequencing space to duplicated or otherwise undesirable areas of the genome.[14]

## PCR-Based Methods

Many PCR-based genotyping methods have been developed. They include direct sequencing of PCR amplicons, long-range PCR sequencing where fragments are sheared in library preparation, and the use of molecular inversion probes to target long regions that are circularized with a ligase before amplification. These methods can pose challenges in scaling marker and sample multiplexing (multiple samples per flow cell or lane) to leverage NGS throughput and minimize costs. Challenges include accurately optimizing multiplex reaction conditions to capture all targeted regions uniformly.[14,16] There are several commercial-based PCR methods that facilitate optimal multiplex conditions, including Illumina TruSeq® Custom Amplicon.

## Hybridization-Based Methods

Hybridization-based approaches include solid substrate as well as liquid hybridization methods, using oligonucleotide specificity to bind to and isolate complementary sequences. To leverage sequencing capacity and optimize costs, these methods rely on multiplexing samples enriched using the same probe sets. Solid phase hybridization is completed after library preparation, where regions of the genome that hybridize are retained and those that do not are washed away. The more common methods of solution hybridization typically take advantage of biotinylated probes or RNA baits to facilitate capture of targets. Hybridization capture has an advantage in genotyping allotetraploids because it can enable homologous genomes to be differentiated.[17]

## Targeted Enrichment

Targeted enrichment approaches are ideal for pristine genomes (eg, bovine, rice) where there is *a priori* knowledge for regions of interest, such as markers for loss of function or trait associations used in marker-assisted selection. They are powerful methods for SNP discovery and fine mapping of recombination breakpoints. For example, researchers studying wheat used a sequence capture assay for targeted resequencing of a 2.2 Mb exon region and identified 4,000 SNPs and 129 indels suitable for differentiating between cultivated and wild wheat populations.[17]

A need to reduce costs has been the primary driver of the evolution in sequence-based approaches to genotyping. Therefore, cost-effective targeted and enrichment methods will be increasingly important to allow researchers to choose their markers of interest as more genomes are assembled and referenced.

## Restriction Enzyme Methods: RE-GBS, RAD-Seq, and ddRADSeq

The biggest advances in NGG affordability have been achieved using restriction enzyme methods of reducing the representation of the library for subsequent sequencing. Restriction enzyme GBS (RE-GBS), restriction site-associated sequencing (RAD-Seq), and ddRADSeq methods use restriction enzymes to generate fragments for sequencing. They provide a reduced, genome-wide representation with data that

