ISSN NO. 2320-5407

Journal homepage: http://www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH

RESEARCH ARTICLE

Next Generation Sequencing Technologies: Applications in Crop Improvement

Youvika Singh[1], Virupaksh U Patil[1]*, Ritu Singh[1] , Pradeep Kumar Shukla[2] and Shashi Rawat[1]

1.Central Potato Research Institute, Shimla-171001, Himachal Pradesh, India

2. Department of Biological Sciences, School of Basic Sciences, Sam Higginbottom Institute of Agriculture, Technology and Sciences, Naini, Allahabad-211007

Manuscript Info

Manuscript History:

Received: 15 July 2015
Final Accepted: 22 August 2015
Published Online: September 2015

Key words:

sequencing by synthesis, throughput, read length, crop improvement and genomic breeding

Corresponding Author

Mandsorwale Deepti

Abstract

Whole genome sequencing of plants with large genome sizes was considered to be a major challenge about a decade ago. Thanks to the rapid developments in sequencing technologies, genome sequencing of any crop plant has now become faster, cheaper and technically less demanding. These sequencing technologies are undergoing never ending revolution with commercialization of next-next-generation technologies capable of sequencing thousands of millions of nucleotide bases in each run. Using these sequencing technologies it is possible to sequence or resequence entire plant genomes or sample entire transcriptomes in greater depth than ever before. Rather than sequencing individual genomes, now scientists envision the sequencing of hundreds of related genomes to sample genetic diversity within and between germplasm pools. The increasing availability of DNA sequence information in large number of crop plants enable the discovery of genes and molecular markers associated with diverse agronomic traits creating new opportunities for crop improvement. Such huge and accurate DNA sequence information impacts many of the current uses of molecular tools in plant evolution, phylogenetics, fingerprinting, linkage mapping and marker assisted selection and breeding. This review provides an overview of various NGS technologies that are currently available and near future arrivals along with their applications in *de nova* sequencing and resequencing, marker development, population genetics and evolutionary biology, organelle sequencing and translational biology of crop plants.

INTRODUCTION

The development of DNA sequencing strategies has been a high priority in genomics research since the unearthing of the structure of DNA and the basic molecular mechanisms of heredity. However, it was not until the works by Maxam and Gilbert (Maxam et al., 1977) and Sanger (Sanger et al., 1977) that the first practical sequencing methods were developed and implemented on a large scale. These two landmark researches were responsible for the introduction of first automated DNA sequencers led by Caltech (Smith et al., 1986) and subsequently commercialized by Applied Biosystems (ABI), European Molecular Biology Laboratory (EMBL), Pharmacia-Amersham and General Electric (GE) healthcare. These are now categorized as first generation sequencing technologies. Despite their popularity as a 'gold standard' among the research community, these suffer certain limitations like limited length of DNA sequenced, biological biasness, higher amount of sample required, lower number of samples that can be analyzed and most important is higher cost of sequencing. With the advances made in the field of micro-fluidics, imaging power, detection power and computational tools, unconventional sequencing technologies with increased throughput and lower sequencing cost are continuously emerging. The completion of the first human genome drafts (Yamey, 2000) was just a start of the modern DNA sequencing era which resulted in

further invention, improved development toward new advanced strategies of high-throughput DNA sequencing, which were collectively called as the "high-throughput next generation sequencing" (HT-NGS) (Varshney et al., 2009). The first of these NGS technologies was pyrosequencing, which was followed by Illumina and SOLiD (Sequencing by Oligo Ligation and Detection). These HT-NGS technologies have the capacity of producing 100 times more data as compared to the first generation sequencers. The horizons and expectations have broadened due to the technological advances in the field of genomics, especially the HT-NGS and its wide range of applications such as: chromatin immune-precipitation coupled to DNA microarray (ChIP-chip) or sequencing (ChIP-seq), RNA sequencing (RNA-seq), whole genome genotyping, *de novo* assembling and re-assembling of genome, genome wide structural variation, mutation detection and carrier screening, DNA library preparation, paired ends and genomic captures, sequencing of mitochondrial and chloroplast genome. Besides the advancement of sequencing techniques, the past decade is remembered as the decade of the genome research. Since the publications of first composite genomes of human (Lander et al., 2001; Venter et al., 2001) many draft genomes from many plant and animal species have been published (www.ensembl.org/info/about/species.html) (Fig 1). Recently, a single molecule based sequencing technologies have hit the market, which is collectively called as Next-Next Generation Sequencing (NNGS) or Third Generation Sequencing (TGS) technologies. Pacific Biosciences Inc. was the first to introduce the NNGS in the global market. The NNGS technologies are more efficient as compared to NGS in terms of throughput, cost of sequencing and time consumption, but their utility and performance is yet to be proved as a real advancement of technologies over NGS.

These never ending advances in sequencing technologies provides opportunities to target not only the model plant species with small genome sizes, but many cultivated and economically important plant species for sequencing to identify millions of novel markers, agronomically important genes, knowledge of which can directly translated into crop improvement. Reference genome sequences for several crop species are now becoming available and this information allows both the rapid identification of candidate genes through bioinformatics analysis and single nucleotide polymorphism (SNP) discovery through comparison of the reference with sequence data from different cultivars. With the advance in the sequencing technologies, reduction in the sequencing cost to US$ 1000 genome sequencing of crop species having larger genomes like wheat, barley and cotton is also made possible (Chandra et al., 2011). The paper reviews the global status regarding NGS and NNGS technologies and their utilization in the crop improvement.

Next Generation sequencing technologies

The completion of first human genome draft (Yamey, 2000) started the modern DNA sequencing era which resulted in further invention, improved development toward new strategies of high throughput DNA sequencing which were collectively called Next Generation Sequencing (NGS) or Second Generation sequencing. There are mainly three leading NGS technologies which dominated the global market for the last 5-6 years are discussed here.

Pyrosequencing

The pyrosequencing technology (http://www.454.com) was the first NGS, derived from technical combination of pyrosequencing chemistry and emulsion PCR. The basis of this technology was sequencing by synthesis (Melamede, 1985), a different approach of DNA sequencing by pyrophosphate detection was also reported (Hyman, 1988). A team lead by Nyren in 1993 came out with a sequencing approach based on chemi-luminiescent detection of pyrophosphate released during deoxynucleotide triphosphate (dNTP) incorporation (Nyren et al., 1993). Later, up gradation of technique by Ronaghi and coworkers laid the foundation stone for the commercial development of pyrosequencing at Royal Institute of Technology, Stockholm in 1996 (Ronaghi et al., 1996). On a separate front, single-molecule PCR in micro compartments comprising water-in-oil emulsion was described by Twafik and Griffith (Twafik et al., 1998). The 454 Life sciences founded by Jonathan Rothberg in 2000, launched first commercially available NGS platform named GS20 in 2005. In the same year Margulies and colleagues sequenced whole genome of *Mycoplasma genitalia* at 96% coverage and 99.96% accuracy in a single run using GS20 (Margulies et al., 2005). The technology is incessantly being upgraded several times into a routine functioning method. The first major technological improvement was replacement of dATP with that of dATPα (Ronaghi et al., 1996) followed by introduction of light phase pyrosequencing (Ronaghi et al., 1998) and addition of ssDNA-binding proteins to pyrosequencing (Ronaghi et al., 2001). In 2007, Roche introduced newer version as GS FLX with a unique flowcell referred 'picotiter plate' (PTP) comprising 3.4 $\times 10^6$ separate sequencing reaction wells allowing hundreds of thousands of sequencing reactions to be carried out in parallel and massive high-throughput way. The current pyrosequencer instrument, the GS FLX+ produces an average read length of approximately 1000 bp and throughput of approx 800Mb to 1Gb of high quality sequence data per 7-8 hr run (www.454.com/). Pyrosequencing

is basically dual step approach. Firstly, single stranded DNA is fractionated into smaller fragments (300-1000bp), polished (made blunt end), and short oligo adapters having 5' biotin tag are ligated to the fragments. These adapters provide priming sequence for the attachment; amplification as well as sequencing the fragment. DNA fragments to be sequenced are then individually immobilized onto streptividin decorated beads which are amplified by the PCR in the water-oil emulsion droplets. These droplets act as individual amplification reactors producing manifold replicas (~10^7) of the same DNA sequence on each bead. Template single stranded DNA is hybridized to sequencing primer and loaded onto the PTP plate along with DNA polymerase, ATP sulfurylase (a recombinant version from *Saccharomyces cerevisiae* (Karamohamed et al., 1999), Luciferase (from firefly *Photinus pyralis*) (Ronaghi et al., 1998) nucleotide degrading enzyme Apyrase (from potato) (Ronaghi et al., 2001) along with the substrates adenosine 5' phosphosulfate (APS) and luciferin. One of the four dNTPs are added and if complementary DNA polymerase incorporates onto the template accompanied with the release of pyrophosphate (PPi) equal to molarity of incorporated nucleotide. The PPi released is quantitatively converted into adenosine tri phosphate (ATP) in presence of APS. The ATP acts as a fuel to lucifease mediated conversion of luciferin to oxyluciferin that generates light in comparative amount of ATP produced. Unincorporated nucleotides and ATPs are continuously washed away by apyrase and the next reaction start with another nucleotide addition cycle. One picomole of DNA in a pyrosequencing reaction yields $6x10^9$ photons at a wavelength of 560 nanometers, which is easily being detected by 16 mega pixel CCD camera maintained at -24 ^{0}C for its higher resolution and performance. The sequence of DNA is yielded in "pyrogram" corresponds to order of nucleotides that has been incorporated.

Pyrosequencing technology works best in cases where longer read length are in demand like *de novo* sequencing assembly and metagenomics. The run time required is also very short (few hours). The present strategy can generate more than 1,000,000 individual reads with improved read length of 500-800 bases per 10 hour run. Additionally, the process requires short sample preparation time approximately two hours. A major limitation of pyrosequencing is difficulty in sequencing homopolymer or repetitive regions. Homopolymer regions cause dephasing due to asynchronous synthesis at the repetitive region. This will increase the sequence length and also effects error rate. Another disadvantage of pyrosequencing is its high cost ($45.0 per Mb data sequenced) as compared to other NGS technologies.

Illumina genome analyzer

In 1997, British chemist Shankar Balasubramanian and David Klenerman conceptualized an approach for sequencing single DNA molecules attached to microspheres. They funded Solexa in 1998; however their goal of sequencing single DNA molecules was not fulfilled. The idea was then shifted towards sequencing clonally amplified templates. Year 2006, marks the commercial launch of the first 'short read' sequencing platform *Solexa Genome Analyzer*. Illumina is one of the high throughput sequencing technology among the NGS and it uses reversible terminator-based sequencing by synthesis approach. The templet DNA sample is fractionated to the average size ~800bp. The fragmented DNA ends are repaired; 5' end phosphorylated while at 3' poly A tail is added. Repairing of DNA is carried out using T_4 DNA Polymerase (digests 3' protruding ends), Klenow DNA polymerase (extension of 3'recessive ends) and T_4 PNK (phosphorylates 5'end and dephosphorylates 3'ends). Like pyrosequencing, illumina also requires template sequence to be converted to special sequencing library which insures the immobilization and amplification for sequencing (Fedurco et al., 2006). Therefore, two unique forked adaptors (adaptor oligonucleotides are complimentary to flow cell anchors) are added at the 5' and 3' end of the DNA fragment. The prepared samples are immobilized on 8 channeled flow cell surface allowing bridge amplification. Hybridization of library fragments and adaptor with that of flow cell occurs by active heating and cooling step. Subsequently, reactants and an isothermal polymerase are incubated to amplify the fragment in a discrete area 'cluster' on flow cell surface (for animation: http://www.illumina.com/) to form small clusters of single stranded fragments called 'bridge amplification'. Clusters are formed impulsively due to the fact that the newly produced copies of the fragment get attached in close proximity to the original fragment. After the bridge amplification is done, densely packed clusters of fragments formed, each cluster consisting of many copies of the same fragment, which begins the sequencing by synthesis step. For single strand sequencing of forward strands, clusters are denatured, chemically cleaved and washed. Sequencing of forward strand starts with the hybridization of sequencing primer complimentary to adaptor sequence followed by addition of DNA polymerase and mixture of four differently colored fluorescent dye terminator nucleotides. All four nucleotides are modified with distinct fluorochrome and reversible terminator group attached at its 3'hydroxyl group is chemically blocked, so that when one nucleotide is incorporated replication stops. This ensures the uniqueness of each event. DNA polymerase incorporates the appropriate nucleotide and unused nucleotides are washed away. After every incorporation cycle imaging step occurs for determining each incorporated nucleotide followed by chemical cleavage step which

removes fluorescent nucleotide and unblocks the 3' end with the help of reducing agent tris (2-carboxymethyl phosphine) for next sequencing cycle. The process of adding nucleotides, imaging and removing the terminator is called a cycle. The illumina sequencing run offers very high throughput with 50×10^6 clusters per flow cell to generate of 2 to 15 GB in 2-8 days of run time and read length of 35 to 75 bases. The paired end module enables sequencing up to 2x100bp of fragments ranging from 200bp to 5kb with output reaching upto 45-50 GB. The latest technology 'Hi-seq 2500' produces around 600 Gb throughput per 11 day run with dual flow cell and another higher version of the same MiSeq® system with much more higher throughput and quality is at the doorstep for the release (http://www.illumina.com/). The run cost offered by this technology is also very low ($5.9 per Mb data sequenced). Currently, illumina is most widely used NGS platforms with 60% of global NGS installations. Though, it has very short read length with relatively higher error rates, but are compensated by coverage and throughput. The massively produced sequence data is difficult to manage and process as huge number of short reads complicate the assembly and alignment algorithm.

