illumina®

Flexible MiSeq® System Supports a Broad Range of Agrigenomics Applications

The compact and economical MiSeq system delivers next-generation sequencing results in hours rather than weeks, providing a higher throughput alternative to CE for agrigenomics studies.

Introduction

Agrigenomics researchers have made significant advancements in our understanding of plants and animals, increasing crop yield and improving livestock breeding. Whole-genome sequences have been generated for many of the major plants and animals, such as corn, wheat, cotton, bovine (dairy and beef), and chicken, supporting research advances to meet the world's food, feed, fiber, and fuel demands. While whole-genome sequencing is performed with next-generation sequencers, many other agrigenomics sequencing applications such as clone checking, amplicon sequencing, and targeted transcript sequencing are routinely conducted using capillary electrophoresis technology (CE). More than 50-years old, CE technology was first used in the 1980s for DNA analysis and involves lengthy and complex workflows, taking several weeks to generate data. The MiSeq® system (Figure 1), with its faster turnaround time and simplified workflows, offers a cost-effective alternative for performing *de novo* and resequencing applications for small to mid-size genomes, RNA sequencing, and epigenetics studies.

Streamlined Workflow and Faster Data Analysis than CE

The MiSeq system offers the first end-to-end sequencing solution, integrating cluster generation, amplification, sequencing, and data analysis into a single instrument. Its small footprint—approximately two feet square—fits easily into a laboratory environment. The MiSeq system employs Illumina sequencing by synthesis technology, the most widely used, proven next-generation sequencing chemistry with over 2,000 publications to date. As with other Illumina sequencers, the MiSeq system is powered by TruSeq® technology, delivering the highest data integrity, with the highest yield of error-free reads and the most base calls above Q30.

Figure 1: MiSeq System

The compact MiSeq system is the ideal platform for many agrigenomics sequencing applications.

The MiSeq system features enhanced fluidics architecture, enabling a five-fold decrease in chemistry cycle time to provide results in hours, rather than the weeks required by CE. Preparing a sequencing library takes just 90 minutes, with clonal amplification and sequencing completed within as little as 4.5 hours. On the integrated instrument computer, data analysis from quality-scored base calls to variant calling and alignment can be completed in less than 2 hours with no user intervention (Figure 2). This data can be stored, analyzed, and shared with BaseSpace™, an Illumina secure, cloud-based resource, enabling unparalleled collaboration, access, and security. This is in sharp contrast to CE's limited data analysis capability that adds to its cost of operation and strains the personnel resources of most agrigenomics research groups.

Uniquely Suited for a Variety of Agrigenomics Sequencing Applications

Despite its size, MiSeq is a powerful sequencer capable of performing demanding sequencing applications such as small genome *de novo* sequencing, targeted sequencing to analyze a discrete section of the genome, RNA sequencing to elucidate gene and protein function, microRNA sequencing to study the role of these short nucleotide sequences in regulating development, and epigenetics studies to determine the role of DNA methylation in plant and animal development. Continued MiSeq performance enhancements will expand these applications to include resequencing of larger organisms, as well as support small genome sequencing at higher coverage. More importantly, its high accuracy, simpler workflow, and minimal hands-on time will make the MiSeq system a viable CE alternative for performing valuable agrigenomics applications, including:

Library Quality Control

Validating a nucleic acid library before sequencing is an important quality control step that ensures the value of downstream data. A poor quality library can undermine the success of large-scale sequencing, and lead to costly and time-consuming repeat experiments. The MiSeq system can perform a front-end library check using 2 × 25 bp sequencing, enabling researchers to assess the evenness of coverage, insert size, diversity, and GC content of the library.

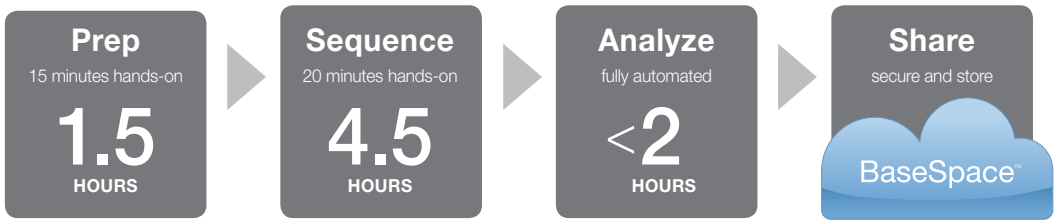
Clone Checking

MiSeq is capable of performing fast analysis of plasmid constructs used to clone DNA. In less than a day, plasmid constructs can be sequenced, confirming the presence and integrity of the inserted piece of DNA.