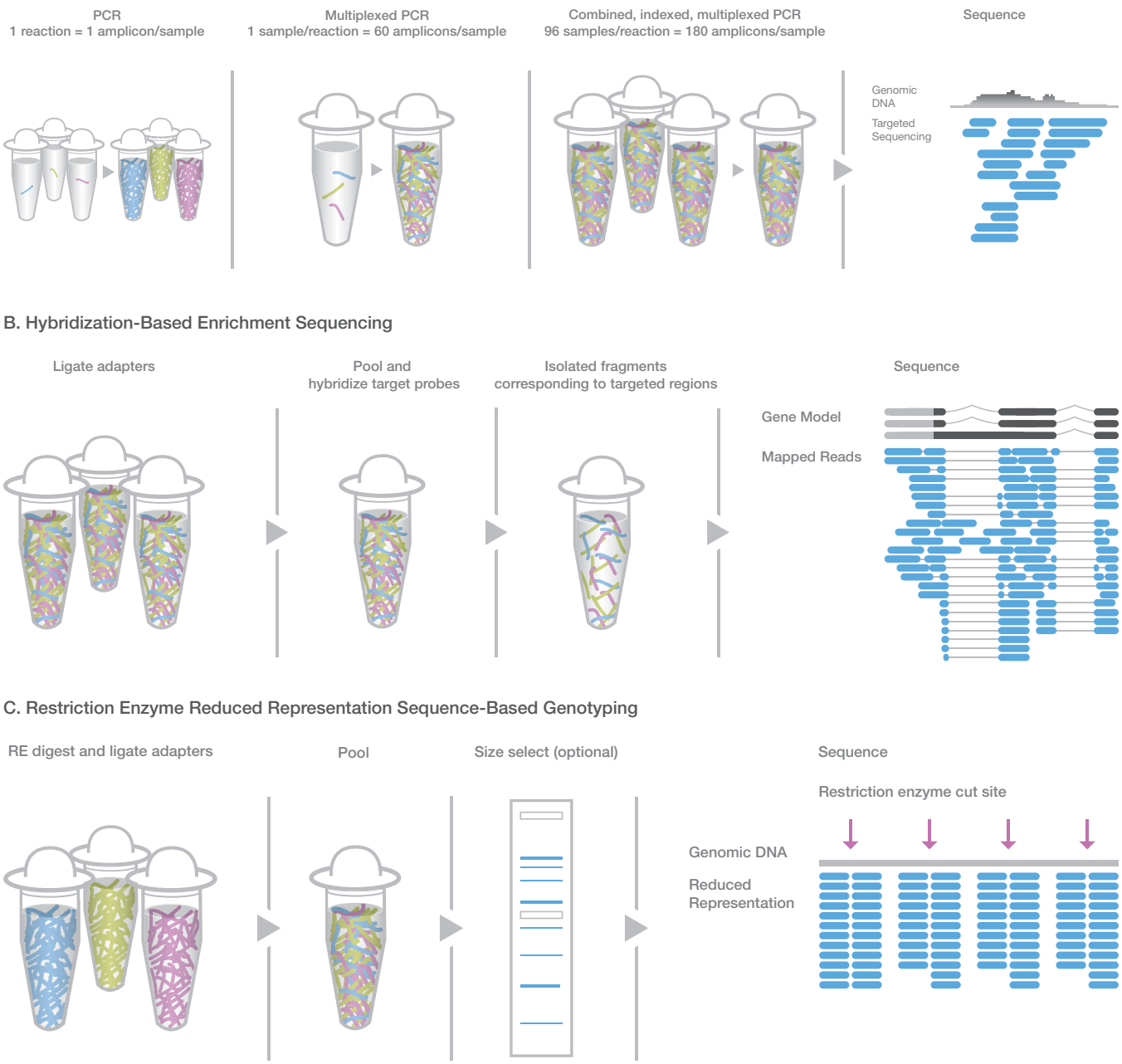


**Figure 1: NGG Methods for Discovering and Genotyping SNPs.** Amplicon-based targeted resequencing methods (panel A) adapted from Mamanova et al, 2010[16] and Liu et al, 2012.[18] Hybridization-based enrichment sequencing methods (panel B) adapted from Mamanova et al, 2010[16] and Cronn et al, 2012.[10] Restriction enzyme reduced representation sequence-based genotyping methods (panel C) adapted from Andolfatto et al, 2011.[19]

can be aligned, compared, and screened for SNP variants (Figure 1C).[5,8,9,20] NGS-compatible fragment libraries enable massively parallel and multiplexed sample sequencing, facilitating the rapid discovery and genotyping of tens to hundreds of thousands of SNPs across large populations.

RE-GBS protocols, initially established for crops like maize and wheat, have advantages in cost per sample and application in species where there is no *a priori* knowledge of the genome. The application of RE-GBS is especially powerful in mapping populations, or closely related groups of samples, such as candidates for

genomic selection. If populations are more divergent than expected or target novel species, RE-GBS protocols can require optimization (beyond published protocols) to customize coverage and minimize missing data. For example, high divergence across targeted samples can result in missing data, complicating downstream analysis, whereas low divergence can result in a lower number of detected SNPs.