Sequencing by Oligo Ligation and Detection (SOLiD)

The SOLiD strategy is the only NGS technology which is based on ligase mediated sequencing chemistry and di-based labeled probes, hence the name sequencing by oligo ligation and detection. The technology was developed at Harvard Medical School and was first time used successfully for the resequncing of *Escherichia coli* genome (Shendure et al., 2005). The technology was further refined and commercially launched by Applied Biosystems in the year 2007. Sample preparation share similarities to that of pyrosequencing and Illumina. Oligo adapter is ligated to the DNA sequence to be determined, and then bound with the 1 µm magnetic beads decorated with complementary oilgos and DNA fragment is clonally amplified by emulsion PCR. After amplification the enriched beads are recovered for their immobilization onto derivitized-glass flow-cell surface by hybridizing adaptor sequence (P1 adaptor) with complimentary sequence on the flow cell. The ligation based sequencing initiate with annealing of universal sequencing primer complementary to the adapter sequence flanking the library fragment. Mixture of fluorescently labeled interrogated octamer probes hybridize to the fragment DNA. These octamers are dibase degenerate fluorescently labeled oligo nucleotides with specific dye at its 5'ends. In these octamers, first 2 nucleotides are interrogated dinucleotide, 3rd, 4th and 5th bases are degenerate bases having cleavable site and 6th, 7th and 8th are universal bases with flouorophore attached to 5'end. The system uses sixteen possible dinucleotide combinations encoding four different fluorescent dyes. Ligase enzyme hybridizes complementary octamer oligonucleotide sequence to the target DNA; imaging signal identifies the attached oligonucleotide by the associated fluorescent dye. A chemical cleavage step cleaves off the 6th through 8th universal base thereby removing attached fluorescent group and enabling the next sequencing cycle to proceed. The cycle is repeated 7 times referred to as 'round' yielding 35bp read length. However, to increase read length additional cycles can also be performed. Second round of sequencing onsets with the stripping off synthesized strand and hybridizing another universal sequencing primer at the n-1 position followed by subsequent ligation and cleavage step. These steps are repeated with n-2, n-3, n-4 primers. Two flow cells in one run produces an overall output of 4 Gb data with two time enquiry of each base "two base encoding" which provides an extra quality check to discriminate measurement errors. High fidelity ligation chemistry coupled with interrogation of each nucleotides base twice yields sequence output with 99.9% accuracy. On an independent platform, the church laboratory at Harvard Medical School, Danaher Motion and Dover systems have collaborated to develop and introduce least expensive sequencing by ligation platform, the Polonator G.007 (http://www.polonator.org). The technology enables analyzing number of samples simultaneously at a smaller amount of reagents and low cost. The Polonator G.007 expected to produces 10-35 Gb of data per 2.5 day run. The only disadvantage with SOLiD technology is short read length which increases cost of data analysis and processing and significantly limits its applications in spite of low error rates than other NGS platforms. Special features of the three major NGS technologies mentioned above have been shown in table 1.

Upcoming developments

Although the NGS platforms have radically abridged the cost of sequencing 10 to 20 folds, these dependent on PCR amplification of DNA fragments to make the signal enough for reliable detection. Although PCR amplification has revolutionized the DNA analysis, some instances it may introduce base sequence errors or favor certain sequences over others. Therefore, scientists tried to bypass this error prone step by interrogating single molecule of DNA. With the dawn of new sequencing technologies called 'next-next generation sequencers' or 'third generation sequencers' genomic research is now on its doorway to further revolutionize (Schadt et al., 2010). These technologies are based upon single molecule sequencing (SMS). The companies introducing these SMS claim that the technologies will be much faster and cheaper than their ancestral generations, which enable the scientists to peruse novel scientific enquiries that are difficult to achieve with the present technologies. Whether or not SMS technologies will succeed

in achieving their promise is debatable. The concept of sequencing by synthesis without prior amplification is currently pursued by number of companies; productive efforts are underway to develop new SMS technologies, the underlying principles of these are discussed below.

True single molecule sequencing (tSMS^{TM})

The first commercially available single molecule sequencing system, introduced by Braslavsky et al., 2003 and commercialized by Helicos Biosciences, USA (http:// www.helicosbio.com/), has been employed to resequence the entire genome of M13 virus (Harris et al., 2008). This is highly sensitive detection system used to interrogate single molecule of the DNA (or RNA without reverse transcription) in parallel without the need of amplification technique. Libraries are prepared by random fragmentation and poly dA tailing of the template allows it to hybridize with surface decorated poly dT oligonucleotide, anchored to glass cover slip. In every cycle called 'quad' DNA polymerase and single species of nucleotide labeled with the cyanine dye, cy5 (a non radioactive fluorescent dye) is sequentially added (ACGT, ACGT….), this causes extension of surface immobilized primer template duplexes. The cleavage of the fluorescent label enables the determination of particular nucleotide (Gupta, 2008). The fluorescent 'virtual terminator' nucleotide (Cy5-12ss-NTPs) prevents the incorporation of any subsequent nucleotide until the nucleotide dye moiety is cleaved (Krzymanska-Olejnik et al., 2009). The images from each quad are assembled to generate an overall set of sequence reads. On a standard run, 120 cycles of nucleotide addition and detection cab be carried out. Presence of 25 channel flow cells in a standard run enables 50 different samples sequencing simultaneously possible.

There are several unique features which makes tSMS an advanced sequencing strategy. Firstly, misincorporation arising from dephasing is not an issue. In tSMS, dephasing is bypassed as each template is monitored individually (Gupta, 2008). Homopolymer is the second most error prone stage as several molecules need to be incorporated in the same cycle. In contrast, tSMS uses single molecule for interrogation therefore, the problem can be controlled by limiting the arêtes of incorporation events. Thirdly, error rate is reduced substantially by performing 'two pass' sequencing. The template strand is sequenced as usual (pass 1); the newly synthesized strand is surface tethered, and original template is denatured. Sequencing primed from distal adaptor yields second sequence for same template with opposite orientation (pass 2) (Shendure et al., 2008). Even the sub-nanogram amounts and very poor quality DNA, including degraded or modified DNA, is sufficient for sequencing (Hart et al., 2010; Thompson et al., 2010). This system has been used for variety of applications involving sequencing RNA sample providing unparalleled quantitative accuracy for RNA expression measurements (Ozsolak et al., 2009). Precise expression measurement to be made with either RNA or cDNA has been possible due to very high read count per sample (Ozsolak et al., 2010a, b) and also all classes of RNA molecules can be detected using this technology which was not possible previously (Kapranov et al., 2010 a,b). The technology has been utilized in sequencing ancient DNA (Orlando et al., 2011), detection of BRCA1 mutation by single step target selection (Thompson et al., 2011) and human gene therapy (Kapranov et al., 2011). Helicos claims the total output to be 25-35 Gb per run with a throughput of 1.1 Gb of data per hour and an average read length of 35 bp. (http://www.helicosbio.com). The only important weakness of tSMS approach is high raw error rate, which can be overcome by repetitive sequencing but increases cost per base for a given accuracy rate.

Single Molecule Real Time Sequencing (SMRT^{TM})

SMRT is a paralleled single molecule DNA sequencing by synthesis developed by Pacific Biosciences (http://www.pacificbiosciences.com) based on the usage of Zero-Mode Waveguide (ZMW), developed in the laboratory of Harold Craighead at Cornell University. SMRT technology is unique in the field of DNA Sequencing and offers the ultimate combination of speed, long reads, and low costs. The axis of this technology involves specialized chip called SMRT chip, made of 100 nm thick metal film containing thousands of zero-mode waveguides (ZMW) which are basically wells of 10-50 nm diameters. Each well contains, a DNA polymerase molecule attached at the bottom, that elongates the primer hybridized template by incorporating the ϒ-labeled dNTPs. All the four bases are fluorescently labeled at the phosphate group with distinguishable fluorophores (Korlach et al., 2010). When a nucleotide is incorporated during DNA synthesis, the laser beam illumination of small detection volume (20 zeptoliters = 20×10^{-21}) causes attached fluorophore to light up allowing the identification of each nucleotide incorporated. During formation of the phosphodiester bond, a nucleotide is held up in the detection volume for a much longer time (milliseconds) than the time (microseconds) needed for a nucleotide to diffuse in and out of the detection volume. This increase in time facilitates proper detection (Gupta, 2008). The other unincorporated nucleotides float in the dark unilluminated volume of ZMW and do not light up. This strategy

incorporates nucleotide at a speed of ten bases per second simultaneously in all thousands of ZMWs located on SMRT chips giving rise to a chain of thousands of nucleotides in length within minutes (Korlach et al., 2008).

SMRT sequencing platform require minimal amount of sample and reagent for a complete run and also there are no laborious scanning and washing steps. Furthermore, this does not need any routine PCR amplification as needed in previous generation sequencing systems thereby avoiding any systematic amplification bias. The company claims the average read length of >3,000bp, which simplifies the assembly and mapping. However, premature termination caused by laser induced photo damage to polymerase and nucleotide causes short reads (Thompson et al., 2011). Turning off laser for short period of time during sequencing helps to produce longer reads named strobe reads (Ritz et al., 2010). It takes less than a day for obtaining the result, starting from sample preparation. SMRT sequencing offers flexibility providing multiple protocols including standard, circular consensus and strobe sequencing. An additional benefit of this system is its ability to potentially detect modified bases. It is possible to detect 5-methylcytosine, although the role of sequence context and other factors affecting the accuracy of such assignments remains to be clarified (Flusberg et al., 2010). Despite many potential advantages of SMRT sequencing, a number of challenges remain, like the raw read error rates can be in excess of 5%, with error rates dominated by insertions and deletions, particularly problematic errors when aligning sequences and assembling genomes. In addition, the throughput of SMRT sequencing will not initially match to that of data achieved by previous generations. Nevertheless, high raw error rates can be overcome by creating SMRT bell templates consisting double stranded region (insert of interest, 40-25,000 bp) flanked by single stranded loops on either side (Travers et al., 2010).

Nanopore sequencing

The launch of nanopore machine put an end to long decade wait as the method was under progress since 1995. The idea behind Nanopore sequencing was first conceived by David Deamer at the University of California. The technology is being commercialized by Oxford Nanopore technologies, which claims to sequence human genome within 15 minute. The technology involves the use of thin membrane that contains nanopore of ~1.5 to 2 nm diameter. The target DNA is placed on one side and current is applied across the membrane. The negatively charged DNA translocates through membrane and blocks the channel which generates alteration in electrical conductance leading to change in current in the range of pico amps (pA). This enables the discrimination of DNA molecules with different sequences (Kasianowicz et al., 1996). Nanopores of great interest includes solid state nanopore like carbon nanotubes and thin films (Garaj et al., 2010) plastic materials (Harrell et al., 2006), biological protein nanopore like α-hemolysin (Stoddart et al., 2010) and MspA (Derrington et al., 2010) have been investigated for sequencing. Of all these, protein nanopore proves to be more advantageous as they can be genetically and chemically modified to optimize the detection of specific bases and translocation rate of DNA through the pore (Stoddart et al., 2010). The α-hemolysin is a toxic protein from *Staphylococcus aureus*, which is highly stable and remains functional even at boiling point (Kang et al., 2005). It was also demonstrated that inner diameter of the protein is as wide as single stranded nucleic acid; and helps locally to unknot the coiled nucleic acid enabling it to translate in strictly single file and sequential order (Kasianowicz et al., 1996).

High translocation speeds (potentially millions of bases per second) creates barrier in the signal. In such cases detecting a signal over background noise can be a challenging task. Variety of methods have been used to control the pace of DNA through nanopores, including attachment of polystyrene beads (Balagurusamy et al., 2010) varied salt concentrations (de Zoysa et al., 2009) viscosity (Kawano et al., 2009) magnetic fields (Peng et al., 2009) introduction of regions of double-stranded DNA on a single-stranded target (Garaj et al., 2010) and the attachment of polymerase to retach DNA through the α-hemolysin pore (Stoddart et al., 2010).Efforts to overcome this came out in the form of modification to basic nanopore technology described as *hybridization assisted nanopore sequencing* (HANS) developed by NABsys (http://nabsys.com/) which combines nanopore sequencing with sequencing by hybridization. The approach involves hybridizing molecule to be sequenced with known probe sequence and the resulting hybrid will pass through a nanopore. The changes in the current can be measured and subsequently hybridization sites can be determined. Another modified version of nanopore sequencing is '*design polymer'-assisted nanopore sequencing* developed by LingVitae (http://www.lingvitae.com/) which involves conversion of target DNA into magnified form called 'design polymer' and encoding the sequence by transformed nucleotide sequence, using binary code of molecular beacons.

