NEW ERA AGRICULTURE MAGAZINE

Exome Sequencing in Crop Improvement

V. Thakur [1], S. Sharma [2] and S. Thakur [3]

Introduction:

Exome sequencing is a powerful genomic technique that selectively targets and sequences the protein-coding regions of an individual's genome. These regions, known as the exome, constitute only about 1-2% of the entire genome but harbor the majority of disease-causing variants (Turner et al. 2009). By focusing on these critical regions, exome sequencing provides a cost-effective approach for identifying genetic variations associated with various diseases and traits. (Warr et al. 2015). Traditionally, the exome has been defined as the region that contains all of the exons of the genome's protein-coding genes. Depending on the species, this sequence can make up 1% to 2% of the total genome. It may also be expanded to target particular potential loci and functional nonprotein coding components (such as microRNA, long intergenic noncoding RNA, etc.). (Turner et al. 2009). A key strategy for simplifying genome analysis is through a method known as Sequence Capture or Targeted Sequencing. This technique involves focusing on specific genes, regions within genes, or even the entire protein-coding portion of the genome, known as the exome. Sequence Capture can be accomplished using one of three approaches: hybridization-based sequence capture, PCR-based amplification, or selective circularization, as outlined by (Dahl et al. 2005; Hodges et al. and Gnirke et al. 2009). Exome capture technology falls into two primary categories: array-based and solution-based. In solution-based, whole-exome sequencing (WES), DNA samples are fragmented and biotinylated oligonucleotide probes (baits) are used to selectively hybridize to target regions in the genome. Magnetic streptavidin beads are used to bind to the biotinylated probes, the nontargeted portion of the genome is washed away, and the polymerase chain reaction (PCR) is used to amplify the sample, enriching the sample for DNA from the target region. The sample is then sequenced before proceeding to bioinformatic analysis. Array-based methods are similar except that the probes are bound to a high-density microarray.

V. Thakur [1], S. Sharma [2] and S. Thakur [3]

[1,2] Department of Genetics and Plant Breeding, CSK HP Krishi Vishvavidyalaya, Palampur-176062, HP, India.

[3]MS Swaminathan School of Agriculture, Shoolini University, Kasauli Hills, Solan, 173229, HP, India.

The array-based method was the first to be used in exome capture (Albert et al. 2007), but it has largely been supplanted by solution-based methods, which require less input DNA and are consequently potentially more efficient; however, studies by Asan et al. (2011) and Bodi et al. (2013) found that Noblemen's Sequence Capture Array performed better than the solution-based alternatives in low GC content regions; had high sensitivity and read mapping rates; and single-nucleotide polymorphism (SNP) detection from these reads was more specific to the target region. This suggests that a niche may remain for the older technology. Array-based capture has been used successfully and accurately to identify rare and common variants and identify candidate genes for monogenic diseases in small cohorts (Ng et al. 2009); however, the array-based methods are less scalable owing to the limitation of the number of probes that can be accommodated on the array and additional equipment and time required to process the microarrays. One of the earliest applications of exome sequencing was in the identification of rare Mendelian disorders, where a single mutation in a protein-coding gene can lead to a specific phenotype. By sequencing the exomes of affected individuals and their family members, researchers can pinpoint the causal genetic variants responsible for the disease phenotype (Sarah *et al.* 2010). Moreover, exome sequencing has been instrumental in advancing our understanding of complex diseases, such as cancer. By comparing the exomes of tumor cells with those of healthy tissues from the same individual, researchers can identify somatic mutations driving tumorigenesis and tumor progression. This information not only aids in the diagnosis and prognosis of cancer but also informs the development of targeted therapies tailored to the individual's genetic profile. (Bousquet et al. 2016). In addition to disease research, exome sequencing has been applied to population genetics studies to elucidate the evolutionary history and genetic diversity of different human populations. By sequencing the exomes of individuals from diverse ethnic backgrounds, researchers can identify population-specific genetic variants and infer migration patterns and demographic history. (Glotov et al. 2023) Exome sequencing is a tool that can be used to assess natural evolution in plants, study host-pathogen interactions, and improve crop production. Exome sequencing focuses on the genomic fraction that encodes for mRNA and a phenotype, which can explain the molecular origin of genetic variation. Coding sequences make up only 1–2% of a genome, depending on the species, and contain a high level of functional variants and low repeat content. (Hashmi et al. 2015). Next-generation

sequencing has revolutionized molecular breeding, introducing innovative concepts and methodologies. Exome sequencing, in particular, has emerged as a crucial tool with diverse applications in plant biology. It enables researchers to assess natural evolutionary processes in plants, investigate interactions between hosts and pathogens, and enhance crop production by facilitating the interpretation of allelic variation in relation to phenotype. This is possible because exons, the coding regions of genes, play a pivotal role in determining traits. Therefore, exome sequencing provides valuable insights into the genetic basis of traits essential for crop improvement and sustainable agriculture. (Vrashney et al. 2012). Overall, exome sequencing has revolutionized biological research by providing a comprehensive and cost-effective means of identifying disease-causing genetic variants, elucidating the genetic basis of complex traits, and unraveling the evolutionary history of human populations. (Goh et al. 2012).