The advantages of RE-GBS are many, making the protocol development for species-specific applications rewarding.[8] Reduced ascertainment bias over array-based methods, the ability to discover and characterize polymorphisms simultaneously, and the generation of valuable genetic information for a low (< $20 USD) cost per sample (excluding bioinformatics) make this a method of choice for those moving from array methods to genotyping by sequencing. RE-GBS data analysis methods are supported with open-source analysis tools (eg, TASSEL) that can be tailored for crops of interest using a command-line interface. Table 2 (reproduced from Nielsen et al, 2011[21]) shows a list of available non-commercial NGS genotyping calling software. Nielsen et al also present a workflow for converting NGS data into SNP calls (Figure 2).

RAD-Seq protocol enhancements have been primarily focused on increasing the level of multiplexing to reduce cost and eliminate expensive steps in the protocol workflow, such as random shearing and the subsequent need for end repair. Examples of methods that eliminate random shearing include MSG,[19] CRoPS,[22] and ddRADSeq.[20] The ddRADSeq method has been used to refine size selection, recovering a “tunable number of regions” distributed randomly throughout the genome at a reported library preparation cost of $5 USD per sample and input amounts as low as 100 ng of starting DNA.[20] This approach also implements a two-index combinatorial multiplex system (n*m individuals using n+m indexes), a sequence filter analysis toolkit, and a sample tracking data management tool available through a Google Docs interface. High-throughput data management and sample tracking are critical for implementing any sample screening method in breeding and germplasm tracking.[11]

Table 3 summarizes published sequence-based genotyping methods, including PCR-based, hybridization-based, and restriction enzyme approaches.

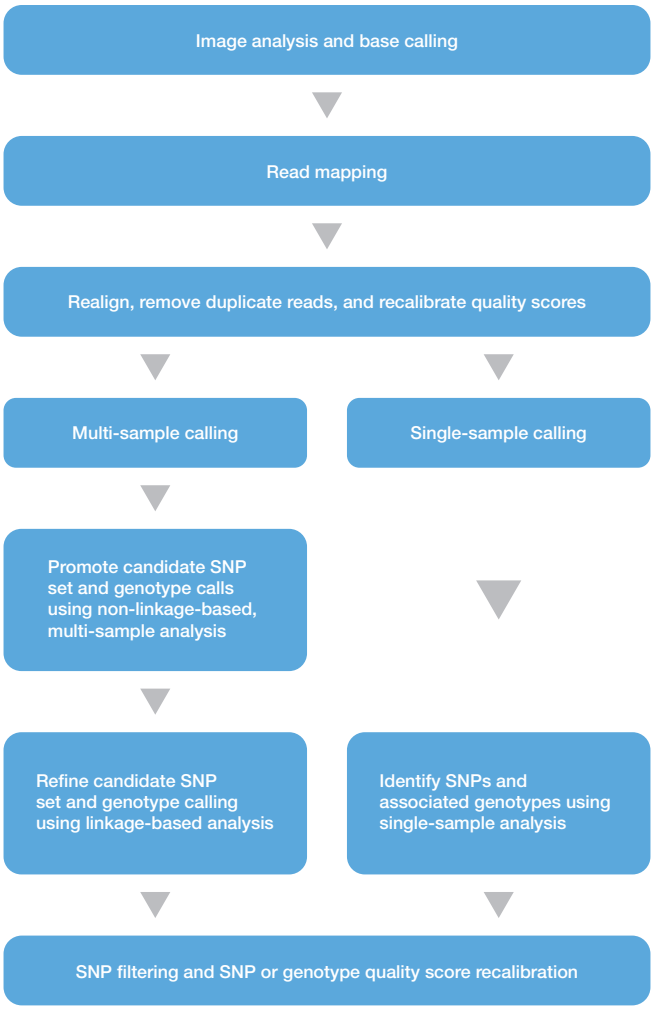


**Figure 2: Converting NGS Data Into Genotype Calls.**
Reproduced from Nielsen et al, 2011.[21] First, pre-processing steps transform NGS data into aligned reads with quality scores that indicate confidence. Next, SNP or genotype calls are made using a multi-sample or single-sample calling procedure, depending on the number of samples and depth of coverage. Finally, post-processing steps filter the called SNPs.