The most important advantage of nanopore sequencing that makes it a unique and inexpensive technology over other methods is undemanding sample preparation, requiring minimal chemistries, eliminating the need of fluorescent nucleotides, enzymes, cloning and amplification steps. Oxford nanopore sequencing technologies expects to start

selling its new machine in second half of this year in the high throughput electronic platform 'GridION' and MinION (a disposable sequencer of a USB memory stick size) (Hayden, 2012), Albeit of many advantages nanopore sequencing also possess some loopholes. Firstly, mechanical instability of lipid bilayer supporting nanopore requires continuous monitoring. Secondly, the sensitivity of biological nanopores to experimental conditions (such as pH, temperature and salt concentration) and finally, difficulty in integrating biological systems into large-scale arrays.

FRET based approach

Life Technologies, a major provider of both first and second generation sequencing systems, is developing fluorescence resonance energy transfer (FRET) based single molecule sequencing by synthesis approach technology, initially introduced by VisiGen (Hardin, 2008). This type of approach could be considered an improvement over the Helicos technology. A theoretical throughput of 1 Mb per instrument second has been given, but right now it is difficult to measure progress. The basic idea behind this strategy was, monitoring of polymerase dependent incorporation of nucleotide bases into the DNA strand. The specialized DNA polymerase contains donar fluorophore, each nucleotide species carries one of the four differently colored acceptor fluorophor. When a nucleotide is incorporated, the proximity of donar-acceptor fluorophor results in a signal called 'FRET'. The FRET signal is specific for a particular nucleotide incorporated at the particular position. After the nucleotide has been incorporated, the pyrophosphate containing fluorophor is released, thereby quenching signal and preparing for next step (Gupta, 2008). Significant advances have been made with the commercial release of the technology in the form of 'starlight' system expected in near future. The current technology consists of a quantum-dot-labeled polymerase that synthesizes DNA using four distinctly labeled nucleotides in a real-time system (Pennisi, 2010). Quantum dots, which are fluorescent semiconducting nanoparticles, have an advantage over fluorescent dyes as they are much brighter, larger and less susceptible to bleaching, although they are much more susceptible to blinking. Interaction of fluorescently labeled nucleotides with the quantum dots (attached with DNA polymerase) causing an alteration in fluorescence of both nucleotide and quantum dots. The quantum dot signal drops, whereas a signal from the dye-labeled phosphate on each nucleotide rises at a characteristic wavelength. The real-time signal is captured and DNA sequence is determined. As each sequence is bound to the surface, it can be reprimed and sequenced again for improved accuracy (Thompson et al., 2011).

Transmission electron microscopy (TEM) for DNA sequencing

Transmission electron microscopy for DNA sequencing is a new born third generation DNA sequencing system, the idea of which was introduced by ZSGenetics (http://zsgenetics.com/). The detail of the technology was first introduced at a sequencing conference in 2008 (http://www.healthtech.com/). The technology involves, denaturing templet DNA followed by heavy labeling. Natural DNA is transparent when viewed with TEM, as elements structuring it (C, O, N, H and P) have a low atomic number (Z=1-15). To circumvent this problem three of the four bases are labeled with heavy elements (bromine Z-35, iodine Z-53, and trichloromethane Z-63) and fourth remains unlabeled. Templet DNA molecules are linearised on a thin solid substrate by molecular cloning, a technique that utilizes the force of receding air-water interface to extend DNA molecules irreversibly bound to silane layer once dried (Bensimon et al., 1994). Image can be viewed as dark and light bright spots on electron micrograph, corresponding to differentially labeled DNA bases. The technology helps in genetic information applications including enabling widespread adoption of genetics-based medicine and driving important innovation in disease research, drug discovery, forensics, environmental impact studies, agriculture, anthropology and history. The ZSG claims for potential of sequencer to be 10-20 Kb with rate of 1.7 Gb per day, at comparatively lower cost than second generation sequencers and has already released images of 23kb DNA (Gupta, 2008). The TEM approach is a technically challenging task. Selective heavy atom labeling and attaching and straightening the labeled DNA to a substrate are a serious technical challenge. Further, the DNA sample should be stable to the high vacuum of electron microscope and irradiation by a focused beam of high-energy electrons.

Ion torrent

A scalable, low cost DNA sequencing technology has always been the choice of scientists for genome sequencing. The next in the list for above attributes is PostLight^{TM} sequencing technology (Ion torrent) introduced by Life technologies (http://www.iontorrent.com/). It is the first commercial sequencing technology which does not depend upon light for determining the sequence. It produces sequencing data quickly and in an unsophisticated manner. The sequencing chemistry is incredibly simple based on a well characterized biochemical process. When a nucleotide is incorporated into the growing DNA strand by polymerase, H^+ ion is release as byproduct causing a detectable local change in the pH. The whole process occurs inside micro-wells on ion torrent sequencing chip, with each well holding a different DNA template. Change in solution pH of the well is detected by the ion sensor placed beneath

the well, essentially going directly from chemical information to digital information. Change in pH is proportional to number of nucleotides added. If voltage is double there must be two identical bases on the DNA strand, and the chip records two identical bases. If the nucleotide that floods the chip is not a match, no voltage change is recorded and no base is called. Technology does not involve the use of fluorescent nucleotides or imaging, base call is relatively fast and cost effective compared to other sequencing technologies. Currently, the company offers one time use Ion 314^{TM} chip (consisting 1.0 million wells and produces 10 Mb of good quality sequence with read length of 200bp), Ion 316^{TM} (6 million wells and 100Mb data) and Ion 318^{TM} (11 million wells and 1 Gb data) (Niedringhaus et al., 2011). In the near future, company claims to release Ion Proton I^{TM} (165 million wells and ~7 GB throughput) and Ion Proton II^{TM} (660 million wells and ~4 Gb throughput) which further scale up the current throughput 1000 times higher (www.appliedbiosystems.com). The technology utilizes natural biochemistry using inexpensive reagents. The major advantage of this technology is its fast run time (2 hours). A small, light weight machine which brings sequencing to the doorstep of every lab. Short read length is the major problem with this sequencing tool which would create a significant burden on assembly process. Also, error accumulation is remarkably high if reaction wells are not properly cleansed between reaction steps. In this technology, homopolymer region is the most error prone area. The sequencing accuracy of 5-mer homopolymer region of *E. coli* DH10B sequenced through Ion torrent reported accuracy of around 97.5%.

Notably, both NGS and NNGS technologies differ in features like template preparation involving clonal amplification of immobilized template by NGS technology, while NNGS technology involves single molecule detection, thereby bypassing the biasness introduced by PCR amplification. The read length offered by NGS technologies are very low ranging from 70-400bp. In contrast, NNGS provides read length greater than 1000 bp and even more. Moreover, few of the NNGS involves real time detection of fluorescent dye in polymerase active site during incorporation. Even two of the technologies (Nanopore and Ion torrent) do not need modified bases, for this reason optics need is eliminated in these NNGS technologies. On the other hand NGS technology needs optics as they are dependent on fluorescently labeled nucleotide for detection. In addition, NNGS technologies offer very high throughput of the rate 100 Gb per hour while NGS provides only 0.4 to 20 Gb per run. Despite moderate read accuracy offered by NNGS technologies they are cost effective when assessed with NGS technologies. Overall, whatever the firms dealing with NNGS technologies claim regarding their performance, it's too early to tell whether these third generation sequencers will truly provide a breakthrough 'if the technology works' and could completely turn the things around for genome sequencing, but there are a lot of ifs and buts.

Applications of NGS technologies in crop improvement

Recent advances in sequencing technologies have revolutionized the field of crop genomics. NGS and NNGS (to make it easy for reading, hereafter both NGS and NNGS technologies will be considered as NGS technologies) technologies provide opportunities to generate vast amount of sequence data in a short lap of time at substantially lower cost. These advances in sequencing technologies are not only used for *de novo* sequencing and resequencing of various crop species but also in the area of plant breeding, mutation mapping, transcript profiling study of small RNAs and protein-DNA interaction. These technologies have also been applied for developing single nucleotide polymorphism (SNP) based markers in N number of plant species irrespective of reference genome availability or not. Efforts are underway employing NGS in metagenomic studies, epigenetic modification and gene mining.

Nascent sequencing of crop species

A major landmark in the history of genomics and molecular biology was the sequencing *Arabidopsis thaliana*, the first plant genome to be sequenced. It is now used as a model plant in molecular biology for understanding several plant traits. *Arabidopsis* offered several merits over other, being smaller in size (125Mbp), short generation time (annual), and high efficiency of transformation. Scientists from Japan and USA joined hands for sequencing this model plant, in 1996 (Bevan, 1997) and within four years completed and published its genome sequence (The Arabidopsis Genome Initiative, 2000). Next in the que was the draft genome sequencing of *indica* and *japonica* rice, an important cereal crop as well as a model monocotyledon that was published in 2002 (Goff et al., 2002; Yu et al., 2002). With the successful completion of whole genome sequencing of arabidopsis and rice, many more genome initiatives were formed to sequence other important food crops of the world (Table 2). In 2008, shortly after 1000 human genome project (www.1000genomes.org), 1000 plant genome sequencing project was initiated under international collaborations to sequence many of economically and medicinally important plant species (www.1000plantgenomes.org). Recently, potato, world's most important non grain food crop, with a genome size of 844Mbp has been sequenced by potato genome sequencing consortium (Xu et al., 2011). Pigeonpea, the first "non-industrial crop" and the second food legume (after soybean), has been sequenced successfully by collaboration of

international scientists as International Initiative for Pigeonpea Genomics (IIPG) (Varshney et al., 2011). Bejing Genome Institute (BGI) has very recently completed the whole genome sequencing of fox tail millet (*Setaria italica*), a minor millet crop providing food and feed for the arid and semi arid countries (Zhang et al., 2012). The whole genome sequencing of many other polyploidy crop plants including barley, peanut, sugar beet, cherry, coffee, water melon, raspberry etc. is under progress. *De novo* sequencing provides a golden path for discovery of novel genes and templet for SNP discovery for many plant species for which reference genome of closely related species are unavailable. For large and complex genome, reduced complexity of sequencing approach provides ample sequence depth for SNP discovery without the need of complete sequence. EST is one of the several methods to reduce the complexity of sequencing templet by reducing the low information content repetitive sequences. EST sequencing is a routine method for gene discovery and EST data is valuable tool for mining of plant SNPs (Batley et al., 2007).

Resequencing of well characterized species

Species for which genome or expressed sequence tag (EST) data are available, resequencing proves a boon to scientist involving in molecular breeding. SNP discovery by whole genome and targeted genome sequencing is the very first application of resequencing. Sequencing of parental genotypes through NGS technologies aligned to reference genome to identify the variation between genotype even at SNP level, which further can be exploited to develop SNP based molecular markers. For instance sequencing of two *Arabidopsis* divergent accessions (Bur-0 and Tsu-1) (using Solexa technology) and aligning with reference accession (Col-1) resulted in 823,325 unique SNP and 79,961 unique 1-3 bp indel polymorphism (Vera et al., 2008). A whole genome resequencing project to discover whole-genome sequence variations in 1,001 accessions of *Arabidopsis* will result in a data set that will become a fundamental resource for promoting future genetics studies to identify alleles in association with phenotypic diversity across the entire genome and species range (http://1001genomes.org/) (Weigel et al., 2009). Resequencing of total 17 wild and 14 cultivated genomes with 5X depth and 90% coverage (using Illumina GA II) and comparing genetic variation between wild and cultivated soybean found higher allelic diversity in wild soybean. A total of 205,614 tag SNPs were identified that can be useful for QTL mapping and association studies (Lam et al., 2010). In resequencing of 6 elite Maize inbred lines including productive commercial hybrid parents from China uncovered more than 1,000,000 SNPs, and 30,000 indel polymorphism (Lai et al., 2010). With the acquaintance of B73 genome sequence features, gene fraction of maize genome was targeted for resequencing in founder inbred lines of the Nested Associated Mapping (NAM) population (McMullen et al., 2009). Two datasets comprising 3.3 million SNPs were used to produce first Haplotype (HapMap) which was utilized to study distribution of recombination and diversity along the maize chromosomes. This HapMap and comparative genome hybridization (CGH) experiments enabled identification of >100 low diversity regions possibly associated with domestication and geographic distribution of maize. Very recently whole genome resequencing of *Phytophthora infestans* (HP1031 strain of A2 mating type) was completed covering ~10x of its genome size (~240Mb) along with three phylotypes of *Ralstonia solanacearum* (RS 50 [phylotype I], RS2 and RS56 [phylotype II], and RS75 [phylotype IV]) covering 50x of the 5.8 Mb genome to decipher the genome wide SNP variations. The preliminary reports show very less to no repeat regions for the bacterial genome (author's work under progress). A broader view of various applications concerning resequencing of well characterized species is shown in Fig 2.