De Novo and Resequencing of Small to Mid-Sized Genomes

In a single workday, the MiSeq system can accurately sequence small whole genomes (< 20 Mb), such as those of pathogens and some

Figure 2: MiSeq Workflow



The MiSeq system's revolutionary workflow enables fast turnaround times for the most demanding agrigenomics sequencing applications.

plants. For example, MiSeq resequenced the 5.2 Mb *Bacillus cereus* genome in eight hours (40 minutes of hands-on time), yielding 175 Mb of data aligned to ATCC10987 with a mismatch rate of 0.06%. Its 5.4 million reads captured > 98% of the *B. cereus* genome, with average coverage of 30×.

The MiSeq system is also an efficient and economical option for performing small genome *de novo* sequencing. *De novo* sequencing of 2.8 Mb methycillin resistant *Staphlococcus aureus* (MRSA) provides an example of the speed and accuracy of the MiSeq system. The project was completed in 22.5 hours (less than 3 hours hands-on time) and yielded *de novo* 2.78 Mb assembly with N50 of 72.9 kb and max contig of 247 kb. Reads were aligned to 2.8 Mb EMRSA15 reference with an average mismatch rate of 0.12%.

Parentage verification

Short tandem repeats (STRs) have been used for many years as genetic markers for identifying animals and understanding the relationship of offspring to parents. Typically, multiple markers are necessary to increase the probability of identifying the true parent. Currently, the use of STRs is being replaced with SNPs, which offer a higher power of discrimination. The multiplex and throughput capabilities of the MiSeq system enable faster parentage studies, providing the long reads necessary to sequence through microsatellites, while simultaneously genotyping parentage SNPs.

RAD-Seq

Restriction site-associated DNA marker sequencing (RAD-Seq) is a genotyping method that enables the interrogation of a fraction of a target genome (0.1% to 15%). Rather than interrogating every base pair, it screens short fragments of DNA (RAD tags) that flank the recognition sites of particular restriction endonucleases for genetic variation. The MiSeq system can be used to sequence RAD tags for SNP discovery, enabling the genotyping of pooled populations for further bulk segregant analysis and multiplexed genotyping of individuals for fine-scale mapping to correlate genotypes with phenotypes.

New Reagents to Support Additional Agrigenomics Applications

Amplicon screening and targeted transcript sequencing has increased in human genomic studies, thanks to next-generation sequencers such as the Illumina HiSeq® systems that offer the throughput necessary to efficiently perform these applications in a matter of days. TruSeq reagents to support these studies on the MiSeq further broaden the agrigenomics applications that can be performed with this compact, economical system.

Amplicon sequencing

Amplicon sequencing can be used for transgene (GMO) detection to ensure the presence of intended, and the absence of unintended, events; genetic purity to assess the similarity of all seeds in a seed lot; or DNA barcoding to identify species. The MiSeq system and TruSeq Custom Amplicon enable these studies to be performed faster and more economically. Amplicon sequencing on 96 samples for 96 targets using the MiSeq system takes less than three days; more than 10× faster than sequencing the same samples using CE (3–4 weeks).

Targeted transcript sequencing

Studying relevant subsets of the transcriptome can provide novel insights into changing expression levels that occur in development, and during disease and stress conditions. Transcriptome studies can identify novel transcripts and sequencing alterations. The MiSeq system enables targeted transcript sequencing studies to be performed in a fraction of the time of CE, offering the bandwidth to multiplex samples at high read depths.

Summary

The MiSeq system offers fast results, simplified workflows, with a minimum of hands-on time, making it an efficient, cost-effective alternative to CE for many agrigenomics sequencing applications. Powered by the same TruSeq technology found in all Illumina sequencing systems, the MiSeq can also perform small genome *de novo* sequencing and resequencing, delivering high-quality data with the highest yield of error-free reads and the most base calls above Q30. Within its small footprint, MiSeq integrates cluster generation, amplification, sequencing, and data analysis, offering a single instrument solution that fits easily into today's agrigenomics laboratory.

Learn More

Go to www.illumina.com/miseq to learn more about the next revolution in sequencing.

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. 770-2011-035 Current as of 12 January 2012

illumina®