HOW TO APPROACH PLANT EXOME

There are two technological options for hybridization-based exome capture: in solution capture (Gnirke et al.2009) and array/chip-based capture (Hodges et al.2007; Okou et al.2007). Both strategies rely on specially made probes or baits to enrich targets from sequencing libraries; however, solution captures need a higher concentration of baits than DNA library, and array captures need a greater quantity of library than probes to accomplish enrichment.

Exome sequencing can be split into two stages: high throughput DNA sequencing is the second phase, and the first involves using probe hybridization to identify a subset of DNA that encodes for a protein (i.e., target enrichment). When designing an exome capture experiment, it is important to take into account a number of factors, including the mode and quality of processing the input DNA sample, the number of targets, the coverage depth for each target, the probe design and GC content, the expected enrichment efficiency, the sequencing technology used, the biological system studied, etc. (Zhou and Holliday. 2012). Of all these variables, depth of sequence coverage is crucial for obtaining high sequencing reliability for exome capture investigations. To properly confirm the variation found, coverage of at least 30X or more is needed (Winfield et al.2012). The factors that determine coverage depth are as follows: probe specificity; genome size; presence of orthologs and paralogs; ploidy level; homology and heterozygosity level; and probe characteristics, such as the genomic regions from which the probe is designed, i.e., whether conserved or unique (Grover et al.2012). For the production of exome

libraries, there are well-proven methods as well as readily used exome kits (NimbleGen, Affymetrix) and user-dependent customizations available for numerous crop species, including soybean, wheat, barley, and maize. Off target capture and the capture of extremely repetitive sequences are recognised to be problems with hybridization capture techniques. In order to avoid sequencing unsuccessfully enriched libraries, a low-cost PCR-based technique has recently been developed. It uses multiplex ligation dependent probe amplification (MLPA probes) of an enriched library followed by capillary electrophoresis to validate the exome library in terms of enrichment efficiency (Klonowska et al.2016).

EXOME SEQUENCING: ADVANTAGE OVER CONVENTIONAL METHODS

Whole exome sequencing (WES) provides certain advantages over similar techniques like as RNA sequencing, which can be biased by transcript quantity and also depends on tissue and stage. Exome sequencing allows for the investigation of genes and alleles likewise, the least amount of control a researcher may apply to particular target regions restricts the sequencing of gene-rich regions by methylation filtration approach, particularly when genotyping different germplasm in parallel. Exome analysis, on the other hand, uses a probe-based enrichment strategy that focuses on particular regions rather than arbitrary euchromatic regions. Other approaches, such as high Cot DNA selection and EST sequencing (Barbazuk et al. 2005), are less effective at delivering particular sequences in a targeted way (Fu et al.2010).

When WGS is not feasible or required, exome sequencing is a significant help as a supplement to later. It offers several advantages over Whole Genome Sequencing (WGS) including: (a) multiplexing of more samples for a given sequencing space; (b) sequencing of targeted informative regions reduces the complexity of data analysis; (c) identification of functional molecular markers; (d) an effective platform to collect genomic data at population level for evolutionary and phylogeny studies; (e) existing databases provides functional context for exome identified SNPs through transcript/exon annotation in contrast to SNPs identified outside coding regions through genotyping based sequencing which are not easily annotated (Scheben et al. 2017); and (f) provides high coverage for identification of low frequency sequence variations. Exome sequencing is a well-established technique for focused resequencing of the gene space, particularly for phylogenetics and divergence studies (Bamshad et al. 2011). This technique improves the sequencing depth of targeted regions by focusing on evolutionary more

conserved regions, such as particular genes or genomic regions, as opposed to surveying the entire genome. Exome sequencing remains a viable alternative technology for trait or target specific studies, even though high-depth WGS is regarded as the gold standard for sequencing and resequencing because it can access and probe all regions of genomes. For species that are unsuitable for WGS research, such sugarcane (Song et al. 2016), pine (Neves et al. 2013), and black cotton wood (Zhou and Holliday. 2012), it has shown promise in producing genome-wide data. Exome sequencing offers a comprehensive perspective of gene regulation by providing insights into coding, intronic, UTRs, and putative regulatory regions, as hybridized captured pieces tend to be longer than the probes employed to retrieve them.

SEQUENCING: TOOLS AND TECHNIQUE

A. FIRST GENERATION SEQUENCING

Molecular biology revolves around the sequencing of the genetic code, which is essential for determining the molecular causes of different types of traits and illnesses. Allan Maxam and Walter Gilbert conducted the first attempts at sequencing DNA in 1976–1977. They created the chemical sequencing method, commonly known as the modification–dependent cleavage method (Maxam and Gilbert. 1977). Sanger's dideoxy chain termination sequencing method is another early technique that gained a lot of traction and is currently being used (Sanger and Coulson. 1975). In order to apply this approach, a developing nucleotide chain that prevents chain extension due to a lacking 3'-OH group must incorporate a dideoxy nucleotide. Using nucleotides containing radioactive phosphorus, the bands can be seen on an SDS-PAGE. Sanger's sequencing technique used less hazardous reagents and was more effective. Eventually, Sanger's sequencing method was used to automate sequencing by using primers that had a fluorescent dye labelled at the 5' terminus. Automation made DNA sequencing easier, more dependable, and more affordable, which is why it was such a significant advancement in the profession. Using all four ddNTPs labelled with distinct fluorescent dyes in a single reaction was one of the later improvements made to Sanger's dideoxy sequencing method (Smith et al., 1986). Another important development was the resolution of the amplicons using capillary electrophoresis, which was followed by LASER-based detection of the integrated ddNTP in amplicons. Because it is substantially faster than a four-reaction sequencing method, this approach is preferred for automated sequencing instrumentations now in use. Even with its great precision, this method's poor quality at the first 20–50 bases,