**Table 2: Available Non-Commercial NGS Genotype-Calling Software.**

| Software | Available From | Calling Method | Prerequisites |
|---|---|---|---|
| SOAP2 | soap.genomics.org.cn/index.html | Single-sample | High-quality variant database (eg, dbSNP) |
| realSFS | 128.32.118.212/thorfinn/realSFS | Single-sample | Aligned reads |
| Samtools | samtools.sourceforge.net | Multi-sample | Aligned reads |
| GATK | www.broadinstitute.org/gatk | Multi-sample | Aligned reads |
| Beagle | faculty.washington.edu/browning/beagle/beagle.html | Multi-sample LD | Candidate SNPs, genotype likelihoods |
| IMPUTE2 | mathgen.stats.ox.ac.uk/impute/impute_v2.html | Multi-sample LD | Candidate SNPs, genotype likelihoods |
| QCall | ftp://ftp.sanger.ac.uk/pub/rd/QCALL | Multi-sample LD | 'Feasible' genealogies at a dense set of loci, genotype likelihoods |
| MaCH | genome.sph.umich.edu/wiki/Thunder | Multi-sample LD | Genotype likelihoods |

Adapted from Nielsen et al, 2011.[21]

## Determining Sequence Depth

High-throughput microarrays (millions of SNPs for thousands of samples) have been used for years to perform genotype screening, with heterozygote detection exceeding 99.99% through optimized probe design. For the detection of heterozygosity, NGG methods depend upon sequencing depth, with increased depth resulting in increased cost per sample. When the goal is to detect parental lines that are fixed for alternate alleles, heterozygotes are infrequent and of little consequence. As a result, multiplexing can be high and coverage per sample as low as 1× to meet project goals. For applications requiring heterozygote detection, missing or ambiguous genotypes can be overcome by resequencing the library in greater depth or by using "soft" bin assignment informatics approaches[19] that facilitate imputation to missing allele states. Li et al, 2011 provide useful modeling analysis for depth of coverage needed to detect SNPs with certainty and detect heterozygotes from sequencing runs with 2×, 4×, 6×, and 30× coverage and a range of minor allele frequencies in the population.[23] Tolerance for missing data can be a critical consideration for sequence-based genotyping decisions.

Independent of the NGG method chosen, there are tradeoffs among factors of marker density, sequence depth, and degree of multiplexing for cost per sample. In RE methods, the more markers targeted (eg, 4-base over a 6-base enzyme cutter), the more fragments are created and the more sequencing is required. Improvements can be expected to lower costs of sequencing with longer reads and evenness of coverage among multiplexed individuals. All of these will allow for quicker associations between genomic regions and traits at a lower cost per sample and improved implementation of marker-assisted breeding in agriculture species.

## The Value of Arrays

While they are no longer the only solution, array methods are often still an excellent fit for screening applications, especially with well-annotated genomes where established trait associations and loss-of-function variants are known. For example, many agriculture research communities need tools for routine testing of known markers with consistent high-throughput data analysis, where volume pricing offers a cost per sample that tips the scale toward an array approach over an NGG approach. When whole communities converge on a common tool, there is an opportunity to leverage the diverse data sets and develop downstream methods for imputation and proprietary custom or Add On content. For example, lower-density arrays (< 50,000 SNPs+indels) are useful as base content for building proprietary Add On beadpools that seed companies can use to build a proprietary array with public and private marker content combined in a single chip. The combination of array- and sequence-based genotyping approaches has already contributed significant value to the dairy cattle breeding industry. As shown by the 1,000 Bull Genomes Project, merging the 2 technologies allows for highly accurate imputation for related individuals in a combined data set.[24] Illumina offers comprehensive sequencing and array solutions that can be tailored to any species.