Molecular breeding

Determining DNA sequence variation within genome is highly informative for crop genetics and breeding. Genetic variation can be assayed using a variety of molecular markers. NGS proves cheap and efficient methods for identification of SNPs and SSRs (simple sequence repeats) marker development (Robinson et al., 2004; Jewell et al., 2006; Duran et al., 2009). Even partial genome sequencing would also facilitate maker assisted breeding programs for an efficient introduction of desired traits (Xu et al., 2011). Through marker assisted selection (MAS) one can select desired lines from large scale population. Once marker has been linked to trait of interest which is more economical using NGS technologies as compared to conventional methods, MAS can be used to modulate the breeding program for crop improvement. Nowadays, plant breeding is dependent on molecular markers for rapid and cost effective analysis of germplasm and trait mapping. Molecular markers enhance understanding of genetic association that can modify breeding strategy. When a desired trait is under genetic control and phenotypic trials are unsuccessful and unreliable, MAS allows breeder not only early selection of trait but also to carry forward the desired allele to a large number of population. The availability of sequence data for identifying genes through various sequencing project has lead to development of 'genic' or 'functional' markers (FMs) from transcribed region of genome which can be well utilized for extracting putative function (Varshney et al., 2007). The starting point of FM development is the structuring of gene sequence with assigned function. The candidate gene approach and

homology searching between plant genomes have been successfully employed for identifying agronomically important traits (Andersen and Lubberstedt, 2003). Moreover, GMMs (genic molecular markers) or FMs have been developed using the transcript sequence data available in public domain. Applications of these GMM are accelerating because their discovery is inexpensive and putative functions can often be extracted by homology searches. NGS methods for developing molecular markers in crop breeding can be effectively used in two circumstances; in major crop species for which genome, or transcriptome sequence data already available, and in less characterized species with no or limited genome resources (Vera et al., 2008). SNP involves finding differences between two sequences even at single nucleotide level. Traditionally SNPs were determined by PCR amplification of genes/genomic region of interest. Individuals are selected from questioned population followed by either direct sequencing of amplicons or cloning, which is further more expensive and time consuming as far as identification of large number of SNPs for application like genetic mapping and association studies are considered. Large amount of data generated by NGS technologies is a treasure for mining SNPs that can be subsequently used in developing molecular markers (Imelfort et al., 2009). Till date numerous examples of SNP discoveries in various crop species using NGS technologies are available and some of the important SNP discoveries in selected crop species are given in Table 3.

Association mapping employ one of the two approaches: candidate gene sequencing (CGS) and whole genome scanning (WGS) of natural population (Rafalski, 2002). Population surveys for haplotypes identified based on either of two approaches trait-marker relationship or linkage disquilibrium (LD) by taking advantage of past recombination events. NGS technologies have the potential to utilize both CGS and WGS approaches although such reports are yet to be published. NGS technologies made it possible to sequence large number of PCR amplicons of candidate gene for several hundred genotypes of natural population. One NGS run can generates sequence data in the form of SNPs and haplotypes for a large number of candidate genes within short time and lower cost as compared to time consuming and expensive traditional sequencing, which involves sequencing PCR amplicons for selected candidate gene from hundreds of genotypes of natural population (Varshney et al., 2009).

Evolutionary genetic studies

In recent years NGS technologies have been used to study whole population rather than just individuals. The study known as population genetics and Handelsman in 1998 coined the term *Metagenomics* for the same (Handelsman, 2004). In today's world metagenomics studies are expanding due to the decreasing cost of sequencing. It has the power of exploring the varying microbial population, community structure and composition with respect to diverse environmental condition like soil (Leininger et al., 2006), deep sea (Sogin et al., 2006) and deep mines (Edwards et al., 2006). There are certain obstacles in studying microbial composition of an environmental sample. One of the first reason is that sheer diversity of microbes that are present in most extreme environments as well as only small portion of the total microbes are culturable. Genomic studies allow molecular phylogenetic analysis of non culturable organisms which can be employed to study taxonomic diversity of organisms present also, analysis of gene content enlightens metabolic potential of an environment. Metagenomics is the field of research that allows the study of genomes recovered from various environmental samples bypassing the need for isolation and laboratory cultivation of individual species (Shakira et al., 2009). Large scale shotgun sequencing approaches allow the discovery of many novel genes found in the environments independent of cultivation efforts. Whole genome shotgun sequencing technology has been applied to study microbial population from Sargasso sea near Bermuda and identified 1.2 million of previously unknown genes including 782 new rhodopsin like receptors (Venter et al., 2004).

It has been estimated that 10 grams of soil would consist of 10^7 distinct microbial populations (Curtis et al., 2005). Large sequencing centre would have to dedicate its entire resource for years to sequence all the genomes of the species present. In soil new highly parallel sequencing technologies offer a cost effective production of genomic data as they can generate much more sequence than classical method. NGS for the first time applied to study genome sequences from two sites in Soudanmine, Minnesota USA. Comparison of microbes and subsystems identified in two samples highlighted important differences in metabolic potential in each environment with respect to carbon utilization, iron acquisition mechanism, nitrogen assimilation and respiratory pathways by assessing gene function and homology searches of sequence reads against the metabolic database. The 16S reads were used to identify species present and it was proved that oxygenated environment possess more diverse species as compared to oxygen poor environment (Edwards et al., 2006). Approximately ~118,000 amplicons were sequenced that spanned the V6 hypervariable region of ribosomal RNAs from the DNA sampled at different depth and location of Atlantic and Pacific Ocean. The resulting sequences were trimmed and compared with reference database (V6RefDB) to identify the closest match for accessing closest taxanomic diversity. Although this study was inadequate to cover the

whole diversity however efficiency of estimates of natural diversity was greater than other methods (Sogin et al., 2006). 'Zebra chip 'disease of solanaceous crops, caused by '*Candidatus liberibacter solanacearum'* (CLs) and transmitted by the potato psyllid, *Bactericera cockerelli*. Recently 16s DNA tag encoded amplicon pyrosequencing was performed to determine total bacterial microbiota in CLs-uninfected and CLs infected potato psyllid strains and potato leaf tissues and identified five bacterial species in *B. cockerelli* including the P-endosymbiont, *C. carsonella ruddii*, the facultative endosymbiont, *C. carsonella wolbachia* plant pathogen (Nachappa et al., 2011). Very recently seasonal pattern of microbial species diversity and richness was shown using samples collected from deep waters of English Channel over 6 years. Deep sequencing of V6 hypervariable region of 16S rDNA using Illumina platform, providing ~1000X coverage indicated the changes in relative abundances of taxas that are always present (Caparaso et al., 2012). NGS technology has accelerated the study of population level of plant diversity. Sequencing of 1001 accession of model plant is accomplished by using Solexa, Roche/454 and ABI SOLiD technologies expected to provide genome wide LD structure and haplotype data that have broad application for evolutionary biology and plant breeding (http://1001genomes.org/index.html).

Mini organ genome sequencing

Mitochondria and chloroplast genome represent a rich source of molecular markers for a range of applications including population genetics, systematics and ecology (Jex et al., 2008). Both these organelles are result of endosymbiosis, mitochondria are the result of group of aerobic bacteria (the α-proteobacteria), while chloroplasts are plant cell organelle of cynobacterial origin. Understanding their structure provides basis of investigating intercellular physiology and biochemistry. They have a major role in performing essential metabolic and biosynthetic functions of global significance including photosynthesis and amino acid biosynthesis (Kleffmann et al., 2004). Sequencing of nuclear as well as organelle genome from phylogenetically diverse species will help to know how these genomes have evolved and importance of their encoded genes. Till date a total of 2075 organelle genomes have been sequenced covering 1386 organisms. Of these, 1898 are complete mitochondrial and 122 are complete chloroplast genome sequences (EnterZ Genomes). Sequencing of such huge number of organelle genomes was possible only after the arrival of NGS technologies.

The first land plant whose mitochondrial DNA sequenced was that of the liverwort (*Marchantia polymorpha*) having genome size of 186 Kb (Oda et al., 1992). Very recently complete chloroplast and mitochondrial genomes of *Boea hygrometrica* was sequenced to have detailed insight into the evolution of plant organellar genomes (Zhang et al., 2012). Sequencing of these organellar genomes revealed that the smaller chloroplast genome (~150kb) contains more coding genes (147 genes covering 72% genome size) and large mitochondrial (~510.5 kb) genome contains less genes (65 genes covering 12% of genome). The study also revealed the horizontal gene transfers between the organelles may have begun early in the land plants lineage. Male sterility genes which are significant for developing hybrid crops are present in mitochondrial genome and therefore sequence analysis of its genome helps improving hybrid crop production (Varshney et al, 2009). Sequencing of chloroplast genome was first accomplished in *Nicotina tabaccum* using Sanger technology (Shinozaki et al., 1986). In photosynthetic plants, an average size of chloroplast genome range from 120 to 217 kb with most angiosperm species having genome size of 135 to 160kbp (Downie et al., 1991) and approximately 130 genes, the chloroplast genome can easily accommodate (Cronn et al., 2008). High degree of sequence conservation, strong purifying selection, action on photosynthetic machinery, imposes clear constraints on nucleotide and structural mutation. Structural changes in non coding regions are often used to study population differentiation (Wolfe et al., 1987; Kapralov et al., 2007). Moore et al., (2006) sequenced chloroplast genome of *Nandina domestica* (24.6X coverage) and *Plantaus accidentalis* (17.3X coverage) using pyrosequensing with 99.7% assembled of predicted genome length. Very recently chloroplast genomes of *Magnolia officinalis* (Li et al., 2012) and *M. kwangsiensis* (Kuang et al., 2011) were sequenced by using HT-NGS and classified phylogenetically based on shared 81 coding genes among the magnoliids. These studies show the efficiency of species classification based on the orgalles genome sequencing.

Plastid transformation is recently developed as an attractive alternative to nuclear gene transformation due to certain advantages like high copy number, high protein expression, feasibility of expressing multiple proteins and gene containment through lack of pollen transmission (Shabir et al., 2010). For plastid engineering, obtaining plastid genome sequence is crucial (Diekmann et al., 2008). Researchers demonstrated nutritional enhancement to various biotic and abiotic food crops by plastid genome engineering in tomatoes by inducing lycopine to provitamin A conversion (Apel et al., 2009). The plastid genome engineering could be useful for improving agronomic traits including phytoremediation (Ruiz et al., 2003), reversible male sterility (Ruiz et al,. 2005) and resistance of stresses like disease (De Gray et al., 2001), insect pest (Chakrabarti, 2006), herbicides (Roudsari et al., 2009), drought and

salinity (Djilianov et al., 2005). Plastid engineering has also been utilized for metabolic pathway engineering and in the field of molecular farming (production of drugs and chemicals through engineered crops) (Bock, 2007) expression and production of biomaterials, biopharmaceuticals, therapeutic proteins and edible vaccines in plants (Millan et al., 2003; Tregoning et al., 2004; Watson et al., 2004; Koya et al., 2005).

Functional genomics in model crops

Year 2000, marks the acceleration of global research on *Arabidopsis* as its sequencing revolutionized the understanding of plant systems and its basic mechanism associated with the development, and for developing tolerance to biotic, abiotic stress and adaptation. Many of these pathways are common to all plants, Arabidopsis genes can be used in heterologous system as a candidate gene for identifying orthologs in other crops. The most successful gene translation to improve crop traits was achieved with Arabidopsis in the abiotic stress tolerance and primarily with transcriptional factors because of axis role in controlling cellular process (Feuillet et al., 2011). CRT (C-repeat) binding factors (CBFs) from *Arabidopsis* were expressed in tomato, rapeseed, strawberry, rice and wheat for improving quality of crops associated with freezing, salt and drought tolerance (Zhang et al., 2004). However, such translational biology found unsuccessful for disease resistance due to existence of two resistance mechanism in plants Pathogen-associated molecular pattern–triggered immunity (PTI) and effector triggered immunity (ETI) as both are superimposed in plants. PTI is more utilized for improving traits in tomato and tobacco by engineering of broad spectrum disease resistance by employing expression of Arabidopsis elongation factor Tu receptor (EFR) (Lacombe et al., 2010). In certain examples candidate gene screening with sequencing of structural/functional orthologs for closely related species like *Arabidopsis* underlying basic traits has been identified. Nelson et al., (2007) obtained drought tolerance after transforming maize with maize ortholog (ZmNF-YB2) of transcription factor AtNF-YB1 that was identified by screening of drought tolerance in *Arabidopsis. In-silico* comparison between sequences enable for speedy and cost effective gene identification. This was well understood after completion of rice genome sequencing and its comparison with *Arabidopsis,* which accelerated rice biology related discoveries (Rensink et al., 2004).