inability to read sequences after 600–1000 bases clearly due to poor size resolution of large-sized DNA fragments by capillary electrophoresis, nonspecific primer binding, and secondary structures in DNA are some of its limitations for high-throughput sequencing. Prior to sequencing, smaller pieces were first cloned into plasmids in an attempt to sequence bigger segments. Nevertheless, these techniques also resulted in vector sequence contamination. Vector contamination problems were reduced by advances in bioinformatics and sequencing techniques that used PCR-cloned fragments. Vector contamination problems were reduced by advances in bioinformatics and sequencing techniques that used PCR-cloned fragments. Longer readings were further enhanced by more recent techniques utilizing coupled amplification (Murphy et al., 2005; Sen Gupta and Cookson, 2010).

The complete hypoxanthineguaninephos phoribosyltransferase (HPRT) gene was sequenced using paired-end sequencing using the first automated fluorescent DNA sequencer (Edwards et al., 1990; Pareek et al.2011). In 1996, the first commercial Sanger's dideoxy sequencing. Sanger's dideoxy method is the basis of the first generation high-throughput sequencers from Applied Biosystems. The template sequence is determined by fluorescent signals from incorporated fluorescent dideoxy nucleotides.

B. Next-Generation Sequencing Technology

In recent years, a number of innovative automated sequencing techniques have been developed and brought to market in response to the constantly increasing need for big sequences. Next-generation or second-generation sequencing platforms are those types of ultrahigh-throughput sequencing platforms that do not make use of Sanger's dideoxy chain termination sequencing method. Among the commercially available next-generation sequencers are the following: Genome Sequencer from Roche/454, Genome Analyzer from Illumina/Solexa, SOLiD! from Applied Biosystems, and Polonator from Dover Systems.

1. Pyrosequencing

The pyrosequencing platform (http://www.lifesequencing.com), which is based on the sequencing-by-synthesis (SBS) approach, is utilised by Roche/454 genome sequencers. In order to use this technology, the DNA library must first be prepared using fragmented genomic DNA (gDNA) (300–500 bp pieces) (Margulies et al. 2005). Subsequently, the fragments undergo blunting and ligation at both ends using brief adaptors that function as primers for the subsequent fragment amplification. The immobilisation of

amplicon on streptavidin-conjugated beads is made possible by a 5'- biotin tag on one of the adapters. In order to create the single-stranded template DNA (sstDNA) library, nick repair releases the nonbiotinylated strand. It is also possible to pool together up to 12 samples by using distinct barcoded adaptors. Following titration to achieve the ideal quantity and quality, the sstDNA library is immobilised onto emPCR (emulsion-based PCR) - useable beads! Next, PCR reagents and the library-immobilized beads are emulsified in water-in-oil emulsions. After that, clonally amplifying each bead containing a single amplicon yields millions of copies of that same single amplicon. In a PicoTiterPlate! apparatus, sstDNA library beads are incubated with polymerase and enzyme beads that contain immobilised ATP sulfurylase and luciferase enzymes. By using this method, it is guaranteed that each well has a single sstDNA library bead. After that, PicoTiterPlate! is put into the pyrosequencing apparatus, where a fluidics system layers the plate in sequential order with sequencing chemicals and individual nucleotides to ensure millions of copies of sstDNA are sequenced in parallel. Next, PCR reagents and the library-immobilized beads are emulsified in water-in-oil emulsions. After that, clonally amplifying each bead containing a single amplicon yields millions of copies of that same single amplicon. In a PicoTiterPlate! apparatus, sstDNA library beads are incubated with polymerase and enzyme beads that contain immobilised ATP sulfurylase and luciferase enzymes. By using this method, it is guaranteed that each well has a single sstDNA library bead. After that, PicoTiterPlate! is put into the pyrosequencing apparatus, where a fluidics system layers the plate in sequential order with sequencing chemicals and individual nucleotides to ensure millions of copies of sstDNA are sequenced in parallel. In a nucleotide run, polymerase will lengthen the expanding polynucleotide chain upon encountering a complementary nucleotide, resulting in the release of an inorganic pyrophosphate (PPi). On the enzyme bead, PPi functions as a substrate for ATP sulfurylase and is transformed into ATP with Adenosine 5'-phosphosulfate (APS). The luciferase enzyme on the enzyme bead uses the ATP thus created to change luciferin into oxyluciferin, which produces light.The sequencer then uses a CCD camera to detect the luminescent light, and the signal strength there shows how many nucleotides were included in a single flow. Sequence assembly is done by bioinformatic analysis. Read lengths of up to 1 kb are now possible because to advancements in the reaction chemistry of GS FLX devices.