**Table 3: Published Sequence-Based Genotyping Methods.**

| Method | Type of Method | Description |
|---|---|---|
| Amplicon sequencing | PCR based | Often used in metagenomics applications where 16S fragments are targeted. Labor intensive to amplify and tag multiple targets to optimize sequencing coverage. Difficult to scale currently to leverage sequencing output to drive down price per sample. |
| LR-PCR | PCR based | Long-range PCR (< 35 kbp, typically 3–10 kbp) can be used to target regions that then require shearing before library preparation. Challenges include equimolar pooling of sample/fragments. Tends to be a drop in coverage at the ends that can be resolved by increasing amplicon overlap to a minimum of 100 bp.[10,16] |
| Molecular inversion probes | Whole genome | Molecular inversion probes, single-stranded oligonucleotides with common linker flanked by target-specific sequences, anneal to target sequence and become circularized by a ligase. PCR amplification and products are sequenced directly. Suited for few targets and high sample numbers (> 100 samples).[25–27] |
| WGS/Genome skimming | Whole genome | Whole-genome sequencing includes DNA shearing and repair before adapter ligation. Low depth or genome skimming of whole genomes is performed for organelle (plastome, mitochondrial, or rDNA), phylogenetic/systematics, or comparative analysis. Can provide partial sequences of low-copy nuclear loci for designing PCR primers or probes for subsequent hybridization-based genome reduction approaches.[13] |
| OS-Seq | Hybridization based | Oligonucleotide-selective sequencing is a targeted genome resequencing in which the lawn of oligonucleotide primers of an Illumina flow cell is modified to function as both a capture and sequence substrate.[28] |
| Array hybridization capture (with or without C0t1) | Hybridization based | Fragment library hybridized to immobilized probe. Non-specific hybrids are removed and targeted DNA is eluted and sequenced. Can be less labor intensive than PCR amplification. Can be followed by a target-specific array that enriches for target in a reduced-complexity sample.[15,16] |
| In-solution hybridization capture (with or without C0t1) | Hybridization based | Specific probes designed to target regions of interest from sequencing library. An excess of probes over template can result in a higher hybridization than with array-based methods. Can be more amenable to scalable throughput.[17] |
| CRoPS | Restriction digest | Complexity reduction using AFLP with next-generation sequencing. Enables SNP discovery using tagged libraries of 2 or more genetically diverse samples. Uses a methylation-sensitive restriction enzyme sequenced at 5–10× redundancy. Use of homozygous lines is encouraged to enable selection of SNPs located in low- or single-copy genome sequences.[22] |
| RAD-Seq | Restriction digest | Genomic DNA digested with a restriction enzyme and a barcoded adapter is ligated to compatible sticky ends. DNA samples, each with a different barcode, are pooled, randomly sheared, and size selected (300–700 bp), and a second adapter is ligated after polishing and filling ends. A Y-adapter ensures that only RAD tags are amplified in the PCR step.[5] |
| Cornell GBS | Restriction digest | Employs unmodified adapters (ie, without the 5' phosphate group and fork), removes fragment size selection. By using a single well for genomic DNA digestion and adapter ligation, it has reduced a number of enzymatic and purification steps. Methylation-sensitive enzymes are used to avoid repetitive regions of plant genomes.[9] |
| Modified Cornell GBS | Restriction digest | Modifies the original Cornell GBS method by use of 2 complementary enzymes (a "rare" cutter and a "common" cutter) and a Y adapter where Adapter 1 and Adapter 2 are on opposite ends of each fragment.[8] |
| ddRADSeq | Restriction digest | Relies on the concept of RAD-Seq, but eliminates the random shearing. Explicitly uses size selection to recover a tunable number of regions distributed randomly through the genome. Provides an index, computational analysis tool kit, and lightweight data management tools to facilitate multiplexing of many hundreds of individuals. Major cost reductions are attributed to removal of random shearing and subsequent end repair requirements.[20] |
| GR-RSC | Restriction digest | Genome reduction based on restriction site conservation. Includes a double digest of DNA with rare and frequent restriction enzymes, labeling a recognition rare cutter site with 5' biotin using paramagnetic bead separation, adding barcode sequences using PCR, equimolar pooling of samples, and size selection using gel isolation.[29–30] |
| MSG | Restriction digest | Multiplex NGS protocol, includes a fragment size-selection step developed to identify recombinant breakpoint of many samples simultaneously at resolution sufficient for most mapping purposes. Incorporates aspects of WGS and RAD-Seq. Uses a more frequent cutter than RAD-Seq and allows ligation of adapters to many small genomic fragments in a single step. Fragment orientation is random regarding the direction of sequencing. No shearing or repair of DNA before adapter ligation.[19] |
| DArTSeq | Restriction digest | Based on genome complexity reduction using restriction enzymes followed by sequencing.[31] |