Sequenced legume genome and many other crop species provide rich opportunities for translational biology. The genome sequence of three legume crops, soybean (*Glycine max*) (Schmutz et al., 2010), barrel medic (*Medicago truncatula*) (http://www.medicago.org/) and birdsfoot trefoil (*Lotus japonicas*) provide wealthy opportunities for translational biology (Sato et al., 2008). Soybean, a valuable protein rich and edible oil source crop while other two are forage crop. Yang et al., (2008) used barrel medic to map based clone of *RCT1* (for resistance to *C. trifolii*) gene that confers the resistance to multiple species of anthracnose (*Colletotrichum trifolii*). TERMINAL FLORAL 1 (TLF1) gene a floral regulatory gene, identified in Arabidopsis was used to find gene responsible for determinacy trait in common bean (*Phaseolus vulagris*) (Kwak et al., 2008). Medicago and lotus have been extensively utilized in studies of nodulation, mycorrhization and plant symbiont signaling while studies are going on to understand the phenylpropanoid and isoflavinoid pathways and secondary metabolites (Farag et al., 2008) various defense responses (Yang et al., 2008), abiotic stress tolerance (Chandran et al., 2008) with the help of barrel medic system.

Agronomic treasure gene mining

Though the DNA sequence reveals precious mysteries about the genome but time, cost and cumbersomeness was the major drawback of earlier versions. With the inception of faster, simple and cheaper sequencing platforms and progress in plant breeding in terms of development of superior and high yielding varieties of agricultural crops, vast amount of sequence information has been available in public databases. It is important to use this genomic information for identification of novel and superior agronomically important genes to develop improved varieties. Structuring of rice genome by clone by clone shotgun approach produced massive amount of sequence data which was very helpful in discovering new functional genes controlling important agronomic traits. Discovery of new genes in rice by sequencing of EST resulted in out characterization of new rice small GTP binding protein coding (*osrab5B*) (Lin et al., 2001) and Glucose-6-phosphate dehydrogenase genes (Huang et al., 2002). In cultivated barley (*Hordeum vulgare* L.) five SNP sites corresponding to substitution in protein sequence of β-amylase gene (*Bmy 1*) were genotyped by pyrosequencing and CAPS assay. As an outcome, six different haplotypes of the gene *Bmy1* were discovered of which four were identified as previously described alleles *Bmy1-Sd1, Bmy1-Sd2L, Bmy1-Sd2H and Bmy1-Sd3,* while two were newly discovered (Malysheva et al., 2006).

Given the availability of low cost, more read length and high throughput sequencing platforms, sequencing-based allele mining in future, would result in faster generation of allelic data at a cheaper cost. In addition, sequencing

based allele mining of specific genes in identified accessions and their association with phenotypic variations provides tremendous impetus to precision breeding programs in crop plants.

Table 1 Special Features of NGS technologies

NGS technologies	Approach	Read Length (bp)	Bp per run	Quality	Cost per Mb ($)	Sources of error
Pyrosequencing	Pyrosequencing	400-800	800-1000 Mb	10^{-4}-10^{5}	45.00	Amplification, mixed beads, intensity thresholding, homoplymers, neighbor interference
Illumina	Sequencing by synthesis with reversible terminators	100	600 Gb	10^{-2}-10^{-3}	5.97	Amplification, mixed clusters/neighbor, interference, phasing, base labeling
SOLiD	Massively parallel sequencing by ligation	50	1-4 Gb	10^{-2}-10^{-3}	5.81	Amplification, mixed beads, signal decline, neighbor interference

Table 2 Applications of NGS technologies for SNP discovery in various crops

Species	Technology involved	Details	SNPs identified	References
Maize	Pyrosequencing	Transcriptomes of apical meristem from 2 inbred lines B73 (260,000 ESTs) and Mo17 (280,000 ESTs) were sequenced and aligned	~7000 SNPs	Barbazuk et al., (2007)
Arabidopsis	Illumina	120-173 million reads from three natural variants yielded were aligned to reference genome	823,325 unique SNPs	Ossowski et al., (2008)
Eucalyptus	Pyrosequencing	148 Mb of ESTs were assembled from multiple genotypes	23,742 SNPs	Evandro et al. (2008)
Wheat	Pyrosequencing	two hexaploid lines were sequenced, assembled and compared with ancestors polyploid	~1000 SNPs	Akhunova et al., (2009)
Soyabean	Illumina	Whole genome of wild soybean (*Glycin soja* IT182932) was sequenced, assembled and compared with *G. max* genome	2,504,985 predicted SNPs	Moon et al., (2010)
Pigeonpea	Illumina	128.9 million transcript reads sequenced and aligned from 12 different genotypes	28,104 novel SNPs	Varshney et al., (2012)
Peach	Pyrosequencing and Illumina	Three parents viz 'Dr. Davis', 'F8, 1-42' and 'Georgia Belle' were sequenced to add SNPs segregating in two breeding populations producing 43.2X coverage	6654 SNPs were discovered on all scaffolds with ~1 SNP/40,000 nucleotide bases	Ahmed et al., (2011)

Rubber tree	Pyrosequencing	Deep sequencing of rubber tree (*Hevea brasiliensis*) using 454 for SNP development	5883 putative SNPs of 10 biallelic SNP markers were validated	Pootakham et al., (2011)
Peach	Illumina, Pyrosequencing	Whole genome resequencing of 56 peach breeding accessions	Of 1,022,354 SNPs identified filtered to get 8,144 SNPs distributed over eight chromosomes with an average spacing 26.7kb	Verde et al., (2012)
Chickpea	Illumina	Transcriptome sequencing and aligning of root tissues of drought tolerant (ICC 4958) and drought sensitive (ICC 1882) yielded 5.2 and 3.6 million sequence reads	~500 SNPs	Azam et al., (2012)

Table 3 *De nova* sequencing of selected crop species since the arrival of NSG technologies.

Species	Genotype	Genome size (Mb)	Ploidy	Sequencing strategy	Coverage	Predicted Genes	Consortium	Year of completion	References
Vitis vinifera (Grapevine)	ENTAV 115	505	Diploid	Sanger and Pyrosequencing	6.5/4.2X	29,585	Multicentre collaboration	2007	Velasco et al., (2007)
Cucumis sativus (Cucumber)	IL 9930	367	Diploid	Sanger and Illumina	72.2X	26,682	Chinese Academy of Agricultural Sciences, Beijing	2009	Huang et al., (2009)
Solanum lycopersicum (Tomato)	Esculentumx pennellii	950	Diploid	Pyrosequencing and Illumina	22X	~40,000	International Solanaceae Genome Project (SOL)	2009	Mueller et al., (2009)
Malus domestica (Apple)	Golden delicious	742	Dihaploid	Sanger and Pyrosequencing	16.9X	57,386	International consortium	2010	Velasco et al., (2010)
Theobroma cacao (Coffee)	B97-61/B2	430	Diploid	Illumina, Pyrosequencing and Sanger	60.7X	28,798	International Cocoa Genome Sequencing consortium (ICGS)	2010	Argout et al., (2010)
Jatropha curcas (Jathropa)	-	410	Diploid	Pyrosequencing, Illumina and Sanger	ND	40,929	Kazusa DNA Research Institute	2010	Sato et al., (2011)
Ricinus communis (castor bean)	-	320	Diploid	Solexa sequencing	4.6x	31,237	Multicentre collaboration	2010	Chan et al., (2010)
Medicago truncatula (Barrel medic)	Jemalong A17	500	Diploid	Sanger and Illumina	40X	62,388	Multicentre collboration	2011	Young et al., (2011)
Solanum tuberosum (Potato)	RH89-039-16	844	Diploid	Pyrosequencing, Illumina and Sanger	~70X	39,031	Potato genome sequencing consortium(PGSC)	2011	Xu et al., (2011)
Cajanus cajan (Pigeonpea)	Asha	833.07	Diploid	Illumina	163.4X	48,680	Multiple collaboration	2011	Varshney et al., (2012)
Phoenix dactylifera (Date palm)	Khalas	658	Diploid	Illumina	53.4X	>25,000	Genomics Core, Qatar	2011	Al-Dous et al. (2011)

Arabidopsis parvula	Tuz Golu line	140	Diploid	Pyrosequencing and Illumina	50X	28,901	multicenter collaboration	2011	Dassanayake et al., (2011)
Brassica rapa (Turnip mustard)	Chiifu-401-42	284	Diploid	Illumina and Sanger	72X	41,174	*Brassica rapa* sequencing project consortium	2011	Wang et al., (2011)
Cannabis sativa (Hemp)	Purple kush	534	Diploid	Pyrosequencing and Illumina	110X	30,074	multicenter collaboration	2011	Bakel et al., (2011)
Fragaria vesca (Wild strawberry)	H4×4	240	Diploid	Pyrosequencing and Illumina	39X	34,809	Multicentre collaboration	2011	Shulaev et al., (2011)
Citrus sinensis (Sweet orange)	Carrizo	367	Tetraploid	Pyrosequencing	3.5X	31,863	USDA Public Citrus Genome Database	2011	Belknap et al., (2011)
Setaria italic (Foxtail millet)	Zhang gu	~490	Diploid	Illumina	8X	38.801	Beijing Genomics institute	2012	Zhang et al., (2012)
Musa acuminata (banana)	DH-Pahang	*523*	Triploid	Pyrosequencing and Illumina	70X	36,542	Multicentre collaboration	2012	D'Hont et al., (2012)
Hordeum vulgare (barley)	Morex	5.1 Gb	Diploid	Pyrosequencing and Illumina	>50X	26,159	The International Barley Genome Sequencing Consortium	2012	The International Barley Genome Sequencing Consortium
Cucumis melo (Muskmelon)	*DHL92*	450	Diploid	Pyrosequencing , Illumina and Sanger	13.5x	27,427	Multicentre collaboration	2012	Garcia-Masa et al., (2012)
Nicotiana benthamiana	Nb-1	3 Gb	Allotetraploid	Illumina	63X	34,739	Multicentre collaboration	2012	Bombarley et al., (2012)
Beta vulgaris (sweet potato)	KWS2320	414-758	Diploid	Pyrosequencing, Illumina and Sanger	84.23X	27,421	Multicentre collaboration	2013	Dohm et al., (2013)
Cicer arientinum (chickpea)	Desi type	740	Diploid	Pyrosequencing	18X	27,571	National Institute of	2013	Jain et al.,

				and Illumina			Plant Genome Research, New Delhi, India		(2013)
Sesamum indicum L	Yuzhi 11	354	Diploid	Illumina and Pyrosequencing	82.9%	86,222	Sesame Genome Working Group (SGWG)	2013	Zhang et al., (2013)
Citrullus lanatus (*water melon*)	inbred line 97103	425	Diploid	Illumina	>100x	23,440	Multicentre collaboration	2013	Guo et al., (2013)

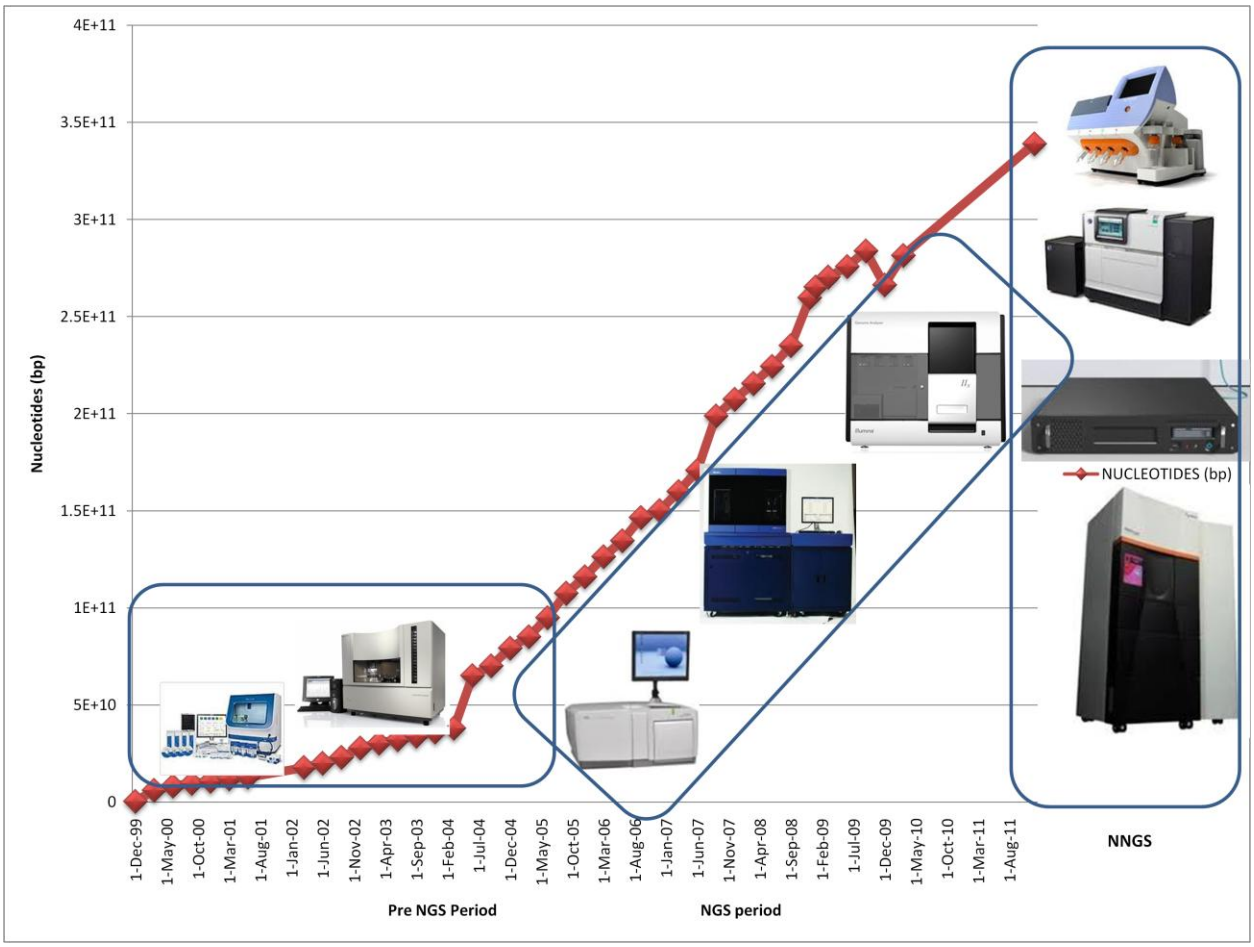

Figure 1 Increase in sequence data deposited per year from 1999 to 2011 (data as per NCBI/EMBL)