2. **Reversible terminator-based sequencing**

In 2007 (project Jim), the genomic sequence of James Watson was found using this sequencing platform (Wheeler et al.2008). Moreover, this approach was successful in sequencing the Neanderthal genome (Green et al.2010). Oligo-primed DNA fragments are extended by simultaneous incubation with all four nucleotides and polymerase within channels of flow cells. Strand extension for sequencing and the creation of a flow cell containing ~108 clusters are caused by bridge amplification. About a High-throughput DNA sequencing technology has become more accessible to a wider range of researchers due to its declining cost and the availability of bench top sequencers. Despite all of the advancements in exome sequencing technology, there are still a number of obstacles that prevent this technology from being used to improve crops. The need for reliable, stable, and well-documented computer techniques and software solutions that researchers can use for extensive genomic sequence data analysis is expanding. For most academics, analyzing the enormous volumes of data produced by the technology still presents a formidable challenge. Despite the development of numerous data analysis tools, their usability remains uncertain. For the alignment tools to handle huge amounts of short reads, more improvement is still required thousand copies of the identical template are present in each cluster. SBS technique is used to sequence the clustered templates using reversible terminators that are attached with a detachable fluorescent dye.A fluorescently labelled terminator is imaged during the addition of each dNTP to determine the sequence, and the terminator is cleaved off to permit more base incorporation. High sensitivity fluorescence detection is achieved by using total internal reflection (TIR) optics when combined with laser excitation. The two strands can be sequenced sequentially due to this technology. After the first strand is sequenced, the template is recreated using the paired-end module, which permits template regeneration (the complimentary strand of the original template) and amplification, to enable a second round of sequencing (75 + base read) from the opposite end.

C. Next-Next Generation Sequencing Technology

Further advancements in sequencing technologies have provided single molecule sequencing, also known as next-next or third generation sequencing technology (TGS). Some of the TGS sequencing platforms are discussed below.

1. **Single molecule, real-time (SMRT) sequencing**

PacBioRS sequencers from Pacific Biosciences (http://www.pacificbiosciences.com/products/s

mrt-technology) use the SMRT sequencing platform. The creation of an SMRTbell library is necessary for this technology (Korlach et al.2010). In essence, this is a real-time SBS technique that detects nucleotide incorporation (Eid et al.2009). To produce the library, DNA fragments are generated, then end repair and hairpin adapter ligations are performed to obtain circular SMRT DNA templates. There is a range in the fragment size from 250 bp to 10 kb. Following the binding of the library to DNA polymerase, DNA sequencing is carried out on SMRT cells, which have a collection of about 75,000 zero-mode waveguides (ZMWs). ZMWs are holes in a 100 nm metal film placed on a glass substrate that are tens of nanometers in size. A single DNA polymerase is present on the glass surface of each ZMW, and a fluorescence pulse is detected whenever a fluorescently labelled nucleotide enters the bottom 30 nm of the ZMW. The fluorescence colour identifies which nucleotide is incorporated, and a longer pulse width, compared to free diffusion, indicates nucleotide incorporation into the DNA. The DNA polymerase cleaves the nucleotide's terminal phosphate linked fluorophore (rather than the typical base linked) before translocating to the next base on the template. Additionally, this technology can identify DNA methylation, which is indicated by an interpulse duration that is five times longer than that of the unmethylated base (Flusberg et al. 2010). Longer read lengths—typically more than 3000 bp—made possible by this technology facilitate simpler mapping and assembly. The typical instrument time is close to thirty minutes, and the sequencing reactions happen quite rapidly.

4. Nanopore sequencing

Oxford Nanopore Technologies' (www.nanoporetech.com) recently released GridION and miniature MinION sequencers use this sequencing platform. The technique of nanopore sequencing, commonly referred to as "Strand sequencing," uses changes in the ion current to identify specific nucleotide sequences when the DNA strand is passed through a protein nanopore implanted into a membrane, one base at a time. The DNA molecule's length, shape, and size all affect how much current changes. The current flow varies noticeably in response to each of the four nucleotides that pass, making it possible to identify each nucleotide. This method can sequence a DNA fragment's sense and antisense by forming hairpins at the end of the fragment. These instruments use an array chip that contains hundreds of thousands of nanopores. Without causing any damage to the DNA, this platform can sequence individual DNA molecules in real time at a very cheap cost and quick speed. The company is expected to start marketing 8000 nanopore-

containing nodes by 2013 that will be have the capability to sequence the entire genome of a human being in 15 minutes. However, the technology currently has a 4% error rate that needs to be significantly reduced.

D. Technical Challenges to High-Throughput Sequencing

High-throughput DNA sequencing technology has become more accessible to a wider range of researchers due to its declining cost and the availability of bench top sequencers. Despite all of the advancements in exome sequencing technology, there are still a number of obstacles that prevent this technology from being used to improve crops. The need for reliable, stable, and well-documented computer techniques and software solutions that researchers can use for extensive genomic sequence data analysis is expanding. For most academics, analyzing the enormous volumes of data produced by the technology still presents a formidable challenge. Despite the development of numerous data analysis tools, their usability remains uncertain. For the alignment tools to handle huge amounts of short reads, more improvement is still required.