## Summary

Genotyping arrays forged the foundation of the genomics movement in agriculture, identifying SNPs associated with desired phenotypic traits that researchers have used to improve livestock breeding and crop yields. The rapid evolution of sequencing technologies is driving the development of lower-cost sequencing-based genotyping methods that will enable agrigenomics researchers to study livestock, crops, and biological systems at a level never before possible. Providing a genome-wide view, NGG methods offer the specificity, reproducibility, and efficiency needed to accelerate agricultural research, advance the development of high-value trait screening methods, and enable the swift deployment of these applications in the real world.

## References

1. Batley, J. and D. Edwards (2007). SNP applications in plants. In Oraguzie NC, Rikkerink EHA, Gardiner SE, and Silva HN (Eds.), Association Mapping in Plants, 95–102. New York, New York, USA: Springer.
2. Wang DG, Fan JB, Siao CJ, et al. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science.* 1998;280:1077-1082.
3. Van Tassell CP, Smith TPL, Matukumalli LK, et al. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nat Methods.* 2008;5:247-252.
4. Boichard D, Chung H, Dassonneville R, et al. Design of a low-density SNP array optimized for imputation. *PLoS One.* 2012;7:e34130.
5. Baird NA, Etter PD, Atwood TS, et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One.* 2008;3:e3376.
6. Kirst M, Resende M, Munoz P and Neves L. Capturing and genotyping the genome-wide genetic diversity of trees for association mapping and genomic selection. *BMC Proceedings.* 2011;5:17.
7. Metzger ML. Sequencing technologies—the next generation. *Nat Rev Genet.* 2010;11:31-46.
8. Poland JA, Brown PJ, Sorrells ME, Jannink J. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotying-by-sequencing approach. *PLoS One.* 2012;7:e32253.
9. Elshire RJ, Glaubitz JC, Sun Q, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One.* 2011;6:e19379.
10. Cronn R, Knaus BJ, Liston A, et al. Targeted enrichment strategies for next generation plant biology. *Am J Bot.* 2012;99:291-311.
11. McCouch SR, McNally KL, Wang W, Sackville Hamilton R. Genomics of gene banks: A case study in rice. *Am J Bot.* 2012;99:407-423.
12. Lorenc MT, Hayashi S, Stiller J, et al. Discovery of single nucleotide polymorphisms in complex genomes using SGSautoSNP. *Biology.* 2012;1:370-382.
13. Huang X, Feng Q, Quian Q, et al. High-throughput genotyping by whole-genome resequencing. *Genome Res.* 2009;19:1068-1076.
14. Grover CE, Salmon A, Wendel JF. Targeted sequence capture as a powerful tool for evolutionary analysis. *Am J Bot.* 2012;99:312-319.
15. Fu Y, Springer NM, Gerhardt DJ, et al. Repeat subtraction-mediated sequence capture from a complex genome. *Plant J.* 2010;62:898-909.
16. Mamanova L, Coffey AJ, Scott CE, et al. Target-enrichment strategies for next-generation sequencing. *Nat Methods.* 2010;7:111-118.
17. Saintenac C, Jiang D, Akhunov ED. Targeted analysis of nucleotide and copy number variation by exon capture in allotetraploid wheat genome. *Genome Biol.* 2011;12:R88.