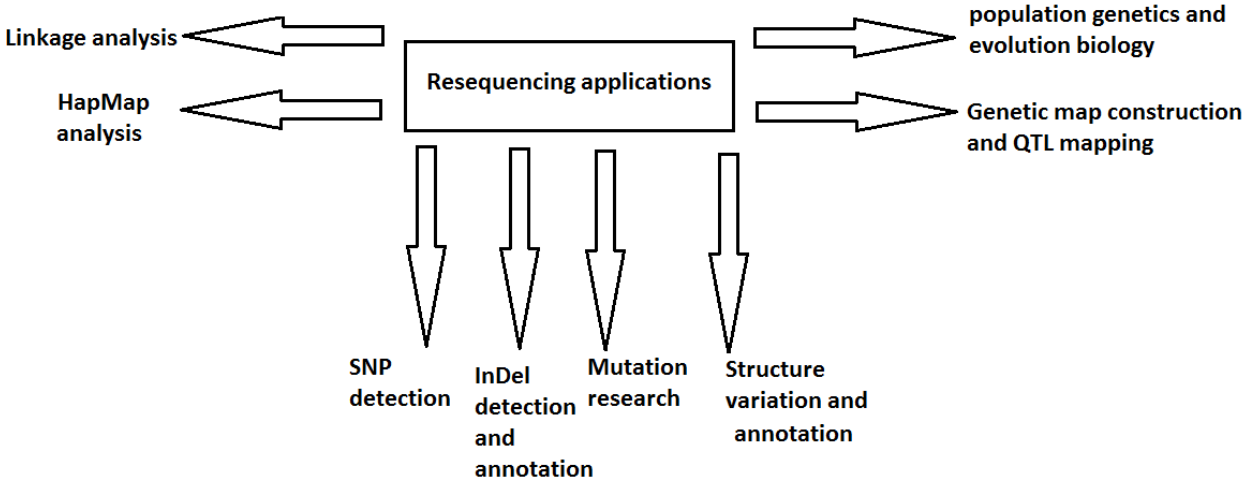

Figure 2. Applications based on resequencing of crop species.

Conclusion

Although HT-NGS and NNGS have read length ranging from 25bp to 1000bp, but for genomic sequencing and for analysis of the ever more important structural genetic variation in genomes such as copy number variations, chromosomal translocations, inversions, large deletions, insertions and duplications it would be a great advantage if sequence read length in the single DNA molecule could be increased to several 1000 bases. Plant breeding has a major role to play in increasing global food production while tackling the issues of limited land and water resources and changing climate. While the molecular era has laid the foundation for molecular breeding, the advent of genomic tools and technologies has been providing unprecedented capabilities for understanding the molecular basis of plant growth, development and key traits towards improving crop productivity. The unending technological advancements in HT-NGS analysis is not only setting benchmark in the advancement of crop genomics researches, but also in the field of proteomics and other omics. It's not only the improvement of sequencing technologies producing thousands of terabytes of data that could solve the problem, but the technologies required for downstream processing of this huge data should also keep in pace with sequencing technologies. The paralleled development of these two would greatly help in genomics assisted breeding for crop improvement and alleviating the world hunger at large.

Acknowledgement

Indian Council of Agriculture Research (ICAR) is thankfully acknowledged for the project funding of 'potato genome sequencing and functional genomics for tuberization and late blight resistance' which helped central potato research institute (CPRI) in collaboration with potato genome sequencing consortium (PGSC) to successfully complete the potato genome sequencing.

References

Ahmad, R., Parfitt, D.E., Fass, J., Ogundiwin, E., Dhingra, A., Gradziel, T.M., Lin, D., Joshi, N.A., Martinez-Garcia, P.J. and Crisosto, C.H. (2011): Whole genome sequencing of peach (*Prunus persica L.*) for SNP identification and selection. BMC Genomics, 12: 569-576.

Akhunova, A.R., Macmil, S.L., Qu, C., Wang, P., Wiley, G.B., Kenton, S., Roe, B.A. and Akhunov, E.D. (2009): SNP Discovery In Polyploid Wheat Using 454 Sequencing Technology. Plant & Animal Genomes XVII Conference, San Diego, CA, pp. 10-14.

Al-Dous, E.K., George, B., Al-Mahmoud, M.E., Al-Jaber, M.Y., Wang, H., Salameh, Y.M., Al-Azwani, E.K., Chaluvadi, S., Pontaroli, A.C., DeBarry, J., Arondel, V., Ohlrogge, J., Saie, I.J., Suliman-Elmeer, K.M., Bennetzen, J.L., Kruegger, R.R., Malek, J.A. (2011): De novo genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). Nature Biotechnology, 29: 521-527.

Andersen, J.R. and Lubberstedt, T. (2003): Functional markers in plants. Trends in Plant Science, 8: 554-560.

Apel, W. and Bock, R. (2009): Enhancement of carotenoid biosynthesis in transplastomic tomatoes by induced lycopene to provitamin A conversion. Plant Physiology, 151: 59-66.

Argout, X., Salse, J., Aury, J.M., Guiltinan, M.J., Droc, G., Gouzy, J., Allegre, M., Chaparro, C., Legavre, T., Maximova, S.N., Abrouk, M., Murat, F., Fouet, O., Poulain, J., Ruiz, M., Roguet, Y., Rodier-Goud, M., Barbosa-Neto, J.F., Sabot, F., Kudrna, D., Ammiraju, J.S.S., Schuster, S.C., Carlson, J.E., Sallet, E., Schiex, T., Dievart, A., Kramer, M., Gelley, L., Shi, Z., Bérard, A., Viot, C., Boccara, M., Risterucci, A.M., Guignon V, Sabau X, Axtell M J, Ma Z, Zhang Y, Brown, S., Bourge, M., Golser, W., Song, X., Clement, D., Rivallan, R., Tahi, M., Akaza, J.M, Pitollat, B., Gramacho, K., D'Hont, A., Brunel, D., Infante, D., Kebe, I., Costet, P., Wing, R., McCombie, W.R., Guiderdoni, E., Quetier, F., Panaud, O., Wincker, P., Bocs, S. and Lanaud C. (2010): The genome of *Theobroma cacao*. Nature Genetics, 43: 101-108.

Azam, S., Thakur, V., Ruperao, P., Shah, T., Balaji, J., Amindala, B., Farmer, A.D., Studholme, D.J., May, G.D., Edwards, D., Jones, J.D. and Varshney, R.K. (2012): Coverage-based consensus calling (CbCC) of short sequence reads and comparison of CbCC results to identify SNPs in Chickpea (*Cicer arietinum* Fabaceae) a crop species without a reference genome. American Journal of Botony, 99(2): 86-92.

Bakel, H., Stout, J.M., Cote, A.G., Tallon, C.M., Sharpe, A.G., Hughes, T.R. and Page, J.E. (2011): The draft genome and transcriptome of *Cannabis sativa*. Genome Biology, 12: 102-120.

Balagurusamy, V.S., Weinger, P. and Ling, X.S. (2010): Detection of DNA hybridizations using solid-state nanopores. Nanotechnology, 21: 335102.

Barbazuk, W.B., Emrich, S.J., Chen, H.D., Li, L. and Schnable, P.S. (2007): SNP discovery via 454 transcriptome sequencing. Plant Journal, 51: 910-918.

Batley, J. and Edwards, D. (2007): SNP applications in plants. Ch 6 In: Batley, J. and Edwards, D. "Association Mapping in Plants". Oraguzie N C, Rikkerink E H A, Gardiner S E and De Silva H N (eds). Springer New York, 6: 95-102.
Belknap, W.R., Wang, Y., Huo, N., Wu, J., Rockhold, D.R., Gu, Y.Q. and Stover, E. (2011): Characterizing the citrus cultivar Carrizo genome through 454 shotgun sequencing. Genome, 54: 1005-15.
Bensimon, A., Simon, A., Chiffaudel, A., Croquette, V., Heslot, F. and Bensimon, D. (1994): Alignment and sensitive detection of DNA by a moving interface. Science, 265: 2096-2098.
Bevan, M. (1997): Objective: the complete sequence of a plant genome. Plant Cell, 9: 476-478.
Bock, R. (2007): Plastid biotechnology: Prospects for herbicide and insect resistance, metabolic engineering and molecular farming. Current Opinion in Biotechnology, 18: 100-106.
Bombarely, A., Rosli, H.G., Vrebalov, J., Moffett, P., Mueller, L.A. and Martin, G.B. (2012): A draft genome sequence of *Nicotiana benthamiana* to enhance molecular plant-microbe biology research. Molecular Plant Microbe Interactions, 25: 1523-30.
Braslavsky, I., Hebert, B., Kartalov, E. and Quake, S.R. (2003): Sequence information can be obtained from single DNA molecules. Proceeding of the National Academy of Sciences USA, 100: 3960-3964.
Caparaso, J.G., Paszkiewicz, K., Field, D., Knight, R. and Gilbert, J.A. (2012): The western English Channel contains a persistent microbial seed bank. ISME Journal, 6: 1089-1093.
Chakrabarti, S., Lutz, K., Lertwiriyawong, B., Svab, Z. and Maliga, P. (2006): Expression of the cry9Aa2 Bt gene in tobacco plastids confers resistance to potato tuber moth. Transgenic Research, 15: 481-488.
Chan, A.P., Crabtree, J., Zhao, Q., Lorenzi, H., Orvis, J., Puiu, D., Melake-Berhan, A., Jones, K. M., Redman, J., Chen, G., Cahoon, E.B., Gedil, M., Stanke, M., Haas, B.J., Wortman, J.R., Fraser-Liggett, C.M., Ravel, J. and Rabinowicz, P.D. (2010): Draft genome sequence of the oilseed species *Ricinus communis*. Nature Biotechnology, 28: 951-956.
Chandra, S.P., Rafal, S. and Andrzej, T. (2011): Sequencing technologies and genome sequencing. Journal of Applied Genetics, 52: 413-435.
Chandran, D., Sharopova, N., Vanden Bosch, K.A., Garvin, D.F. and Samac, D.A. (2008): Physiological and molecular characterization of aluminum resistance in *Medicago truncatula*. BMC Plant Biology, 8: 89-106.
Cronn, R., Liston, A., Parks, M., Gernandt, D.S., Shen, R. and Mockler, T. (2008): Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. Nucleic Acids Research, 36: 122-133.
Curtis, P. and Sloan, W.T. (2005): Microbiology. Exploring microbial diversity- a vast below. Science, 309: 1331-1333.
D'Hont, A., Denoeud, F., Aury, J.M., Baurens, F.C., Carreel, F., Garsmeur, O. et al. (2012): The anana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. Nature, 488: 213-19.
Dassanayake, M., Oh, D.H., Haas, J.S., Hernandez, A. and Hong, H. (2011): The genome of the extremophile crucifer *Thellungiella parvula*. Nature Genetics, 43: 913-918.
De Gray, G., Rajasekaran, K., Smith, F., Sanford, J. and Daniell, H. (2001): Expression of an antimicrobial peptide via the chloroplast genome to control phytopathogenic bacteria and fungi, Plant Physiology, 127: 852-62.
de Zoysa, R.S., Jayawardhana, D.A., Zhao, Q., Wang, D., Armstrong, D.W. and Guan, X. (2009): Slowing DNA translocation through nanopores using a solution containing organic salts. Journal of Physical Chemistry, 113: 13332-13336.
Derrington, I.M., Butler, T.Z., Collins, M.D., Manrao, E., Pavlenok, M., Niederweis, M. and Gundlach, J.H. (2010): Nanopore DNA sequencing with MspA. Proceedings of the National Academy of Sciences USA, 107: 16060-16065.
Diekmann, K., Hodkinson, T.R., Fricke, E. and Barth, S. (2008): An optimized chloroplast DNA extraction protocol for grasses (Poaceae) proves suitable for whole plastid genome sequencing and SNP detection. PLoS One, 3: 2813-2818.
Djilianov, D., Georgieva, T., Moyankova, D., Atanassov, A., Shinozaki, K., Smeeken, S.C.M., Verma, D.P.S. and Murata, N. (2005) Improved abiotic stress tolerance in plants by accumulation of osmoprotectants-gene transfer approach. Biotechnology, 19: 63-71.
Dohm, J.C., Minoche, A.E., Holtgräwe, D., Capella-Gutiérrez, S., Zakrzewski, F., Tafer, H., Rupp, O., Sörensen, T.R., and Stracke, R. (2013): The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). Nature, 505: 7484.
Downie, S.R., Olmstead, R.G., Zurawski, G., Soltis, D.E., Soltis, P.S., Watson, J.C. and Palmer, J.D. (1991): Six independent losses of the chloroplast DNA rpl2 intron in dicotyledons: Molecular and phylogenetic implications. Evolution, 45: 1245-1259.
Duran, C., Appleby, N., Clark, T., Wood, D., Imelfort, M., Batley, J. and Edwards, D. 2009. AutoSNPdb: an annotated single nucleotide polymorphism database for crop plants. Nucleic Acids Research, 37: 951-953.