Exome capture Platforms

1. Among all the platforms, NimbleGen'sSeqCap EZ Exome Library has the highest bait density and covers the target region with short (55-105 bp), overlapping baits (Clark et al. 2011). According to Clark et al. (2011), this strategy has been shown to be an effective way for enrichment with the least amount of sequencing required to cover the target region and detect variants with sensitivity. It also has a high level of specificity, displaying fewer off-target reads than other platforms. Significantly, compared to the other platforms, this bait design has been shown to exhibit higher genotype sensitivity and more uniformity of coverage in hard-to-sequence regions, like those with high GC content (Asan et al. 2011; Sulonen et al.011; Bodi et al. 2013).
2. The only platform that uses RNA probes is Agilent's Sure Select Human All Exon Kit; all other platforms utilize DNA probes. The target sequences are next to one another rather than overlapping, and the baits utilised are longer (114-126 bp) than those employed in NimbleGen's platform (Clark et al. 2011). Longer baits can tolerate larger mismatches, which makes this design effective at identifying insertions and deletions (indels) (Clark et al., 2011; Bodi et al., 2013; Chilamakuri et al., 2014). It has been suggested that this may also lessen reference allele bias at heterozygous sites when compared to other bait designs, but in actual use, the allele bias has been comparable to other

platforms (Asan et al., 2011; Hedges et al). According to research, the platform generates less high-quality reads but also fewer duplicate reads than NimbleGen (Sulonen et al. 2011). Agilent was determined by Bodi et al. (2013) to have less uniform coverage but a higher alignment rate and fewer PCR duplicates than NimbleGen.

3. The TruSeq Exome Enrichment Kit from Illumina uses 95-bp probes that create tiny gaps in the target region. To fill in these gaps, paired end reads are sequenced outside of the bait sequence. According to Clark et al. (2011), this architecture has a significant rate of off-target enrichment, which lowers its target efficiency when compared to the other platforms. Compared to the other platforms, this kit finds more single-nucleotide variations in the untranslated regions (UTRs) (Clark et al., 2011). However, performance comparisons with NimbleGen and Agilent's "+UTR" kits have not yet been completed. Chilamakuri et al. (2014) discovered that after removing duplicates, multiple mappers, incorrect pairs, and off-target reads, this platform retained 54.8% fewer reads than either NimbleGen (66%) or Agilent (71.7%). At elevated read countsThe TruSeq Exome Enrichment Kit from Illumina uses 95-bp probes that create tiny gaps in the target region. To fill in these gaps, paired end reads are sequenced outside of the bait sequence. According to Clark et al. (2011), this architecture has a significant rate of off-target enrichment, which lowers its target efficiency when compared to the other platforms. Compared to the other platforms, this kit finds more single-nucleotide variations in the untranslated regions (UTRs) (Clark et al., 2011). However, performance comparisons with NimbleGen and Agilent's "+UTR" kits have not yet been completed. Chilamakuri et al. (2014) discovered that after removing duplicates, multiple mappers, incorrect pairs, and off-target reads, this platform retained 54.8% fewer reads than either NimbleGen (66%) or Agilent (71.7%). At elevated read counts.

4. The probe designs of Illumina's Nextera Rapid Capture Exome and Expanded Exome kits are comparable to those of the TruSeq kit. They are different from the other kits in that they use transposomes instead of ultrasonication to fragment the genomic DNA. The only comparative research that has included Nextera in relation to these kits is the one conducted by Chilamakuri et al. (2014). There hasn't been a thorough comparison with the other platforms. The Expanded Exome version

of Nextera, which has the bigger target region of the two kits, was the only kit available at the time of the study. The TruSeq kit and the Expanded Exome kit share a target region that contains miRNAs and UTRs. According to Chilamakuri et al. (2014), the Nextera kit had more high GC coveragethe overall consistency of content areas due to modified bias in the transposome technology utilised during fragmentation; however, recent protocol modifications in the current versions may have rectified this. Additionally, they discovered that, at 40.1%, the Nextera platform kept the fewest reads of all the platforms evaluated after removing duplicates, multiple mappers, incorrect pairs, and off-target reads.

Impact on Crop Improvement

Exome sequencing's application in agriculture, particularly in crop improvement, is emerging as a transformative tool. The impact of this technology in the agricultural sector includes:

1. Accelerating Gene Identification: Exome sequencing facilitates the rapid identification of genes associated with desirable traits such as disease resistance, drought tolerance, and yield, significantly speeding up the breeding process [Devi Singh et al. 2012]
2. Enhancing Breeding Programs: The technology enables the development of more precise and efficient breeding programs, integrating genetic insights directly into practical applications [Devi Singh et al. 2012].
3. Introduction of Novel Traits: Through genetic engineering, exome sequencing assists in the introduction of novel traits into crops, potentially leading to enhanced agricultural outputs. [Devi Singh et al. 2012].
4. Understanding Plant-Host Pathogen Interactions: Exome sequencing plays a crucial role in identifying the gene pool involved in symbiotic and other co-existential systems. This enhances the understanding of plant-host pathogen interactions and assesses the process of natural evolution in crops [Hashmi et al.2015].
5. Assessing Natural Evolution in Plants: The technology has emerged as a significant tool for studying host-pathogen interactions and improving crop production, providing deeper insights into the genetic mechanisms underlying these interactions