18. Liu S, Yeh CT, Tang HM, Nettleton D, Schnable PS. Gene mapping via bulked segregant RNA-Seq (BSR-Seq). *PLoS One.* 2012;7:e36406.
19. Andolfatto P, Davison D, Erezyilmaz D, et al. Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Res*. 2011;21: 610-617.
20. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. Double digest RADSeq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS One.* 2012;7:e37135.
21. Nielsen R, Paul JS, Albrechtsen A, Song YS. Genotype and SNP calling from next-generation sequencing data. *Nat Rev Genet.* 2011;12:443-451.
22. Van Orsouw NJ, Hogers RC, Janssen A, et al. Complexity reduction of polymorphic sequences (CRoPS): a novel approach for large-scale polymorphism discovery in complex genomes. *PLoS One.* 2007;2:e1172.
23. Li Y, Sidore C, Kang HM, Boehnke M, Abecasis GR. Low-coverage sequencing: implications for design of complex trait association studies. *Genome Res.* 2011;21:940-951.
24. 1000 Bull Genomes Project (www.1000bullgenomes.com) Accessed 18 December 2014.
25. Hardenbol P, Baner J, Jain M, et al. Multiplexed genotyping with sequence-tagged molecular inversion probes. *Nat Biotechnol.* 2003;21:673-678.
26. Hardenbol P, Yu F, Belmont J, et al. Highly multiplexed molecular inversion probe genotyping: over 10,000 targeted SNPs genotyped in a single tube assay. *Genome Res.* 2005;15:269-275.
27. Dahl F, Gullberg M, Stenberg J, Landegren U, Nilsson M. Multiplex amplification enabled by selective circularization of large sets of genomic DNA fragments. *Nucleic Acids Res.* 2005;33:e71.
28. Myllykangas S, Buenrostro JD, Natsoulis G, Bell JM, Ji HP. Efficient targeted resequencing of human germline and cancer genomes by oligonucleotide-selective sequencing. *Nat Biotechnol.* 2011;29:1024-1029.
29. Maughan PJ, Yourstone SM, Jellen EN, Udall JA. SNP discovery via genomic reduction, barcoding and 454-pyrosequencing in Aramanth. *Plant Genome.* 2009;2:260-270.

30. Maughan PJ, Yourstone SM, Byers RL, Smith SM, Udall JA. Single-nucleotide polymorphism genotyping in mapping populations via genomic reduction and next-generation sequencing: proof of concept. *Plant Genome.* 2010;3:166-178.

31. Sansaloni C, Petroli C, Jaccoud D, et al. Diversity arrays technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. *BMC Proceedings.* 2011;5:P54.

32. Gore MA, Wright MH, Ersoz ES, et al. Large-scale discovery of gene-enriched SNPs. *Plant Genome.* 2009;2:121-133.

33. Deschamps S, la Rota M, Ratashak JP, et al. Rapid genome-wide single nucleotide polymorphism discovery in soybean and rice via deep resequencing of reduced representation libraries with the Illumina Genome Analyzer. *Plant Genome.* 2010;3:53-68.

34. Straub SCK, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A. Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. *Am J Bot.* 2012; 99:349-364.

**Illumina** • 1.800.809.4566 toll-free (U.S.) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

FOR RESEARCH USE ONLY


Pub. No. 370-2012-015 Current as of 22 December 2014

illumina®