Edwards, R.A., Rodriguez-Brito, B., Wegley, L., Haynes, M., Breitbart, M., Peterson, D.M. et al. (2006): Using pyrosequencing to shed light on deep mine microbial ecology. BMC Genomics, 7: 57-70.
Evandro, N., Derek, R.D., William, G.F, Georgios J.P. et al. (2008): High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. BMC Genomics, 9: 312-324.
Farag, M.A., Huhman, D.V., Dixon, R.A. and Sumner, L.W. (2008): Metabolomics reveals novel pathways and differential mechanistic and elicitorspecific responses in phenylpropanoid and isoflavonoid biosynthesis in *Medicago truncatula* cell cultures. Plant Physiology, 146: 387-402.
Fedurco, M., Romieu, A., Williams, S. et al. (2006). BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. Nucleic Acids Research, 34: 22-35.
Feuillet, C., Leach, J.E., Rogers, J., Schnable, P.S. and Eversole, K. (2011): Crop genome sequencing: lessons and rationale. Trends in Plant Science, 16: 77-88.
Flusberg, B.A., Webster, D.R, Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., Korlach, J., and Turner, S.W. (2010): Direct detection of DNA methylation during single-molecule, real-time sequencing. Nature Methods, 7: 461-65.
Garaj, S., Hubbard, W., Reina, A., Kong, J., Branton, D. and Golovchenko, J.A. (2010): Graphene as a subnanometre trans-electrode membrane. Nature, 467: 190-93.
Garcia-Masa, J., Benjaka, A., Sanseverinoa, W., Bourgeoisa, M., Mira, G., Gonzálezb, V.M. et al. (2012): The genome of melon (*Cucumis melo L.)*. Proceedings of the National Academy of Sciences USA, 109: 11872-11877.
Goff, S.A., Ricke, D., Lan, T.H., Presting, G. et al. (2002): A draft sequence of the rice (*Oryza sativa L*. spp. *japonica*). Science, 196: 92-100.
Guo, S., Zhang, J., Sun, H., Salse, J., Lucas, W.J. et al. (2013): The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. Nature Genetics, 45: 51-58.
Gupta, P.K. (2008): Single-molecule DNA sequencing technologies for future genomics research. Trends in Biotechnology, 26: 602-11.
Handelsman, J. (2004): Metagenomics: application of genomics to uncultured microorganisms. Microbiology and Molecular Biology Reviews, 68: 669-85.
Hardin, S.H. (2008): ch8 Real-time DNA sequencing. In Next-Generation Genome Sequencing: Towards Personalized Medicine, Janitz M (eds). Weinheim: Wiley-VCH. 8: 97-102.
Harrell, C.C., Choi, Y., Horne, L.P., Baker, L.A., Siwy, Z.S. and Martin, C.R. (2006): Resistive-pulse DNA detection with a conical nanopore sensor. Langmuir, 22: 0837-43.
Harris, T.D., Buzby, P.R., Babcock, H., Beer, E., Bowers, J. et al. (2008): Single-molecule DNA sequencing of a viral genome. Science, 320: 106-109.
Hart, C., Lipson, D., Ozsolak, F., Raz, T., Steinmann, K., Thompson, J. and Milos, P.M. 2010. Single-molecule sequencing: sequence methods to enable accurate quantitation. Methods in Enzymology, 472: 407-30.
Hayden, E.C. (2012): Nanopore genome sequencer makes its debut. Nature letter doi:10.1038/nature.10051.
Huang, J., Wang, J.F., Zhang, H.S., Cao, Y.J., Lin, C.F., Wang, D. and Yang, J.S. (20020: *In silico* cloning of glucose-6-phosphate dehydrogenase cDNA from rice (*Oryza sativa L*.). Acta Genetica Sinica, 29: 1012-16.
Huang, S., Li, R., Zhang, Z., Li, L., Gu, X. et al. (2009): The genome of the cucumber, *Cucumis sativus* L. Nature Genetics, 41: 1275-81.
Hyman, E.D. (1988): A new method of sequencing DNA. Analytical Biochemistry, 174: 423-36.
Imelfort, M., Duran, C., Batley, J. and Edwards, D. (2009): Discovering genetic polymorphisms in next-generation sequencing data. Plant Biotechnology Journal, 7: 312-17.
Jain, M., Misra, G., Patel, R.K., Priya, P., Jhanwar, S., Khan, A.W., Shah, N., Singh, V.K. et al. (2013): A draft genome sequence of the pulse crop chickpea (*Cicer arietinum* L.). Plant Journal, 74: 715-29.
Jewell, E., Robinson, A., Savage, D., Erwin, T., Love, C.G., Lim, G.A.C., Li, X., Batley, J., Spangenberg, G.C. and Edwards, D. (2006): SR Primer and SSR taxonomy tree: Biome SSR discovery. Nucleic Acids Research, 34: 656-59.
Jex, A.R., Hu, M., Timothy, D., Littlewood, J., Waeschenbach, A. and Gasser, R.B. (2008): Using 454 technology for long-PCR based sequencing of the complete mitochondrial genome from single *Haemonchus contortus* (Nematoda). BMC Genomics, 11: 9-11.
Kang, X.F., Gu, L.Q., Cheley, S. and Bayley, H. (2005): Single protein pores containing molecular adapters at high temperatures. Angewandte Chemie, 44: 1495-99.
Kapralov, M.V. and Filatov, D.A. (2007): Widespread positive selection in the photosynthetic Rubisco enzyme. BMC Evolutionary Biology, 7: 73-83.
Kapranov, P., Chen, L., Dederich, D., Dong, B., He, J., Steinmann, K.E., Moore, A.R., Thompson, J.F. et al. (2011): Native molecular state of AAV vectors revealed by single molecule sequencing. Human Gene Therapy, 23: 46-55.

Kapranov, P., Ozsolak, F., Kim, S.W., Foissac, S., Lipson, D., Hart, C., Roels, S., Borel, C., Antonarakis, S.E., Monaghan, A.P., John, B. and Milos P.M. (2010): New class of gene-termini associated human RNAs suggests a novel RNA copying mechanism. Nature, 466: 642-46.
Karamohamed, S., Nilsson, J., Nourizad, K., Ronaghi, M., Pettersson, B. and Nyren, P. (1999): Production, purification, and luminometric analysis of recombinant *Saccharomyces cerevisiae* MET3 adenosine triphosphate sulfurylase expressed in *Escherichia coli*. Protein Expression and Purification, 15: 381-88.
Kasianowicz, J.J., Brandin, E., Branton, D. and Deamer, D.W. (1996): Characterization of individual polynucleotide molecules using a membrane channel. Proceedings of the National Academy of Sciences USA, 93: 13770-73.
Kawano, R., Schibel, A.E., Cauley, C. and White, H.S. (2009): Controlling the translocation of single-stranded DNA through alpha-hemolysin ion channels using viscosity. Langmuir, 25: 1233-37.
Kleffmann, T., Russenberger, D., von Zychlinski, A., Christopher, W., Sjolander, K., Gruissem, W. and Baginsky, S. (2004): The *Arabidopsis thaliana* chloroplast proteome reveals pathway abundance and novel protein functions. Current Biology, 14: 354-62.
Korlach, J., Bjornson, K.P., Chaudhuri, B.P., Cicero, R.L., Flusberg, B.A., Gray, J.J., Holden, D., Saxena, R., Wegener, J. and Turner, S.W. (2010): Real-time DNA sequencing from single polymerase molecules. Methods in Enzymology, 472: 431-55.
Korlach J, Marks P J, Cicero R L, Gray J J, Murphy D L, Roitman D B, Pham T T, Otto G A, Foquet M and Turner S W. 2008. Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures. Proc Natl Acad Sci USA, 105:1176-81.
Koya, V., Moayeri, M., Leppla, S.H. and Daniell, H. (2005): Plant-based vaccine, mice immunized with chloroplast-derived anthrax protective antigen survives anthrax lethal toxin challenge. Infection and Immunity, 73: 8266-74.
Krzymanska-Olejnik, E., Kung, L., Lipson, D., Lowman, G.M., Marappa, S. et al. (2009): Virtual terminator nucleotides for next-generation DNA sequencing. Nature Methods, 6: 593-95.
Kuang, D.Y., Wu, H., Wang, Y.L., Gao, L.M., Zhang, S.Z. and Lu, L. (2011): Complete chloroplast genome sequence of *Magnolia kwangsiensis* (Magnoliaceae): implication for DNA barcoding and population genetics. Genome, 54: 663-73.
Kwak, M., Velasco, D. and Gepts, P. (2008): Mapping homologous sequences for determinacy and photoperiod sensitivity in common bean (*Phaseolus vulgaris*). Journal of Heredity, 99: 283-91.
Lacombe, S., Cardoso, A.R., Sherwood, E., Peeters, N., Dahlbeck, D., van Esse, H.P., Smoker, M., Rallapalli, G., Thomma, B.P.J., Staskawicz, B., Jones, J.D.G. and Zipfel, C. (2010): Inter family transfer of a plant pattern-recognition receptor confers broad-spectrum bacterial resistance. Nature Biotechnology, 28: 365-69.
Lai, J., Li, R., Xu, X., Jin, W., Xu, M. et al. (2010): Genome-wide patterns of genetic variation among elite maize inbred lines. Nature Genetics, 42: 1027-30.
Lam, H.M., Xu, X., Liu, X., Chen, W., Yang, G., Wong, F.L. et al. (2010). Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nature Genetics, 42: 1053-59.
Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C. and Zody, M.C. (2001): Initial sequencing and analysis of the human genome. Nature, 409: 860-921.
Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G.W., Prosser, J.I., Schuster, S.C. and Schleper, C. (2006): Archaea predominate among ammonia oxidizing prokaryotes in soils. Nature, 442: 806-09.
Li, X.W., Hu, Z.G., Lin, X.H., Li, Q., Gao, H.H., Luo, G.A. and Chen, S.L. (2012): High-throughput pyrosequencing of the complete chloroplast genome of *Magnolia officinalis* and its application in species identification. Yao Xue Xue Bao, 47: 124-30.
Lin, H., Liu, X., Li, F., Luo, W. and Liu, L. (2001): Cloning and characterization of a rice small GTP-binding protein gene Osrab5B. High Technology Letters, 3: 9-14.
Malysheva-Otto, L.V. and Roder, M.S. (2006): Haplotype diversity in the endosperm specific β- amylase gene Bmy1 of cultivated barley (*Hordeum vulgare L.*). Molecular Breeding, 18: 143-56.
Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S. et al. (2005): Genome sequencing in micro-fabricated high-density picolitre reactors. Nature, 437: 376-80.
Maxam, A.M. and Gilbert, W. (1977): A new method for sequencing DNA. Proceedings of the National Academy of Sciences USA, 74: 560-64.
McMullen, M.D., Kresovich, S., Villeda, H.S., Bradbury, P., Huihui, Li., Qi, Sun, Q. et al. (2009): Genetic properties of the maize nested association mapping population. Science, 325: 737-40.
Melamede R J, Automatable process for sequencing nucleotide. US Pat 4863849. 5 Sep, 1985.
Millan, A.F.S, Castel, A.M., Miller, M. and Daniell, H. (2003)A chloroplast transgenic approach to hyper-express and purify human serum albumin, a protein highly susceptible to proteolytic degradation. Plant Biotechnology Journal, 1: 71-79.