Limitations

Although WES is not without limits, it has significant diagnostic promise and can identify the causative mutations in rare

monogenic disorders. Deep intronic variations are missed due to its narrow target capture, which only covers 1%–2% of the genome and ignores clinically significant alleles that exist outside of these regions [Dhir et al. 2010, King et al. 2002]. Cost is still a major concern, although it might be outweighed by the needless expense of "reflex testing." The indifference to epigenetic alterations, the subjectivity of secondary filtering during data processing, read depth and alignment issues, minor CNVs and cryptic indels (poorly resolved and aligned), and variability in in silico sequence capture by different platforms are some other drawbacks. Various laboratories may employ different thresholds for the inclusion or exclusion of certain variants and will have their own methods for prioritizing variants.[Yeo et al. 2004] In a broader sense, WES is a game-changing technique that puts the conventional wisdom of clinical genetics to the test. If suitable retraining programmes are not put in place, the inevitable shift from conventional procedures to WES may put the careers of technologists, cytogeneticists, and other specialists trained prior to the advent of NGS in jeopardy. However, teaching medical professionals about genomics will help to alleviate some of the current deficiencies in communication between the clinical and research domains; for example, the analysis of exome data necessitates the involvement of genomic informaticians with little clinical expertise, and a large number of clinicians lack familiarity with this quickly developing field of technology and thus need ongoing education. Clinicians clearly need to be trained in genetic informatics in order to bridge the gap between the two distinct fields and actually demonstrate personalized, translational medicine[Eleanor et al.2016].

Conclusion

Through the exploration of whole exome sequencing (WES), its refined focus on the exome reveals a powerful tool in the realms of genetics, medicine, and agriculture. We've witnessed how WES strikes a balance between depth, efficiency, and cost, providing vital insights into genetic variations tied to a range of diseases and traits. As we've seen, this technology holds the promise of advancing personalized medicine by enabling precise diagnoses and tailoring treatment plans, while in agriculture, it accelerates breeding programs and enhances crop and livestock quality through genetic insights. This underscores the crucial role of WES in contemporary research and diagnostics, highlighting its significance in driving forward our understanding and manipulation of genetic information for health and agricultural advancements. Moreover, the ongoing advancements in exome sequencing technology and its applications underscore a future rich with potential for groundbreaking

discoveries in both human health and agricultural sciences. The challenges and limitations inherent to WES, ranging from data management to ethical considerations, prompt a continuous evolution of practices and technologies to maximize its benefits. As industry leaders innovate and refine sequencing tools, and as researchers apply these advancements to unravel the complexities of the genome, the implications for future research, diagnostics, and treatmentareboundless. Embracing both the power and the challenges of WES ensures that its journey from a promising technique to a cornerstone of genetic analysis continues to transform our approach to understanding and improving life.

References

1. Albert TJ, Molla MN, Muzny DM, Nazareth L and Wheeler D. 2007. Direct selection of human genomic loci by microarray hybridization. *Nature Methods* 4: 903–905
2. Allen AM, Barker GL, Wilkinson P, Burridge A and Winfield M. 2013. Discovery and development of exome-based, co-dominant single nucleotide polymorphism markers in hexaploid wheat (Triticum aestivum L.). *Plant Biotechnology Journal* 11: 279–295
3. Asan Y, Xu H, Jiang C, Tyler-Smith Y and Xue. 2011. Comprehensive comparison of three commercial human whole-exome capture platforms. *Genome Biology* 12: R95
4. Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond M J and Nickerson DA. 2011. Exome sequencing as a tool for Mendelian disease gene discovery. *Nature Reviews Genetics* 12: 745–755
5. Barbazuk WB, Bedell JA, and Rabinowicz PD. 2005. Reduced representation sequencing: a success in maize and a promise for other plant genomes. *Bioessays*27:839–848
6. Bodi K, Perera AG, Adams PS, Bintzler D and Dewar K. 2013. Comparison of commercially available target enrichment methods for next-generation sequencing. *Journal of Biomolecule Techniques* 24: 73–86
7. Bousquet M, Noirot C, Accadbled F, De Gauzy JS, Castex MP, Brousset P and Gomez-Brouchet A. 2016. Whole-exome sequencing in osteosarcoma reveals important heterogeneity of genetic alterations. *Annals of Oncology* 27(4): 738-744
8. Bustos-Korts D, Dawson IK, Russell J, Tondelli A and Guerra D. 2019. Exome sequences and multi-environment field trials elucidate the genetic basis of adaptation in barley. *Plant Journal* 99(6): 1172-1191

9. Carroll RK, Shelburne SA, Olsen RJ, Suber B, Sahasrabhojane P, Kumaraswami M, Beres SB, Shea PR, Flores AR and Musser JM. 2011. Naturally occurring single amino acid replacements in a regulatory protein alter streptococcal gene expression and virulence in mice. *Journal of Clinical Investigation* 121: 1956–1968

10. Chilamakuri CSR, Lorenz S, Madoui MA, Vodák D and Sun J. 2014. Performance comparison of four exome capture systems for deep sequencing. *BMC Genomics* 15: 449-458

11. Clark MJ, Chen R, Lam HY, Karczewski KJ and Euskirchen G. 2011. Performance comparison of exome DNA sequencing technologies. *Nature Biotechnology Journal* 29: 908–914

12. Dahl F, Gullberg M, Stenberg J, Landegren U and Nilsson M. 2005. Multiplex amplification enabled by selective circularization of large sets of genomic DNA fragments. *Nucleic Acids Research* 33-40

13. Deschamps S, Llaca V and May GD. 2012. Genotyping-by-sequencing in plants. *Biology* 1(3): 460-483

14. Dhir A and Buratti E. 2010. Alternative splicing: role of pseudoexons in human disease and potential therapeutic strategies. *The FEBS journal* 277(4): 841-855

15. Doebley JF, Gaut BS and Smith BD. 2006. The molecular genetics of crop domestication. *Cell* 127(7): 1309-1321

16. Edwards A, Voss H, Rice P, Civitello A, Stegemann J, Schwager C, Zimmermann J, Erfle H, Caskey CT and Ansorge W. 1990. Automated DNA sequencing of the human HPRT locus. *Genomics* 6: 593–608.

17. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter, A., Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J and Turner S. 2009. Real-time DNA sequencing from single polymerase molecules. *Science* 323: 133–138

18. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J and Turner SW. 2010. Direct

detection of DNA methylation during single- molecule, real-time sequencing. *Nature Methods* 7: 461–465

19. Fu Y, Springer NM, Gerhardt DJ, Ying K, Yeh CT and Wu W. 2010. Repeat subtraction-mediated sequence capture from a complex genome. *The Plant Journal* 62, 898–909

20. Glotov OS, Chernov AN and Glotov AS. 2023. Human exome sequencing and prospects for predictive medicine: analysis of international data and own experience. *Journal of Personalized Medicine* 13(8): 1236-1240

21. Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM and Brockman W. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology Journal* 27: 182–189

22. Go G and Choi M 2012. Application of whole exome sequencing to identify disease-causing variants in inherited human diseases. *Genomics & Informatics* 10(4): 214-228

23. Grover CE, Salmon A, and Wendel JF. 2012. Targeted sequence capture as a powerful tool for evolutionary analysis. *American Journal Botanical* 99: 312–319.

24. Hashmi U, Shafqat S, Khan F, Majid M, Hussain H, Kazi AG, John R and Ahmad P. 2015. Plant exomics: concepts, applications and methodologies in crop improvement. *Plant Signaling and Behavior* 10(1): 976-986

25. Hedges DJ, Guettouche T, Yang S, Bademci G and Diaz A. 2011. Comparison of three targeted enrichment strategies on the SOLiD sequencing platform. PLoS One 6: 8595-8609

26. Hodges E, Xuan Z, Balija V, Kramer M, Molla M N and Smith SW. 2007. Genome-wide in situ exon capture for selective resequencing. *Nature Genetics* 39: 1522–1527

27. Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, Smith SW and McCombie WR. 2007. Genome-wide in situ exon capture for selective resequencing. *Nature Genetics* 39: 1522–1527

28. Kaur P and Gaikwad K. 2017. From genomes to GENE-omes: exome sequencing concept and applications in crop improvement. *Frontiers in Plant Science* 8: 304-310

29. Kaur P and Gaikwad K. 2017. From genomes to GENE-omes: exome sequencing concept and applications in

crop improvement. *Frontiers in Plant Science8:* 304-307

30. Kiezun A, Garimella K, Do R, Stitziel NO, Neale BM and McLaren PJ. 2012. Exome sequencing and the genetic basis of complex traits. *Nature Genetics* 44: 623–630

31. King K, Flinter FA, Nihalani V and Green PM. 2002. Unusual deep intronic mutations in the COL4A5gene cause X linked Alport syndrome. *Human genetics 111*: 548-554

32. Klonowska K, Handschuh L, Swiercz A, Figlerowicz M, and Kozlowski P. 2016. MTTE: an innovative strategy for the evaluation of targeted/exome enrichment efficiency. *Oncotarget*7: 67266–67276

33. Korlach J, Bjornson KP, Chaudhuri BP, Cicero RL, Flusberg BA, Gray JJ, Holden D, Saxena R, Wegener J and Turner SW. 2010. Real-time DNA sequencing from single polymerase molecules. *Methods in Enzymology* 472: 431–455

34. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, Helgesen S, Ho, CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF and Rothberg JM. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437: 376–380

35. Mascher M, Richmon TA, Gerhardt DJ, Himmelbach A and Clissold L. 2013. Barley whole exome capture: a tool for genomic research in the genus Hordeum and beyond. *Plant Journal* 76(3): 494-505

36. Maxam AM and Gilbert W. 1977. A new method for sequencing DNA. *Proceeding of the National Academy Science of United states of America* 74: 560–564

37. Mertes F, ElSharawy A, Sauer S, Helvoort Van, JMLM, Vander Zaag, PJ and Franke A. (2011). Targeted enrichment of genomic DNA regions for next- generation sequencing. *Briefing in Functional Genomics* 10: 374–386

38. Mitsui J, Fukuda Y, Azuma K, Tozaki H, Ishiura H, Takahashi Y, Goto J and Tsuji S. 2010. Multiplexed resequencing analysis to identify rare variants in pooled DNA with barcode indexing using next-generation sequencer. *Journal of Human Genetics* 55: 448–455

39. Murphy KM, Berg KD and Eshleman JR. 2005. Sequencing of genomic DNA by combined amplification and cycle sequencing reaction. *Clinical Chemistry* 51: 35–39

40. Neves LG, Davis JM, Barbazuk WB, and Kirst M. 2013. Whole-exome targeted sequencing of the uncharacterized pine genome. *The Plant Journal* 75: 146–156

41. Ng SB, Turner EH, Robertson PD, Flygare SD and Bigham AW. 2009. Targeted capture and massively parallel sequencing of twelve human exomes. *Nature* 461: 272–276

42. Okou DT, Steinberg KM, Middle C, Cutler DJ, Albert TJ and ZwickME. 2007. Microarray-based genomic selection for high-throughput resequencing. *Nature Methods* 4: 907–909.