Moon, Y.K., Sunghoon, L., Kyujung, V. et al. (2010): Whole-genome sequencing of the undomesticated soybean (*Glycine soja Sieb*. *And Zucc*.) genome. Proceedings of the National Academy of Sciences USA, 107: 22032-37.
Moore, M., Dhingra, A., Soltis, P., Shaw, R., Farmerie, W., Folta, K. and Soltis, D. (2006): Rapid and accurate pyrosequencing of angiosperm plastid genomes. BMC Plant Biology, 6: 17.
Mueller, L.A., Lankhorst, R.K., Tanksley, S.D., Giovannoni, J.J., White, R. et al. (2009): A Snapshot of the emerging tomato genome sequence. The Plant Genome, 2: 8-92.
Nachappa, P., Levy, J., Pierson, E. and Tamborindeguy, C. (2012): Diversity of Endosymbionts in the Potato Psyllid, *Bactericera cockerelli* (Hemiptera: Triozidae), Vector of Zebra Chip Disease of Potato. Current Microbiology, 62: 1510-20.
Nelson, D.E., Repetti, P.P., Adams, T.R., Creelman, R.A., Wu, J., Warner, D.C., Anstrom, D.C. et al. (2007): Plant nuclear factor Y (NF-Y) B subunits confer drought tolerance and lead to improved corn yields on water-limited acres. Proceedings of the National Academy of Sciences USA, 104: 16450-55.
Niedringhaus, T.P., Milanova, D., Kerby, M.B., Snyder, M.P. and Barron, A.E. (2011): Landscape of next-generation sequencing technologies. Analytical Chemistry, 83: 4327-41.
Nyren, P., Pettersson, B. and Uhlen, M. (1993): Solid phase DNA minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay. Analytical Biochemistry, 208: 171-75.
Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N. et al. (1992): Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA, A primitive form of plant mitochondrial genome. Journal of Molecular Biology, 223: 1-7.
Orlando, L., Ginolhac, A., Raghavan, M., Vilstrup, J., Rasmussen, M. et al. (2011): True single-molecule DNA sequencing of a Pleistocene horse bone. Genome Research, 21: 1705-19.
Ossowski, S., Schneeberger, K., Clark, R.M., Lanz, C., Warthmann, N. and Weigel, D. (2008): Sequencing of natural strains of *Arabidopsis thaliana* with short reads. Genome Research, 12: 2024-33.
Ozsolak, F., Ting, D.T., Wittner, B.S., Brannigan, B.W., Paul, S., Bardeesy, N., Ramaswamy, S., Milos, P.M., and Haber, D.A. (2010): Amplification-free digital gene expression profiling from minute cell quantities. Nature Methods, 7: 619-21.
Ozsolak, F., Platt, A.R., Jones, D.R., Reifenberger, J.G., Sass, L.E., McInerney, P. et al. (2009): Direct RNA sequencing. Nature, 461: 814-18.
Peng, H. and Ling, X.S. (2009): Reverse DNA translocation through a solid-state nanopore by magnetic tweezers. Nanotechnology, 20: 18510-1.
Pennisi, E. (2010): Genomics. Semiconductors inspire new sequencing technologies. Science, 327: 1190.
Pootakham, W., Chanprasert, J., Jomchai, N., Sangsrakru, D., Yoocha, T., Therawattanasuk, K. and Tangphatsornruang, S. (2011): Single nucleotide polymorphism marker development in the rubber tree, *Hevea brasiliensis* (Euphorbiaceae). American Journal of Botony, 98: 337-38.
Rafalski, A. (2002): Novel genetic mapping tools in plants: SNPs and LD-based approaches. Plant Science, 162: 329-33.
Rensink, W.A. and Buell, C.R. (2004): Arabidopsis to rice. Applying knowledge from a weed to enhance our understanding of a crop species. Plant Physiology, 135: 622-29.
Ritz, A., Bashir, A. and Raphael, B.J. (2010): Structural variation analysis with strobe reads. Bioinformatics, 26: 1291-98.
Robinson, A., Love, C.G., Batley, J., Barker, G. and Edwards, D. (2004): Simple sequence repeat marker loci discovery using SSR primer. Bioinformatics, 20: 1475-76.
Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlen, M. and Nyren, P. (1996): Real-time DNA sequencing using detection of pyrophosphate release. Analytical Biochemistry, 242: 84-9.
Ronaghi, M., Uhlen, M. and Nyren, P. (1998): A sequencing method based on real-time pyrophosphate. Science, 281: 363-65.
Ronaghi, M. (2001): Pyrosequencing sheds light on DNA sequencing. Genome Research, 11: 3-11.
Roudsari, M.F., Salmanian, A.H., Mousavi, A., Sohi, H. H. and Jafari, M. (2009): Regeneration of glyphosate-tolerant *Nicotiana tabacum* after plastid transformation with a mutated variant of bacterial aroA gene. Iranian Journal of Biotechnology, 7: 247-53.
Ruiz, O.N. and Daniell, H. (2005): Engineering cytoplasmic male sterility via the chloroplast genome by expression of ketothiolase. Plant Physiology, 138: 1232-46.
Ruiz, O.N., Hussein H S, Terry N and Daniell H. 2003. Phytoremediation of organomercurial compounds via chloroplast genetic engineering. Plant Physiology, 132: 1344-52.
Sanger F, Nicklen S and Coulson A R. 1977. DNA sequencing with chain terminating inhibitors. Proceedings of the National Academy of Sciences USA, 74: 5463-67.

Sato S, Hirakawa H, Isobe S, Fukai E, Watanabe A et al. 2011. Sequence analysis of the genome of an oil-bearing tree, *Jatropha curcas L*. DNA Research, 18: 65-76.
Sato, S., Nakamura, Y., Kaneko, T. et al. (2008): Genome structure of the legume, *Lotus japonicus*. DNA Research, 15: 227-39.
Schadt, E.E., Turner, S. and Kasarskis, A. (2010): A window into third-generation sequencing, *Hum* A window into third-generation sequencing. Human Molecular Genetics, 19: 227-40.
Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson. W., Hyten, D.L., Song, O., Thelen, J.J et al. 2010. Genome sequence of the palaeopolyploid soybean. Nature, 463: 178-83.
Shabir, H.W., Nadia, H., Hitesh, K. and Singh, N.B. (2010): Plant plastid engineering. Current Genomics, 11: 500-12.
Shakira, G. and Azim, A. (2009): Metagenomics and its application in rumen ecosystem: potential biotechnological prospects. Pakistan Journal of Nutrition, 8: 1309-15.
Shendure, J. and Ji, H. (2008): Next generation DNA sequencing. Nature Biotechnology, 26: 1135-45.
Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X., Mc-Cutcheon, J.P. et al. (2005): Accurate multiplex polony sequencing of an evolved bacterial genome. Science, 309: 1728-32.
Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N. et al. (1986): The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO Journal, 5: 2043-49.
Shulaev, V., Sargent, D.J., Crowhurst, R.N., Mockler, T.C., Folkerts, O. et al. (2011); The genome of woodland strawberry (*Fragaria vesca*). Nature Genetics, 43: 109-16.
Singh, N.K., Gupta, D.K., Jayaswal, P.K., Mahato, A.K., Dutta, S., Singh, S., Bhutani, S., Dogra, V., Singh, B.P. et al. (2012): The first draft of the pigeonpea genome sequence. Journal of Plant Biochemistry and Biotechnology, 21: 98-112.
Smith, L.M., Sanders, J.Z., Kaiser, R.J., Huges, P. et al. (1986): Fluorescence detection in automated DNA sequence analysis. Nature, 321: 67479.
Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M. et al. (2006): Microbial diversity in the deep sea and the under-explored ‘‘rare biosphere’’. Proceedings of the National Academy of Sciences USA, 103: 12115-20.
Stoddart, D., Heron, A.J., Klingelhoefer, J., Mikhailova, E., Maglia, G. and Bayley, H. (2010): Nucleobase recognition in ssDNA at the central constriction of the alpha-hemolysin pore. Nano Letters, 10: 3633-37.
Tawfik, D.S. and Griffiths, A.D. (1998): Man-made cell-like compartments for molecular evolution. Nature Biotechnology, 16: 652-6.
The Arabidopsis Genome Initiative. (2000): Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature, 408: 796-15.
The International Barley Genome Sequencing Consortium. (2012): A physical, genetic and functional sequence assembly of the barley genome. Nature, 491: 711-717.
Thompson, J.F. and Steinmann, K.E. (2010): Single molecule sequencing with a HeliScope genetic analysis system. Current protocols in molecular biology, 92: 7.10.1-7.10.14
Thompson, J.F., Reifenberger, J.G., Giladi, E., Kerouac, K., Gill, J. et al. (2011). Single-step capture and sequencing of natural DNA for detection of BRCA1 mutations. Genome Research, 22: 340-45.
Travers, K.J., Chin, C.S., Rank, D.R., Eid, J.S. and Turner, S.W. (2010): A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Research, 38:159.
Tregoning, J., Maliga, P., Dougan, G. and Nixon, P.J. (2004): New advances in the production of edible plant vaccines: chloroplast expression of a tetanus vaccine antigen, TetC. Phytochemistry, 65: 989-94.
Varshney, R.K., Chen, W., Li, Y., Bharti, A.K., Saxena, R.K., Schlueter, J.A. et al. (2012): A Jackson Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nature Biotechnology, 30: 83-9.
Varshney, R.K., Mahendar, T., Aggarwal, R.K. and Borner, A. (2007). Genic molecular markers in plants: development and applications. Ch 2, In: Genomics- Assisted Crop Improvement: Genomics Approaches and Platforms. Varshney, R.K and Tuberosa, R. (eds) Vol. I, Springer. 2: 13-30.
Varshney, R.K., Nayak, S.N., May, G.D. and Jackson, S.A. (2009): Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends in Biotechnology, 27: 522-30.
Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A. et al. (2010): The genome of the domesticated apple (*Malus domestica Borkh.*). Nature Genetics, 42: 833-39.
Velasco, R., Zharkikh, A., Troggio, M., Cartwright, D.A., Cestaro, A. et al. (2007): A high quality draft consensus sequence of the genome of a heterozygous grape vine variety. PLoS One, 2: 1326.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W. and Mural, R.J. (2001): The Sequence of the human genome. Science, 291: 1304-51.
Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E. et al. (2004): Environmental genome shotgun sequencing of the Sargasso Sea. Science, 304: 66-74.
Vera, J.C., Wheat, C.W., Fescemyer, H. W., Frilander, M.J., Crawford, D.L., Hanski, I. and Marden, J.H. (2008): Rapid transcriptome characterization for nonmodel organism using 454 pyrosequencing, Molecular Ecology, 17: 1636-47.
Verde, I., Bassil, N., Scalabrin, S., Gilmore, B. and Lawley, C.T. (2012): Development and evaluation of a 9K SNP array for peach by internationally coordinated SNP detection and validation in breeding germplasm. PloS One, 7: 35668.
Wang, X., Wang, H., Wang, J., Sun, R, Wu J et al 2011. The genome of the mesopolypoid crop species *Brassica rapa*. Nature Genetics, 43: 1035-39.
Watson, J., Koya, V., Leppla, S.H. and Daniell, H. (2004): Expression of *Bacillus anthracis* protective antigen in transgenic chloroplasts of tobacco, a non-food/feed crop. Vaccine, 22: 4374-84.
Weigel, D. and Mott, R. (2009): The 1001 genomes project for *Arabidopsis thaliana*. Genome Biology, 10: 107-12.
Wolfe, K.H., Li, W.H. and Sharp, P.M. (1987): Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proceedings of the National Academy of Sciences USA, 84: 9054-58.
Xu, P., Wu, X., Luo, J., Wang, B., Liu, Y., Ehlers, J.D., Wang, S., Lu, Z. and Li, G. (2011): Partial sequencing of the bottle gourd genome reveals markers useful for phylogenetic analysis and breeding. BMC Genomics, 12: 467-77.
Xu, X., Pan, S., Cheng, S., Zhang, B., Mu, D. et al. (2011): Genome sequence and analysis of the tuber crop potato. Nature, 475: 189-95.
Yamey, G. (2000): Scientists unveil first draft of human genome. British Medical Journal, 321: 7.
Yang, S., Gao, M., Xu, C., Gao, J., Deshpande, S., Lin, S., Roe, B.A. and Zhu, H. (2008): Alfalfa benefits from *Medicago truncatula*: the RCT1 gene from *M. truncatula* confers broad-spectrum resistance to anthracnose in alfalfa. Proceedings of the National Academy of Sciences USA, 105: 12164-9.
Young, N.D., Debelle, F., Oldroyd, G.E.D., Geurts, R. and Cannon, S.B. (2011): The Medicago genome provides insight into the evolution of rhizobial symbioses. Nature, 480: 20-24.
Yu, J., Hu, S., Wang, J., Wong, G.K. et al. (2002): A draft sequence of the rice genome (*Oryza sativa L.* spp. *indica*). Science, 296: 79-92.
Zhang, G., Liu, X., Quan, Z., Cheng, S., Xu, X., Pan, S., Xie, M., Zeng, P., Yue, Z. et al. (2012): Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. Nature Biotechnology, 30: 549-54.
Zhang, H., Miao, H., Wang, L., Qu, L., Liu, H., Wang, Q. and Yue, M. (2013): Genome sequencing of the important oilseed crop *Sesamum indicum L*. Genome Biology, 14: 401.
Zhang, J.Z., Creelman, R.A. and Zhu, J.K. (2004): From laboratory to field. Using information from Arabidopsis to engineer salt, cold and drought tolerance in crops. Plant Physiology, 135: 615-21.
Zhang, T., Fang, Y., Wang, X., Deng, X., Zhang, X., Hu, S. and Yu, J. (2012): The complete chloroplast and mitochondrial genome sequences of *Boea hygrometrica*: insights into the evolution of plant organellar genomes. PloS One, 7: 30531.