43. Pankin A, Campoli C, Dong X, Kilian B and Sharma R. 2014. Mapping-by-sequencing identifies HvPHYTOCHROME C as a candidate gene for the early maturity 5 locus modulating the circadian clock and photoperiodic flowering in barley. *Genetics* 198: 383–396

44. Pareek CS, Smoczynski R andTretyn A. 2011. Sequencing technologies and genome sequencing. *Journal of Applied Genetics* 52: 413–435

45. Sanger F and Coulson AR. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology* 94: 441–448

46. Scheben A, Batley J and Edwards D. 2017. Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. *Plant Biotechnology Journal* 15: 149–161

47. Seaby EG, Pengelly RJ and Ennis S. 2016. Exome sequencing explained: a practical guide to its clinical application. *Briefings in Functional Genomics* 15(5): 374–384

48. SenGupta DJ and Cookson BT. 2010. SeqSharp: a general approach for improving cycle- sequencing that facilitates a robust one-step combined amplification and sequencing method. *The Journal of Molecular Diagnostic* 12: 272–277

49. Singh D, Singh, Chaudhary PK, Mehla K and Kumar S. 2012. Exome sequencing and advances in crop improvement. *Advances in Genetics* 79: 87-121

50. Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SB and Hood LE. 1986. Fluorescence detection in automated DNA sequence analysis. *Nature* 321: 674–679

51. Song J, Yang X, Resende MFR, Neves LG, Todd J and Zhang J. 2016. Natural allelic variations in highly polyploidy saccharum complex. *Frontiers Plant Science* 7:804

52. Sulonen AM, Ellonen P, Almusa H, Lepisto M and Eldfors S. 2011. Comparison of solution-based exome capture methods for next generation sequencing. *Genome Biol*ogy 12: 94-103

53. Tuskan GA, DiFazio S, Jansson S, Bohlmann J and Grigoriev I. 2006. The genome of Black Cottonwood, Populus trichocarpa (Torr. & Gray). Science, 313, 1596–1604

54. Van Dijk EL, Jaszczyszyn Y and Thermes C. 2014. Library preparation methods for next-generation sequencing: tone down the bias. *Experimental Cell Research* 322(1): 12–20

55. Varshney RK and Singh VK. 2017. Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends in Biotechnology* 35(6): 522-532

56. Warr A, Robert C, Hume D, Archibald A, Deeb N and Watson M. 2015. Exome sequencing: current and future perspectives. *G3: Genes, Genomes, Genetics* 5(8): 1543-1550

57. Washida H, Kaneko S, Crofts N, Sugino A, Wang C and Okita TW. 2009. Identification of cis-localization elements that target glutelin RNAs to a specific subdomain of the cortical endoplasmic reticulum in rice endosperm cells. *Plant and Cell Physiology* 50(9): 1710-1714

58. Wendler N, Mascher M, Noh C, Himmelbach A and Scholz U. 2014. Unlocking the secondary gene-pool of barley with next-generation sequencing. *Plant BiotechnologyJournal* 12: 1122–1131

59. Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT, Gomes X, Tartaro K, Niazi F, Turcotte CL, Irzyk GP, Lupski JR, Chinault C, Song XZ, Liu Y, Yuan Y, Nazareth L,

Qin X, Muzny DM, Margulies M, Weinstock GM, Gibbs RA and Rothberg JM. 2008. The complete genome of an individual by massively parallel DNA sequencing. *Nature* 452: 872–876

60. Winfield MO, Allen AM, Burridge AJ, Barker GLA, Benbow HR and Wilkinson PA. 2016. High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant Biotechnology Journal* 14: 1195–1206.

61. Witek K, Jupe F, Witek AI, Baker D and Clark MD. 2016. Accelerated cloning of a potato late blight-resistance gene using RenSeq and SMRT sequencing. *Nature Biotechnology Journal* 34(6): 656-660

62. Xin M, Wang Y, Yao Y, Song N, Hu Z, Qin D, Xie C, Peng H, Ni Z and Sun Q. 2011. Identification and characterization of wheat long non-protein coding RNAs responsive to powdery mildew infection and heat stress by using microarray analysis and SBS sequencing. *BMC Plant Biology* 11: 1-13

63. Xiong H, Guo H, Fu M, Xie Y, Zhao L, Gu J, Zhao S, Ding Y, Du Q, Zhang J and Qiu L. 2023. A large-scale whole-exome sequencing mutant resource for functional genomics in wheat. *Plant Biotechnology Journal* 21(10): 2047-2056

64. Zho L and Holliday JA. 2012. Targeted enrichment of the black cottonwood (*Populus trichocarpa*) gene space using sequence capture. *BMC Genomics* 13: 703-719.